

Estimas de los niveles de *autozigosis* en el hombre basadas en registros históricos y análisis genómicos. Retos y oportunidades de la tecnología molecular de alta resolución

HERNÁNDEZ C.L., ORTEGA A. Y CALDERÓN R.

Departamento de Biodiversidad, Ecología y Evolución.
Facultad de Biología. Universidad Complutense de Madrid. España.

Corresponding Author: Dra. Candela L. Hernández
Departamento de Biodiversidad, Ecología y Evolución
Facultad de Biología. Universidad Complutense de Madrid
Ciudad Universitaria 28040 - Madrid (España)
clhernan@bio.ucm.es

ABSTRACT

Keywords:

Consanguinity
Family Trees
Surnames
Genomic Inbreeding
ROHs
Demographic Histories

From the second half of 19th century onwards, population genetics has attempted to explain through theoretical models, concepts and empirical data, the evolutionary consequences and biological cost of mating between relatives. The genetic load linked to the consanguineous marriages has been traditionally assessed from pedigree reconstruction and identity of surnames of both members of the couple as a reflection of a potential biological relationship (isonymy method). These kind of results are being enriched by molecular approaches based on new high-resolution analyses of the human genome. The present survey provide chronological comparisons among those strategies used for estimating human autozygosity. Inbreeding estimates based on pedigrees and genomic data show a rather positive correlation, mainly in populations considered as human isolates. The implementation of isonymy methodologies entail limitations that generate overrated inbreeding estimates. Genomic studies have described autozygous chromosomal segments, known as *Runs of Homozygosity (ROHs)*. Analyses of these tracts allow not only assessments of recent kinship structure of populations but also a better knowledge of human demographic dynamics in an evolutionary perspective. Furthermore, *ROHs* can be analyzed in any population, not being restricted to those ones defined by a low demographic size. In conclusion, the current genomic approach aimed to evaluate human inbreeding provides an interesting, promising challenge regarding other traditional methodologies based on historical records.

Recibido: 25-04-2018

Aceptado: 02-06-2018

RESUMEN

Palabras clave:

Consanguinidad
Árboles Familiares
Apellidos
Inbreeding Genómico
ROHs
Historias Demográficas

Desde la segunda mitad del siglo XIX hasta la actualidad, la Genética de Poblaciones ha explicado mediante modelos teóricos, conceptos y datos empíricos las implicaciones evolutivas y el coste biológico de los cruza- mientos entre parientes. La carga genética ligada al fenómeno de la consanguinidad humana se ha evaluado tradicionalmente a partir de la reconstrucción de árboles familiares (o pedigríes) y mediante el análisis de la identidad de apellidos entre los miembros de una pareja, como reflejo de su relación biológica (método de la isonimia). Estos resultados se están enriqueciendo gracias a los enfoques moleculares basados en las últimas tecnologías de análisis de alta resolución del genoma humano. El presente estudio ofrece una comparación detallada y cronológicamente secuencial entre las estrategias para evaluar la autozigosis humana. Se han detectado correlaciones positivas entre las estimas del *inbreeding* basadas en pedigríes y datos genómicos, fundamentalmente en poblaciones consideradas como aislados humanos. Por otro lado, la aplicación de la isonimia alberga ciertas limitaciones que repercuten en estimas sobrevaloradas del *inbreeding* en la población general. Los estudios genómicos han descrito bloques cromosómicos de autozigosis, denominados *Runs of Homozygosity (ROHs)*, cuyo análisis permite, no solo evaluar la estructura del parentesco reciente en una población, sino también conocer dinámicas demográficas humanas con una perspectiva evolutiva. Además, los *ROHs* se pueden estudiar no solo en cualquier población sino también en aquellas de tamaño demográfico reducido. En consecuencia, el nuevo enfoque molecular dirigido al estudio del *inbreeding* humano está aportando una rica información con respecto a las metodologías tradicionales fundamentadas en los registros históricos.

Introducción

El fenómeno de la consanguinidad, entendido como la práctica de cruzamientos entre parientes biológicos, ha sido un comportamiento usual en la historia de la humanidad y constituye un claro ejemplo de interacción entre Biología y Cultura. Las reglas socioculturales prevalentes durante generaciones en la historia de una población han podido condicionar la elección de una pareja potencial. En la actualidad, el estudio de la consanguinidad en el hombre tiene un carácter multidisciplinar siendo un campo científico de especial interés para la Antropología Genética, la Genética de Poblaciones, la Antropología Social, la Genética Médica y, en general, para la gran área temática de la Biomedicina.

Los cruzamientos entre parientes biológicos o consanguíneos constituyen un componente central de la estructura marital de una población y conllevan implicaciones evolutivas de impacto indirecto. En este contexto, el proceso evolutivo orgánico se analiza a través de los cambios en las frecuencias alélicas a corto o a largo plazo (procesos evolutivos directos) y/o de las frecuencias genotípicas (procesos evolutivos indirectos). En consecuencia, la preferencialidad hacia los cruzamientos consanguíneos puede conducir a alteraciones del equilibrio genético Hardy-Weinberg, desembocando en una mayor proporción de genotipos homocigotos que los esperados bajo condiciones de panmixia.

La carga genética derivada del *inbreeding* asociado a una pérdida de eficacia biológica *-depresión endogámica* (Hartl y Clark, 2007)- puede ser relevante, en la medida en que genes mutantes recesivos y deletéreos se manifiesten más frecuentemente en los descendientes de padres emparentados y, por tanto, haya una mayor prevalencia de enfermedades mendelianas recesivas. Este escenario aparece con mayor incidencia en aquellas poblaciones de tamaño demográfico reducido, altamente endógamas y consanguíneas.

La homocigosis (autozigosis), con probabilidad variable, en los descendientes de padres consanguíneos se conoce en la literatura científica como *inbreeding* (F). El valor de F (coeficiente de *inbreeding*) es dependiente del parentesco biológico (ϕ) que une a los

miembros de la pareja. En otras palabras, cuanto más estrecho sea el parentesco de la pareja, más probabilidad de ocurrencia de riesgo aparecerá en sus descendientes, teniendo un coste biológico notable aquellos cruzamientos que albergan una consanguinidad estrecha: tío(a)-sobrina(o) (M12) y primos hermanos (M22) (Cavalli-Sforza y Bodmer, 1981). Los conceptos de los coeficientes ϕ y F son distintivos, aunque en la práctica sus valores se equiparan. El coeficiente ϕ hace referencia a la probabilidad de que los dos miembros de la pareja emparentada posean en el mismo locus un alelo homólogo, recibido simultáneamente de alguno de los antecesores comunes que comparten. Esos alelos o genes homólogos son idénticos por descendencia (IBD, *Identity by Descent*, Malécot y Blaringham, 1948). Por el contrario, el coeficiente F -descrito por S. Wright en la década de 1920- es la probabilidad de que un individuo, descendiente de padres emparentados, reciba en un *locus* dado dos alelos IBD.

Tradicionalmente, en las poblaciones donde predomina la religión católica las metodologías aplicadas para estimar la consanguinidad y el *inbreeding* se han basado en los Registros Históricos de la Iglesia Católica, en concreto, en las Dispensas Eclesiásticas. Estos documentos contienen una rica información demográfica, genética y paragenética de cada una de las parejas consanguíneas. La disponibilidad de ese importante bagaje de información ha permitido desarrollar investigaciones, con gran riqueza informativa, acerca del proceso de la consanguinidad humana, sus riesgos biológicos poblacionales y los factores que han podido influenciar su registro en el tiempo y en el espacio.

Por otro lado, el análisis de la estructura de los apellidos (la isonimia entre los miembros de una pareja) ha sido y, continúa siendo, una información de alto interés para evaluar niveles de *inbreeding* en las poblaciones humanas (Crow y Mange, 1965; Lasker y Kaplan, 1985; Pinto-Cisternas et al., 1985).

En la actualidad, existen avanzadas metodologías para estimar el *inbreeding* a nivel individual o poblacional, basadas en técnicas de secuenciación de nueva generación (*Next Generation Sequencing*, NGS) o en el genotipado masivo del genoma humano a través de marcadores bialélicos o SNPs (*Genome-Wide analysis*, GW).

Este trabajo tiene como objetivo realizar un análisis comparativo detallado y, cronológicamente secuencial, de las estrategias que conducen a la evaluación de los niveles de *inbreeding* en las poblaciones humanas, contrastando los valores de autozigosis basados en diferentes metodologías. Como fin último, se propone una aproximación al método más idóneo y efectivo para el análisis y conocimiento de la consanguinidad humana, visto este proceso como un componente esencial de la estructura marital de una población, con fuerte impacto sobre su estructura genética.

Estado actual del tema

Los árboles familiares o pedigrees como fuente de datos para evaluar la consanguinidad y el inbreeding en las poblaciones humanas

Las fuentes de información más frecuentemente utilizadas para la valoración del *inbreeding*, F en las poblaciones humanas han sido los pedigrees (adaptación al castellano del término anglosajón *pedigree*) o árboles familiares. La información sobre el parentesco que une a los dos miembros de la pareja puede ser obtenida a través de diferentes fuentes de datos, siendo, sin duda, las Dispensas Eclesiásticas, los registros más completos, los cuales están disponibles consistentemente desde el Concilio de Trento (1545-1563).

Una gran mayoría de estudios sobre consanguinidad en grandes y pequeñas áreas geográficas y poblaciones del continente europeo (p.e. Península Ibérica, Francia e Italia) han utilizado las Dispensas como fuente prioritaria (Cavalli-Sforza y Bodmer, 1981; Cavalli-Sforza et al., 2004). A partir de 1917, la Iglesia Católica eximió a los matrimonios consanguíneos con relación de parentesco lejano [tío(a)-sobrina(o) tercero (M34) ($F=1/128$) y primos terceros (M44) ($F=1/256$)] de solicitud de Dispensa. Desde 1982, únicamente las parejas con una consanguinidad estrecha (M12 y M22) requieren este permiso. La Figura 1 presenta algunos ejemplos de árboles familiares que albergan relaciones de parentesco comunes entre las poblaciones humanas.

Para el cálculo del coeficiente de *inbreeding* sólo es necesario conocer el tipo de parentesco (ϕ) de la pareja (Figura 2). El método habitualmente utilizado se basa en la siguiente formulación (Wright, 1951):

$$f = \sum_{i=1}^m \left(\frac{1}{2}\right)^k (1 + f_{CA})$$

Donde m es el número de rutas posibles a partir de todos los antecesores comunes; k es el número de individuos en cada ruta genealógica (excluyendo al descendiente del cruzamiento consanguíneo) y f_{CA} es el coeficiente de *inbreeding* del antepasado común en cada ruta genealógica. El coeficiente de *inbreeding* se denota, indistintamente, como F o f .

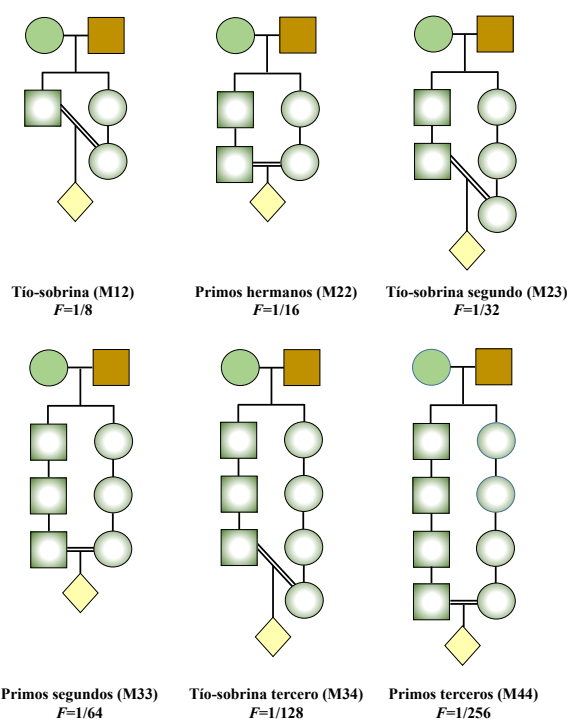


Figura 1: Algunos ejemplos de *pedigrees* o árboles familiares referidos a tipos de matrimonios consanguíneos que se registran en las poblaciones humanas. Cada tipo de *pedigree* está asociado a una nomenclatura específica, así como a su coeficiente de *inbreeding*, F para genes autosómicos. Dentro de cada árbol existen diversos subtipos genealógicos en función del sexo de los antecesores intermedios. Para el ejemplo de primos hermanos (M22) se generarían cuatro subtipos genealógicos posibles.

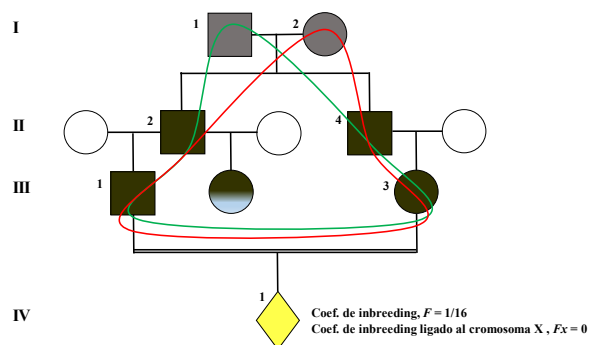


Figura 2: Árbol familiar mostrando un cruzamiento consanguíneo entre los individuos III-1 y III-3, que son primos hermanos (M22). Las líneas roja y verde trazan las rutas genealógicas conformadas por todos los miembros que contribuyen en la transmisión de la copia del gen desde alguno de los antepasados comunes (I-1 o I-2) al descendiente consanguíneo o *inbred* (IV-1). Se indica también el valor del coeficiente de *inbreeding* para genes autosómicos, F y el referido al coeficiente de *inbreeding* ligado al cromosoma X, F_x .

Asimismo, es posible calcular el coeficiente de *inbreeding* ligado al cromosoma X, F_x para lo que es necesario conocer la estructura que distingue al árbol familiar en estudio. El valor del F_x puede variar entre 0 y un valor positivo, dependiendo de la distribución de los sexos de los antepasados intermedios (Hedrick, 2005).

La fórmula se asemeja mucho a la utilizada para genes autosómicos:

$$f_x = \sum_{i=1}^n \left(\frac{1}{2}\right)^l (1 + f_{CA})$$

Donde n es m menos todas las rutas genealógicas en las que aparecen dos o más individuos masculinos consecutivos; l es el número de mujeres en cada ruta, y f_{CA} es el coeficiente de *inbreeding* del antepasado común en cada ruta.

Sobre la base de los tipos de matrimonios consanguíneos registrados en una población y el período de tiempo seleccionado por el investigador, se puede calcular el coeficiente medio de *inbreeding* \bar{F} de la población general. Dicho parámetro, omnipresente en las fórmulas de la Genética de Poblaciones, nos indica el riesgo biológico o la carga genética por consanguinidad inherente a la población analizada. El concepto del parámetro \bar{F} nos estaría revelando la probabilidad

de que, tomado un individuo al azar en la población, ese individuo haya recibido en un *locus* dado dos alelos IBD. Esta estima tiene claras implicaciones epidemiológicas.

$$\bar{F} = \sum p_i F_i$$

Siendo p_i la frecuencia del parentesco ϕ_i en relación al número de matrimonios totales (M_t) (consanguíneos + no consanguíneos) contraídos en la población bajo estudio y, F_i el coeficiente de *inbreeding* asociado al descendiente potencial de la pareja.

Haciendo uso de esta metodología, los patrones de consanguinidad y los niveles de *inbreeding* han sido estudiados con bastante profusión dentro de la Península Ibérica y, particularmente, en España, a lo largo de los dos últimos siglos (McCullough y O'Rourke 1986; Gómez, 1989; Fuster y Colantonio 2003; Calderón et al. 1993, 1998, 2009; Varela et al. 1997, 2001, 2003; Blanco Villegas et al., 2004; Alfonso-Sánchez et al., 2004, 2005). Numerosos trabajos apuntan de manera convergente a que España es el país de Europa occidental con mayores niveles de *inbreeding* y que esos niveles han disminuido más tardía y rápidamente que en el resto de Europa (Calderon et al., 1998). Las diferencias entre regiones españolas son importantes, sobre todo, entre aquellas que conforman la franja cantábrica con respecto a las no-cantábricas (ver más abajo).

Otros resultados de interés revelan que las regiones españolas predominantemente rurales presentan una alta representación de los emparejamientos del tipo M33, mientras que, en los núcleos urbanos y semiurbanos, la consanguinidad estrecha (M22 y M12) es la gran protagonista (Calderón, 1989; Fuster y Colantonio, 2003). Excepciones a estos escenarios se encuentran en la población del País Vasco (Calderón et al., 1998, 2009; Alfonso-Sánchez et al., 2003, 2005), donde la estructura marital consanguínea es distintiva, con altas frecuencias relativas de M22 y M12, así como unos niveles apreciables de matrimonios consanguíneos múltiples, es decir, con más de un parentesco. La Tabla 1 presenta algunos resultados sobre las variables claves del proceso consanguíneo para grandes áreas geográficas de España.

Tabla 1.- Estructura de la consanguinidad en un conjunto de regiones y poblaciones españolas seleccionadas. Asimismo, se muestran los valores del Índice de Preferencialidad (relación entre M22/M33) y los niveles medios de *inbreeding*, para genes autosómicos basados en el análisis de *pedigrees*.

| Población | Región | Período | Mt | Mc ^a | Tipos Mc | | | | | M22/M33 ^c | Referencia |
|--|--------------------|-----------|---------|-----------------|------------------|-------|-------|-------|-------|----------------------|----------------------------------|
| | | | | | M12 ^b | M22 | M23 | M33 | MCM | | |
| España Septentrional | | | | | | | | | | | |
| Arzob. Santiago de Compostela | Galicia | 1900-1979 | 307 094 | 15 739 | 501 | 4 970 | 1 758 | 7 857 | 653 | 0.63 | Varela et al., 1997 |
| | | | | | 3.18 | 31.58 | 11.17 | 49.92 | 4.15 | | |
| Ourense (rural) | Galicia | 1900-1979 | 110 128 | 9 010 | 171 | 2 901 | 968 | 4 400 | 570 | 0.56 | Varela et al., 2003 |
| | | | | | 1.90 | 32.20 | 10.74 | 48.83 | 6.33 | | |
| Álava (provincia) | País Vasco | 1831-1980 | 122 125 | 3 739 | 98 | 1 177 | 336 | 1 431 | 694 | 0.82 | Calderón et al., 1993 |
| | | | | | 2.62 | 31.48 | 8.99 | 38.27 | 18.56 | | |
| Guipúzcoa (provincia) | País Vasco | 1862-1995 | 300 084 | 6 638 | 270 | 2 671 | 561 | 2 185 | 951 | 1.22 | Alfonso-Sánchez et al., 2005 |
| | | | | | 4.07 | 40.24 | 8.45 | 32.92 | 14.33 | | |
| León (provincia) | Castilla y León | 1918-1968 | 71 740 | 4 500 | 0 | 1 163 | 587 | 2 523 | 227 | 0.46 | Gómez, 1989 |
| | | | | | 0 | 25.84 | 13.04 | 56.07 | 5.04 | | |
| España Central | | | | | | | | | | | |
| Toledo (provincia) / Guadalaajara (vertiente occidental) | Castilla-La Mancha | 1900-1979 | 235 984 | 10 570 | 22 | 2 688 | 793 | 6 591 | 475 | 0.38 | Calderón et al. (en preparación) |
| | | | | | 0.20 | 25.43 | 7.50 | 63.35 | 4.49 | | |
| Dióc. Sigüenza-Guadalaajara | Castilla-La Mancha | 1918-1979 | 65 720 | 6 172 | 32 | 1 729 | 787 | 3 279 | 345 | 0.52 | Morales, 1992 |
| | | | | | 0.52 | 28.01 | 12.75 | 53.13 | 5.59 | | |
| España Meridional | | | | | | | | | | | |
| Archidió. Granada | Andalucía | 1900-1979 | 277 676 | 15 440 | 78 | 5 456 | 2 002 | 7 139 | 765 | 0.76 | Núñez, 2015 |
| | | | | | 0.51 | 35.34 | 12.97 | 46.24 | 4.95 | | |

^aNúmero de matrimonios consanguíneos totales, expresados en valores absolutos y relativos (Mc/Mt x 100)

^bNúmero de matrimonios consanguíneos de cada tipo, en valores absolutos y relativos (M**/Mc x 100)

^cLa relación M22/M33 bajo equilibrio Hardy-Weinberg es 0.25

El trabajo de McCullough y O'Rourke (1986) permitió detectar en Europa la existencia de una correlación negativa ($r = -0.738$) entre el coeficiente medio de inbreeding (\bar{F}) y la latitud, interpretando este resultado mediante factores históricos y culturales. Los coeficientes de inbreeding más altos parecían registrarse en el Mediterráneo europeo y, los más bajos, en la vertiente atlántica más septentrional. En el caso del área mediterránea, se observa, en general, un declive de las uniones consanguíneas desde 1900, lo cual se atribuye al aumento de la movilidad, industrialización y otros cambios socioeconómicos.

Como el desarrollo de este tipo de trabajos conlleva especiales dificultades y costes en aquellas poblaciones de tamaño demográfico elevado, el análisis del *inbreeding* humano basado en árboles familiares se ha visto enriquecido por el uso de otros métodos, basados en el estudio de los apellidos.

El cálculo de los niveles de inbreeding mediante el estudio de la semejanza de apellidos (isonimia) entre los individuos de una población

El valor predictivo que poseen los apellidos para evaluar el nivel de *inbreeding* ha llevado a muchos investigadores a realizar estudios sobre la *isonimia*. La primera investigación de naturaleza estadística sobre este tema se remonta a 1875, cuando G.H. Darwin (hijo de Charles Darwin) estimó la frecuencia de cruzamientos entre primos hermanos mediante la proporción de individuos que compartían el mismo apellido en una población (Lasker, 1985).

Los apellidos se han comparado con genes neutrales, ya que no afectarían al fenotipo y, por tanto, no estarían sometidos a la acción de la selección natural. Asimismo, el patrón de herencia patrilineal de los apellidos les convierte en buenos indicadores de movimientos migratorios humanos. Dentro de este marco, uno de los enfoques más modernos sobre apellidos ha sido la observación de su posible relación con marcadores del cromosoma Y (C-Y) (Sykes e Irven, 2000; King y Jobling, 2009; Larmuseau et al., 2012; Martínez-Cadenas et al., 2016). Un estudio recientemente publicado por nuestro equipo de investigación ha puesto de manifiesto que cada apellido está generalmente asociado a varios linajes (haplogrupos) del C-Y o que,

por el contrario, cada linaje paterno está relacionado con varios apellidos, observándose, por tanto, una baja concordancia entre apellidos y marcadores moleculares patrilineales. La razón central podría estar en la condición mayoritariamente *polifilética* de los primeros (Calderón et al., 2015).

Existen numerosos Registros, de distintas características, a partir de los cuales se puede obtener información sobre la estructura de los apellidos y su evolución histórica dentro de un área geográfica, siendo los Registros de Matrimonios (*Libros de Nupcialidad*) los más valiosos. Por otro lado, también se emplean los *Libros de Defunciones*, los *Libros de Bautismo* (*Libros de Nacimiento*, en el Registro Civil) y los *Directorios Telefónicos* o los *Censos* poblacionales. En estos últimos, se registran todos los individuos de la población, aunque poseen la limitación (en muchos países) de que las mujeres casadas quedan ligadas al apellido del esposo.

Crow y Mange (1965) propusieron por vez primera el método de la isonimia para la estimación de los niveles de inbreeding en una población humana. Los autores se basaron en la isonimia marital, es decir, la frecuencia de cruzamientos en los que la pareja comparte el mismo apellido, asumiendo que el mecanismo de transmisión de los apellidos, de una generación a la siguiente, se asemeja al patrón de herencia genética patrilineal. La ratio (proporción) que existe entre la probabilidad de que dos individuos posean el mismo apellido (p) y su potencial coeficiente de inbreeding (F) es ≈ 4 . Por tanto, sería posible estimar el inbreeding promedio de una población seleccionada, a partir de la frecuencia de cruzamientos isónimos, con la fórmula $F = p/4$ (Lasker, 1985).

El método de Crow y Mange está basado en registros biodemográficos de países cuyos miembros son identificados por un único apellido. Otros investigadores (Pinto-Cisternas et al., 1985) intentaron extrapolar ese procedimiento a otros países cuyos ciudadanos se identifican por dos apellidos, como es el caso de España y aquellos de su ámbito cultural, en los cuales se adoptó el denominado sistema de apellidos iberoamericanos (*IASS*, *Iberoamerican Surname System*). El *IASS* consiste en que una persona hereda sus apellidos tanto del padre como de la madre, siendo el del padre el que se posiciona en primer lugar. Las mujeres casa-

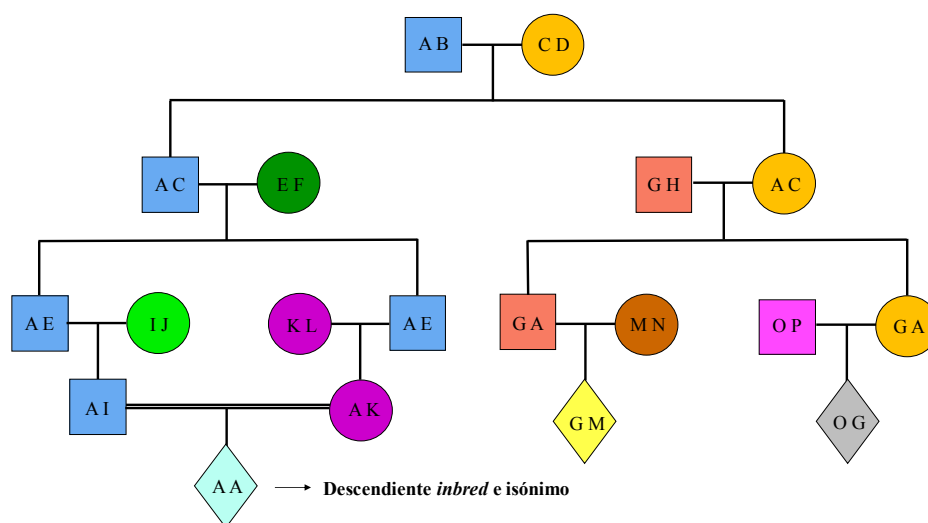


Figura 3: Mecanismo de transmisión del Sistema Iberoamericano de Apellidos (IASA)

Cada letra hace referencia a un apellido distinto, observándose una herencia patrilineal de éstos a la generación siguiente. El apellido paterno siempre se hereda en primer lugar con respecto al materno, lo que hace que este último se pierda en la siguiente generación. El individuo “AA” es descendiente de un cruzamiento consanguíneo del tipo M33 (primos segundos).

das no adquieren, en estos países, el apellido del marido y, al coincidir con el de su abuelo, permite estudiar la isonimia de dos generaciones atrás. El segundo apellido de cada progenitor se pierde en la generación siguiente (ver Figura 3).

Por tanto, la identidad de los apellidos de una persona puede indicarnos a priori si es descendiente de un cruzamiento consanguíneo o no. Sin embargo, para confirmar su validez se han de cumplir tres premisas: i) Los apellidos han de transmitirse de manera regular a lo largo de las generaciones, ii) El número de hombres y mujeres de la población ha de ser similar y, iii) Los apellidos analizados deben tener origen monofilético. Estas premisas raramente se cumplen de manera estricta.

Un número considerable de trabajos sobre isonimia han sido realizados en poblaciones españolas de tamaño demográfico reducido y sometidas a un relativamente alto grado de aislamiento. Algunos ejemplos, como la población de la Comarca de La Cabrera en la región de Castilla y León (Boattini et al., 2007; Rodríguez-Díaz y Blanco-Villegas, 2011) o Sierra de Gredos en la España Central (Fuster y Colantonio, 2006), revelan paralelamente una sobrevaloración del *inbreeding* estimado mediante isonimia (F_t) respecto al basado en pedigrees (F_{ped}).

En un marco territorial y poblacional más amplio, Rodríguez-Larralde et al. (2003) analizaron la distribución de apellidos en 283 ciudades españolas con datos tomados de los registros de 6328 usuarios telefónicos. La mayor estima del coeficiente de *inbreeding* basado en isonimia (F_t) corresponde a Asturias (0.01246), seguida por la región de Castilla y León (0.00899) y Galicia (0.00744). Los valores más bajos corresponden a Cataluña (0.00462) y Navarra (0.00394). Este parámetro muestra una correlación negativa con el tamaño poblacional, ya que cuanto más reducida sea la población, mayor sería su *inbreeding* y también se verían potenciados los efectos de la deriva genética.

La evaluación del nivel de homocigosis (autozigosis) mediante el genotipado masivo de marcadores bialélicos (SNPs). La detección de segmentos cromosómicos idénticos (ROHs) a lo largo del genoma humano

En la actualidad, el enfoque más novedoso para evaluar y caracterizar la autozigosis en un individuo se basa en la detección de segmentos cromosómicos idénticos por descendencia, los cuales se distribuyen a lo largo del genoma en “carreras” o “tracts” y que se denominan *Runs of Homozygosity (ROHs)*, ver Figura 4) (Gazal et al., 2014a). Esos bloques genómicos sur-

gen cuando dos copias de un haplotipo ancestral coexisten en un individuo. Dichos haplotipos serán más largos si se heredan a partir de antecesores comunes recientes y, más cortos, si los antecesores son más lejanos en el tiempo. Estudios empíricos muestran que los *ROHs* son importantes reservorios de variación homocigótica deletérea (Ceballos et al., 2018).

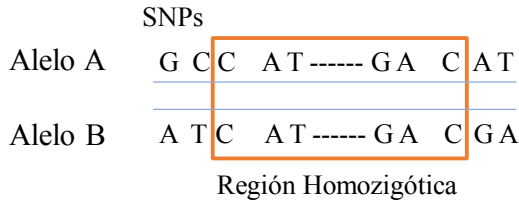


Figura 4: Un ejemplo de arquitectura molecular de un segmento cromosómico idéntico (*ROHs*) definido por secuencias de SNPs contiguos que son *IBDs*.

La aproximación molecular al análisis de la consanguinidad humana se fundamenta en genotipados masivos de alta resolución a nivel genómico (*GW analysis*), lo que supone una fuente de información clave para obtener conocimientos refinados, no solo acerca de la historia evolutiva humana reciente sino también sobre etiologías de enfermedades genéticas complejas.

Del mismo modo que en las metodologías consideradas tradicionales (análisis de árboles familiares e isonimia), para realizar una medida de la autoziosis humana a nivel genómico se utiliza el estadístico *F*. El parámetro *F_{roh}* refleja aquellos segmentos cromosómicos (*tracts*) que presentan estados de homocigosis (*ROHs*) y, además, son *IBDs*:

$$F_{roh} = \sum \frac{L_{roh}}{L_{auto}}$$

Siendo *L_{roh}* la longitud total de todos los *ROHs* presentes en el genoma de una persona y *L_{auto}* la longitud del genoma autosómico cubierto por SNPs, excluyendo los centrómeros. Por tanto, estamos hablando de la proporción del genoma autosómico compuesto por *ROHs* y que, por tanto, sería autozigoto. En la práctica, para el cálculo de ese índice se implementan

algoritmos que escanean cada cromosoma en toda su longitud, mediante ventanas de un tamaño fijo en pares de bases, en la búsqueda de segmentos de SNPs homocigotos consecutivos. Un fragmento cromosómico será considerado como un *ROH* si el número de SNPs consecutivos excede un umbral prediseñado. Este límite inferior, en cuanto a tamaño, va a permitir filtrar aquellos fragmentos que se deben a fenómenos de desequilibrio de ligamiento (*LD*) (Ceballos et al., 2018). El tratamiento estadístico se lleva a la práctica a partir de diversas herramientas informáticas, siendo los más frecuentemente utilizadas PLINK (Purcell et al., 2007) y *FSuite* (Gazal et al., 2014b).

La información adicional extraída a partir de esos programas (*software*) se centra en la estructura de estos segmentos, en el número de *ROHs* que se detectan en el genoma de un determinado individuo y en el tamaño de cada uno de ellos (tanto en pares de bases como en el número de SNPs que contienen). Estos datos se pueden procesar con el objetivo de calcular la suma total de regiones *ROHs* por individuo y el tamaño medio de *ROHs*. Los parámetros que definen a los *ROHs* permiten establecer sólidas inferencias acerca de las historias demográficas de las poblaciones humanas [ver Figura 1 en Ceballos et al. (2018)].

Los *ROHs* no se distribuyen de manera uniforme a lo largo del genoma y, además, la variación de *F* dentro de cada cromosoma puede ser muy grande. El hecho de que el grado de autoziosis pueda variar entre cromosomas e individuos, supone que esos coeficientes *F* podrían ser dependientes de la dinámica poblacional (p.e. consanguinidad o mestizaje) y/o de los efectos de la selección natural (Schraiber et al., 2012). Sería interesante incluso enfatizar que los *ROHs* son más comunes en regiones con alto desequilibrio de ligamiento *LD* y baja recombinación, siendo, además, particularmente frecuentes en el cromosoma X y en regiones de baja diversidad genética. De manera más concreta, existen en cada cromosoma regiones que muestran exceso de *ROHs* (*ROH islands*) junto con otras que muestran defecto de ellos (*coldspots*) (Nothnagel et al., 2010; Ceballos et al., 2018). Se han descrito unas 77 islas *ROH* en poblaciones de Europa, y entre el pool de genes que contienen, destaca el gen de la lactasa (*LCT*) en el cromosoma 2, el cual ha estado sometido a una elevada presión selectiva en el continente europeo (Beja-Pereira et al., 2003). El hecho de

que estos puntos calientes de *ROHs* alberguen genes esenciales para el organismo humano se asociaría con fenómenos de selección positiva, lo que genera nuevas perspectivas a la hora de estudiar el carácter evolutivo de los *ROHs*.

Historias demográficas, consanguinidad y enfermedades. Endogamia, exogamia e inbreeding genómico (Froh)

Como se ha comentado anteriormente, el número total de *ROHs*, sus longitudes y la suma total de *ROHs* presentes en el genoma de las poblaciones humanas contemporáneas constituyen un buen registro de sus dinámicas demográficas y una rica base interpretativa para desvelar razones causales de enfermedades. Pemberton et al. (2012) establecieron categorías para los *ROHs*: cortos (clase A), intermedios (clase B) y largos (clase C), relacionando el tamaño del fragmento con el número de generaciones transcurridas desde el antepasado común. Cuanto más cercanos en el tiempo se encuentren estos antecesores comunes (p.e. un matrimonio M22), mayor tamaño registrarán estos segmentos de autozigosidad. Por el contrario, carreras (*tracts*) de homocigosidad más cortas (clase A) serían el resultado de fenómenos sucesivos de recombinación a lo largo de las generaciones que han permitido la fragmentación de dichos segmentos (Kirin et al., 2010).

Las nuevas herramientas bioinformáticas están permitiendo estimar, no solo los niveles de *ROHs* a nivel individual en personas descendientes de cruzamientos consanguíneos y, originarias de poblaciones de tamaño demográfico relativamente pequeño, sino también en aquellas parejas “exógamas” seleccionadas aleatoriamente y procedentes de poblaciones demográficamente grandes. Varios trabajos han mostrado cómo *ROHs* de longitudes más cortas aparecen en prácticamente todas las poblaciones, mientras que aquellos de longitudes más largas, suelen detectarse en poblaciones sometidas a un elevado aislamiento geográfico (Kirin et al., 2010; Nothnagel et al., 2010; Pemberton et al., 2012). Ejemplos de estos escenarios se encuentran en el trabajo de McQuillan et al. (2008). Los autores revelaron que en las poblaciones nativas objeto de estudio (Islas Orcadas en Escocia, y la Isla Dálmeta en Croacia) se registró un elevado porcentaje de *ROHs* de

todas las longitudes, destacando aquellos ≥ 2.5 Mb, como un reflejo de la fuerte endogamia que caracteriza a estos grupos insulares. La endogamia es una variable demográfica de alto interés para la Antropología Social, así como para la Antropología Genética y se refiere a la proporción de cruzamientos en los que los dos miembros de la pareja son originarios de la misma localidad (p.e. Municipio, Parroquia). En la medida en que la endogamia puede ser examinada con carácter transversal, permite al investigador adoptar otras escalas geográficas o incluso el criterio de pertenencia o no a la misma población.

Un exhaustivo estudio publicado recientemente (Karafet et al., 2015) nos permite una muy buena aproximación al conocimiento de los niveles de *ROHs* actuales en poblaciones de Eurasia. Entre el conjunto de muestras poblacionales analizadas, aquellas que presentaban los niveles más altos de *ROHs* fueron las de Daguestán (Rusia) y Líbano (ver Figura 5). Estos resultados moleculares parecen asociarse a factores lingüísticos (como barreras al flujo génico) o a estructuras consanguíneas muy estrechas. Este último escenario sería el caso del Líbano, donde el $\approx 25\%$ de los matrimonios presentan una relación de parentesco de primos hermanos (Khat, 1989). Estos datos y resultados podrían explicar la prevalencia de enfermedades genéticas raras en dicha población (Jalkh et al., 2015).

Desde la emergencia de *Homo sapiens*, la dinámica de las poblaciones modernas se ve reflejada en los tipos de *ROHs* y sus longitudes. La Figura 6 ilustra las características de este tipo de segmentos cromosómicos en grandes áreas continentales. Kirin et al. (2010) evaluando el coeficiente de *inbreeding Froh*, observaron cómo los niveles más elevados del *F* asociado a *ROHs* > 5 Mb (*Froh* 5) se detectan en poblaciones nativas americanas, lo que indicaría una gran prevalencia de cruzamientos entre parientes cercanos y un reducido tamaño efectivo poblacional, *Ne*. En lo que concierne al continente europeo y sus poblaciones se ha observado que, tanto la media del número de *ROHs* por individuo como la media de la longitud de *ROHs* acumulada, presentan gradientes (*clines*) positivos de sur a norte. Estos parámetros muestran elevadas correlaciones con la latitud (más intensa en el primer caso, $r = 0.84$ y $r = 0.61$, respectivamente) (Nothnagel et al., 2010). Los datos son consistentes con un mayor *Ne* en el sur de Europa, como reflejo de las migracio-

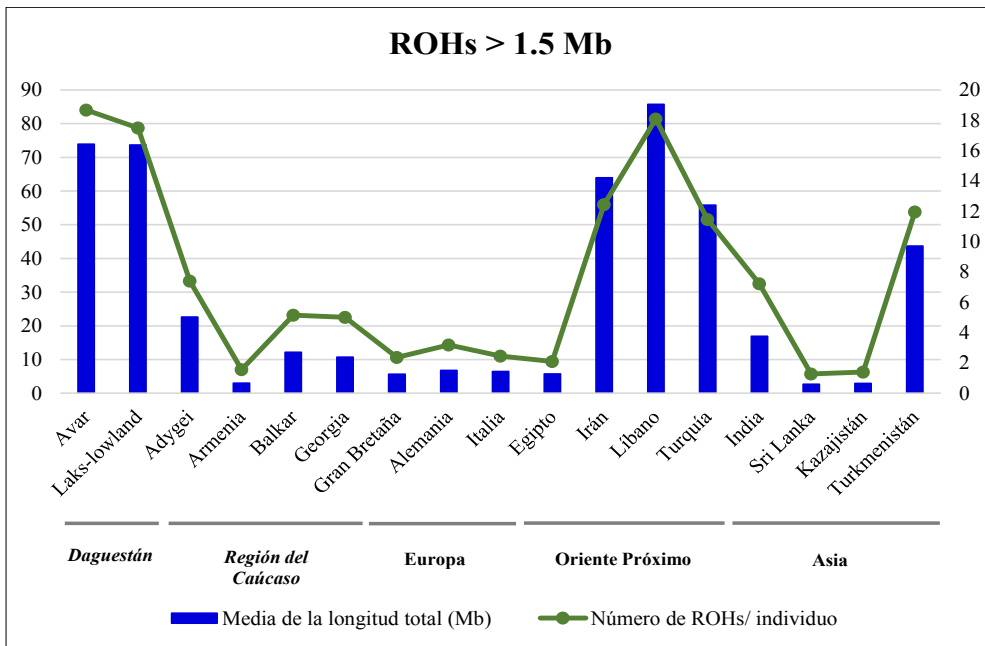


Figura 5: Media de la longitud total de ROH (Mb) y número total de ROHs por individuo de tamaño mayor a 1.5 Mb en distintas poblaciones de Eurasia. Información tomada de Karafet et al. (2015).

nes humanas ocurridas en el continente desde su colonización, lo que habría dado lugar a una disminución de la diversidad genética a medida que se incrementa la latitud.

En este contexto, es interesante resaltar el caso de la Península Ibérica, donde se registran valores que no son demasiados coincidentes con la distribución clinal observada para el resto del continente. El nivel elevado del número de ROHs podría ser debido a una tendencia hacia el matrimonio consanguíneo en muchas regiones de España, escenario conducente a un aumento sensible del nivel de *inbreeding* de la población general (Calderón et al., 2009; Fuster y Colantoni, 2003).



Figura 6: Características de los ROHs en las poblaciones humanas de grandes áreas geográficas en función de sus tamaños e implicaciones evolutivas.

Adaptada de <http://pembertonlab.med.umanitoba.ca/research.html>.

Las estimas tradicionales del coeficiente de *inbreeding* frente a las herramientas genómicas

Sobre la base de todas las evidencias presentadas en este trabajo, resulta de especial interés contextualizar las estimas del coeficiente de *inbreeding* basadas en datos genómicos con resultados previos basados en metodologías tradicionales. Una de las grandes ventajas de los análisis de segmentos cromosómicos es que permite hacer distinción entre *ROHs* de diferente tipo en función de su longitud. Como se ha podido ver anteriormente, las diferentes longitudes de *ROHs* nos indicarían orígenes de muy distinta naturaleza, lo que proporciona valiosos datos acerca de la historia demográfica humana en relación con la consanguinidad.

Recientes y consistentes investigaciones han mostrado altas y positivas correlaciones entre *Fped* y *Froh*. Por tanto, ambas fuentes de información parecen constituir medidas fiables del *inbreeding* registrado en una población. Con estos objetivos, McQuillan et al. (2008) haciendo uso de esas dos metodologías, estimaron el coeficiente de *inbreeding* en diferentes subpoblaciones de las Islas Orcadas (Escocia, Reino Unido) (ver Tabla 2).

A pesar de la existencia de diferencias entre ambas valoraciones, el análisis de los resultados mostró una alta correlación entre *Fped* y *Froh* ($r > 0.70$ para las tres categorías de *ROHs* en base a la longitud). Sin embargo, este estudio se centró en una población de tamaño demográfico pequeño y elevado aislamiento geográfico. Ambas características presentan un escenario idóneo para el estudio de su estructura marital

(*Fped*). Sin embargo, no todas las poblaciones cumplen esta condición. La realización de este tipo de trabajos en poblaciones más abiertas y con tamaños poblacionales más elevados permitiría contrastar de manera precisa la existencia de esa alta correlación entre ambas estimas del *inbreeding*.

Discusión

En la actualidad, el registro de matrimonios entre parientes biológicos en las sociedades económicamente avanzadas es muy bajo en contraposición a aquellas más tradicionales (Ahmad et al., 2016). Este hecho ha justificado un alto interés por seguir investigando en aquellas poblaciones y territorios donde la prevalencia de matrimonios consanguíneos continúa siendo elevada. Ejemplos de estos escenarios los encontramos en el Norte de África, Oriente Medio y Asia occidental (Hamamy et al., 2011). Los datos biodemográficos han sido los utilizados más frecuentemente para estudiar los patrones y niveles de *inbreeding* en las poblaciones humanas y sus variaciones temporales. La obtención de esta información es muy costosa en términos de tiempo, recursos y esfuerzo, por lo que el análisis de la consanguinidad humana resulta más asequible en las poblaciones demográficamente pequeñas, calificadas como “aislados humanos”.

Dentro de las fuentes biodemográficas, los apellidos han sido frecuentemente utilizados como marcadores genéticos. Uno de los factores limitantes del método de la isonimia como una aproximación al aná-

Tabla 2: Valores medios de *Fped* y *Froh* estimados en subpoblaciones de las Islas Orcadas (Escocia, Reino Unido). Información tomada de McQuillan et al. (2008)

| Subpoblaciones Islas Orcadas | N | Media (SE) <i>Fped</i> | Equivalente al grado de relación entre primos | Media (SE) <i>Froh 0.5</i> | Media (SE) <i>Froh 1.5</i> | Media (SE) <i>Froh 5</i> |
|------------------------------|---------|---------------------------|---|-------------------------------|-------------------------------|-----------------------------|
| Descendencia de M22 o M33 | 42 | 0.01820 (0.00140) | M33 | 0.05690 (0.00240) | 0.02710 (0.00220) | 0.01690 (0.00170) |
| Orcadianos endógamos | 11 4 | 0.00150 (0.0004) | M44 y M55 | 0.03790 (0.00080) | 0.00870 (0.00070) | 0.00300 (0.00040) |
| Medios orcadianos | 49 | 0 | Ninguno | 0.03150 (0.00040) | 0.00210 (0.00020) | 0.00010 (0.00007) |
| Total | 24 9 | 0.00380 (0.00050) | M44 | 0.03900 (0.00080) | 0.00980 (0.00070) | 0.00450 (0.00050) |

lisis de la consanguinidad humana, tal y como hemos podido ver, estaría en que los cálculos del coeficiente de *inbreeding* total (F_t) se encuentran sobreestimados (10-15%, como promedio) en relación a los que emergen desde el análisis de pedigrías (F_{ped}). Este hecho se debe al incumplimiento de las premisas sobre las cuales se fundamenta el método basado en los apellidos (p.e. su carácter polifilético) (Esparza et al., 2015). En esta dirección, los estudios que intentan relacionar apellidos con haplogrupos del C-Y han demostrado una asincronía entre ambos marcadores patrilineales (Calderón et al., 2015), una evidencia que nos lleva a plantear una serie de cuestiones acerca del método de la isonimia para estudiar el *inbreeding* en nuestra especie.

Sin duda, el análisis de la estructura marital consanguínea basada en árboles familiares o pedigrías ha sido la metodología más habitual para la estimación de las relaciones de parentesco biológico en poblaciones de muy diferentes orígenes geográficos. Sin embargo, una de las limitaciones intrínsecas a las estimas del *inbreeding* basadas en pedigrías, a partir de las Dispensas Eclesiásticas, es que, en general, no se encuentra reflejado el posible nivel de *inbreeding* que pudiese albergar alguno de los progenitores, así como el posible coeficiente de parentesco (ϕ) entre ellos (Speed y Balding, 2015). Dado que los estudios basados en isonimia no necesitan cumplir este requisito, es obvio esperar su relativa abundancia en los países nortatólicos, hecho que se observa ampliamente en las publicaciones científicas.

McCullough y O'Rourke (1986) detectaron una correlación negativa entre el F_{ped} y la latitud en Europa, conclusión que se contrapone al escenario observado por Nothnagel et al. (2010) a través del análisis del *inbreeding* genómico. Este último trabajo mostró un gradiente positivo del sur al norte del continente europeo, con algunas excepciones como el caso de la Península Ibérica. Dichos resultados parecen correlacionarse con las corrientes migratorias, en dirección sur-norte, que tuvieron lugar durante el proceso del repoblamiento de Europa después del Último Máximo Glacial. Las inferencias basadas en el enfoque molecular podrían considerarse como más fiables, posiblemente debido al elevado alcance de los *ROHs* que permitirían retrotraerse muchas generaciones atrás.

Los avances metodológicos en la estimación de las relaciones de parentesco intrapoblacionales ha desembocado en la detección de una alta correlación positiva entre F_{ped} y F_{roh} (McQuillan et al., 2008). El parámetro F_{ped} disminuye en aquellas subpoblaciones que registran un menor grado de endogamia, llegando a alcanzar el valor de cero en otras subpoblaciones con parientes exógamos (ver Tabla 2). Por el contrario, la subpoblación con mayor nivel de endogamia estudiada por ese equipo (Orcadianos endógamos) alberga los niveles más elevados de todos los parámetros analizados. El resto de poblaciones se distingue por tener niveles de F_{roh} elevados en comparación con los que resultan a partir de los pedigrías, F_{ped} . Estos resultados son de nuevo indicativos del poder de resolución del *inbreeding* genómico. La correlación entre F_{ped} y el F_{roh} , es sensiblemente más acusada para longitudes de *ROHs* superiores a 1.5 Mb. Futuros estudios en este ámbito y focalizados en poblaciones abiertas y con tamaño demográfico elevado podrán mostrar si estas correlaciones se manifiestan no solo en los aislados poblacionales.

Existen una serie de ventajas adicionales y altamente convenientes relacionadas con los análisis de los *ROHs* y que no aportan las otras metodologías consideradas y analizadas en el presente estudio. En primer lugar, los programas utilizados han resultado ser herramientas bioinformáticas esenciales que permiten desarrollar grandes análisis en datos genómicos masivos. Otro aspecto innovador a tener en cuenta es que los *ROHs* pueden detectar *inbreeding* no solo en poblaciones demográficamente pequeñas, aisladas y consanguíneas como, en general, lo hacen el resto de métodos, sino también en poblaciones más cosmopolitas con tamaño poblacional elevado. En consecuencia, cualquier individuo elegido al azar y procedente de cualquier población es susceptible de ser estudiado. Además, los distintos tipos de *ROHs* pueden ser analizados en función de su longitud (Pemberton et al., 2012), lo que nos ofrece información acerca de la existencia de fenómenos de *inbreeding* reciente o lejano, siendo también indicadores del número de generaciones que han transcurrido desde el surgimiento del *Homo sapiens* en África y de su dinámica poblacional (Kirin et al., 2010).

Existen millones de individuos que aparentemente comparten el mismo árbol familiar o pedigrí

pero se distinguen en cuanto a la distribución de los segmentos homocigotos (autocigotos) en sus genomas, debido al proceso aleatorio de segregación mendeliana (Wang, 2016). Por tanto, ambas fuentes de información deben conjugarse para evaluar su impacto en el marco de la Genética Médica y de la Biomedicina. Los coeficientes de *inbreeding* basados en pedigríes, *F_{ped}* especificarían la distribución esperada de homocigosis en el genoma, mientras que los *ROHs* nos muestran su distribución real (Speed y Balding, 2015). La cartografía de la autocigosis (*ROHs*) y los estudios de asociación entre fenotipos y genes (*GWAs*, *Genome Wide Analysis*) permiten rastrear aquellas regiones genómicas implicadas en el desarrollo de enfermedades de etiología compleja, un campo de gran dinamismo hoy entre la comunidad científica.

El potencial de detección de la autocigosis en el genoma por parte de los *ROHs* está abriendo nuevas perspectivas en el estudio del *inbreeding* tanto a nivel individual como poblacional. La comparación del número, longitud y suma de los *ROHs* en las poblaciones humanas es una representación útil de la historia demográfica de los antepasados de un individuo y de la dinámica de la población a la que pertenece. Por tanto, se puede afirmar hoy que la medida de la homocigosis humana basada en el método de estudio de los *ROHs* (*F_{roh}*) está aportando una gran cantidad de información complementaria de primer orden que solo con las metodologías tradicionales, fundamentadas en los registros históricos, no hubiera sido posible conseguir.

Conclusiones

Las estimas del coeficiente de *inbreeding* basadas en árboles familiares resultan más sólidas que las que emergen del análisis de la isonimia. Por otro lado, no se puede afirmar con rotundidad hasta el momento una clara correlación entre *F_{ped}* y *F_{roh}*, debido a la escasez de estudios basados en pedigríes en aquellas poblaciones más cosmopolitas. Futuros trabajos podrán aportar más datos sobre esa correspondencia. El *inbreeding* genómico (*F_{roh}*) ofrece una firme y clara estimación del *inbreeding* y aporta una rica información a distintos niveles, lo que lo convierte hoy en una herramienta indispensable para el conocimiento del fenómeno de la consanguinidad humana.

Agradecimientos

Este trabajo ha sido financiado por el Ministerio de Economía y Competitividad (MINECO) a través del proyecto de investigación CGL2014-53985-R concedido a RC como Investigador Principal.

Referencias bibliográficas

- Ahmad B., Rehman A.U., Malik S. (2016) Consanguinity and inbreeding coefficient in tribal Pashtuns inhabiting the turbulent and war-affected territory of Bajaur Agency, North-West Pakistan. *J Biosoc Sci* 48:113–128.
- Alfonso-Sánchez M.A., Peña J.A., Calderón R. (2003) Time Trends and Determinants of Completed Family Size in a Rural Community from the Basque Area of Spain (1800-1969). *J Biosoc Sci* 35:481–497.
- Alfonso-Sánchez M.A., Aresti U., Peña J.A., Calderón R. (2005) Inbreeding levels and consanguinity structure in the Basque province of Guipúzcoa (1862-1980). *Am J Phys Anthropol* 127:240–252.
- Beja-Pereira A., Luikart G., England P.R., Bradley D.G., Jann O.C., Bertorelle G., Chamberlain A.T., Nunes T.P., Metodieiev S., Ferrand N., Erhardt G. (2003) Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nat Genet* 35:311–313.
- Boattini A., Blanco Villegas M.J., Pettener D. (2007) Genetic structure of La Cabrera, Spain, from surnames and migration matrices. *Hum Biol* 79:649–666.
- Calderón R., Peña J.A., Morales B., Guevara J.I. (1993) Inbreeding Patterns in the Basque Country (Alava Province, 1831-1980). *Hum Biol* 65:743–770.
- Calderón R., Peña J.A., Delgado J., Morales B. (1998) Multiple kinship in two Spanish regions: New model relating multiple and simple consanguinity. *Hum Biol* 70:535–561.
- Calderón R., Aresti U., Ambrosio B., González-Martín A. (2009) Inbreeding coefficients for X-linked and autosomal genes in consanguineous marriages in Spanish populations: the case of Guipúzcoa (Basque Country). *Ann Hum Genet* 73:184–95.
- Calderón R., Hernández C.L., Cuesta P., Dugoujon J.-M. (2015) Surnames and Y-chromosomal markers reveal low relationships in Southern Spain. *PLoS One* 10:e0123098.
- Cavalli-Sforza L.L., Bodmer W.F. (1981) *Genética de las poblaciones humanas*. Barcelona: Ediciones Omega.
- Cavalli-Sforza L.L., Moroni A., Zei G. (2004) *Consanguinity, Inbreeding and Genetic Drift in Italy*. Princeton University Press.
- Ceballos F.C., Joshi P.K., Clark D.W., Ramsay M., Wilson J.F. (2018) Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet* 19:220–234.
- Crow J.F., Mange A.P. (1965) Measurement of inbreeding from the frequency of marriages between persons of the same surname. *Eugen Q* 12:199–203.

- Esparza M., Martínez-Abadías N., Sjøvold T., González-José R., Hernández M. (2015) Comparison between Inbreeding Analyses Methodologies. *Coll Antropol* 39:843–846.
- Fuster V., Colantonio S.E. (2003) Inbreeding coefficients and degree of consanguineous marriages in Spain: A review. *Am J Hum Biol* 15:709–716.
- Fuster V., Colantonio S.E. (2006) Isonymic Analysis of Population Structure. *Coll Antropol* 30:199–203.
- Gazal S., Sahbatou M., Perdry H., Letort S., Génin E., Leutenegger A.L. (2014a) Inbreeding coefficient estimation with dense SNP data: Comparison of strategies and application to HapMap III. *Hum Hered* 77:49–62.
- Gazal S., Sahbatou M., Babron M.C., Génin E., Leutenegger A.L. (2014b) FSuite: Exploiting inbreeding in dense SNP chip and exome data. *Bioinformatics* 30:1940–1941.
- Gomez P.G. (1989) Consanguinity: Geographical variation and temporal evolution in the North of the Iberian peninsula, 1918–1968 (León, Spain). *Int J Anthropol* 4:119–124.
- Hamamy H., Antonarakis S.E., Cavalli-Sforza L.L., Temtamy S., Romeo G., Kate L.P., Bennett R.L., Shaw A., Megarbane A., van Duijn C., Bathija H., Fokstuen S., Engel E., Zlotogora J., Dermitzakis E., Bottani A., Dahoun S., Morris M.A., Arsenault S., Aglan M.S., Ajaz M., Alkalamchi A., Alnaqeb D., Alwasiyah M.K., Anwer N., Awwad R., Bonnefin M., Corry P., Gwanmesia L., Karbani G.A., Mostafavi M., Pippucci T., Ranza-Boscardin E., Reversade B., Sharif S.M., Teeuw M.E., Bittles A.H. (2011). Consanguineous marriages, pearls and perils: Geneva International Consanguinity Workshop Report. *Genet Med* 13:841–847.
- Hartl D., Clark A.G. (2007) Principles of population genetics. Sudbury, Massachusetts: Sinauer Associates.
- Hedrick P.W. (2005) Genetics of populations. Sudbury, Massachusetts: Jones and Bartlett publishers.
- Jalkh N., Sahbatou M., Chouery E., Megarbane A., Leutenegger A.L., Serre J.L. (2015) Genome-wide inbreeding estimation within Lebanese communities using SNP arrays. *Eur J Hum Genet* 23:1364–1369.
- Karafet T.M., Bulayeva K.B., Bulayev O.A., Gurganova F., Omarova J., Yepiskoposyan L., Savina O.V., Veeramah K.R., Hammer M.F. (2015) Extensive genome-wide autozygosity in the population isolates of Daghestan. *Eur J Hum Genet* 23:1405–1412.
- Khlat M. (1989) Inbreeding effects on fetal growth in Beirut, Lebanon. *Am J Phys Anthropol* 80:481–484.
- King T.E., Jobling M.A. (2009) What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. *Trends Genet* 25:351–360.
- Kirin M., McQuillan R., Franklin C.S., Campbell H., Mckeigue P.M., Wilson J.F. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5:e13996.
- Larmuseau M.H.D., Vanoverbeke J., Gielis G., Vanderheyden N., Larmuseau H.F.M., Decorte R. (2012) In the name of the migrant father—analysis of surname origins identifies genetic admixture events undetectable from genealogical records. *Heredit (Edinb)* 109:90–95.
- Lasker G.W. (1985) Surnames and Genetic Structure. Cambridge: Cambridge University Press.
- Lasker G.W., Kaplan B.A. (1985) Surnames and Genetic Structure: Repetition of the same Pairs of Names of Married Couples, a Measure of Subdivision of the Population. *Hum Biol* 57:431–440.
- Malécot G., Blaringhem L. (1948) Les mathématiques de l'hérédité. Paris: Masson et Cie.
- Martinez-Cadenas C., Blanco-Verea A., Hernando B., Busby G.B., Brion M., Carracedo A., Salas A., Capelli C. (2016) The relationship between surname frequency and Y chromosome variation in Spain. *Eur J Hum Genet* 24:120–128.
- McCullough J.M., O'Rourke D.H. (1986) Geographic distribution of consanguinity in Europe. *Ann Hum Biol* 13:359–367.
- McQuillan R., Leutenegger A.-L., Abdel-Rahman R., Franklin C.S., Pericic M., Barac-Lauc L., Smolej-Narancic N., Janicijevic B., Polasek O., Tenesa A., Macleod A.K., Farrington S.M., Rudan P., Hayward C., Vitart V., Rudan I., Wild S.H., Dunlop M.G., Wright A.F., Campbell H., Wilson J.F. (2008) Runs of homozygosity in European populations. *Am J Hum Genet* 83:359–372.
- Morales B. (1992) Estructura de la consanguinidad en la Diócesis de Sigüenza-Guadalajara (1855–1980). Variación histórica, microgeográfica y genealógica. Tesis Doctoral. Universidad del País Vasco, Bilbao.
- Nothnagel M., Lu T.T., Kayser M., Krawczak M. (2010) Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Hum Mol Genet* 19:2927–2935.
- Núñez A.M. (2015) Un siglo de matrimonios consanguíneos en la Archidiócesis de Granada (1900-1999). Un análisis antropológico. Tesis Doctoral. Universidad de Granada.
- Pemberton T.J., Absher D., Feldman M.W., Myers R.M., Rosenberg N.A., Li J.Z. (2012) Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet* 91:275–292.
- Pinto-Cisternas J., Pineda L., Barraí I. (1985) Estimation of Inbreeding by Isonymy in Iberoamerican Populations: An Extension of the Method of Crow and Mange. *Am J Hum Genet* 37:373–385.
- Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J., Sham P.C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* 81:559–575
- Rodríguez-Díaz R., Blanco-Villegas M.J. (2011) Consanguinity in Two Spanish regions: La Cabrera and Fuentes Carrionas. Dispersions and Isonymy. *Coll Antropol* 35:1251–1257.
- Rodríguez-Laralde A., Gonzales-Martin A., Scapoli C., Barraí I. (2003) The names of Spain: A study of the isonymy structure of Spain. *Am J Phys Anthropol* 121:280–292.
- Schraiber J.G., Shih S., Slatkin M. (2012) Genomic tests of variation in inbreeding among individuals and among chromosomes. *Genetics* 192:1477–1482.
- Speed D., Balding D.J. (2015) Relatedness in the post-genomic era: is it still useful? *Nat Rev Genet* 16:33–44.
- Sykes B., Irvén C. (2000) Surnames and the Y chromosome. *Am J Hum Genet* 66:1417–1419.
- Varela T.A., Lodeiro R., Fariña J. (1997) Evolution of consanguinity in the Archbishopsric of Santiago de Compostela (Spain) during 1900–1979. *Hum Biol* 69:517–531.

- Varela T.A., Aínsua R.L., Fariña J. (2003) Consanguinity in the Bishopric of Ourense (Galicia, Spain) from 1900 to 1979. *Ann Hum Biol* 30:419–433.
- Wang J. (2016) Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient? *Theor Popul Biol* 107:4–13.
- Wright S. (1951) The Genetical Structure of Populations. *Ann Eugen* 15:322–354.