

Aplicación de los apellidos al estudio de la estructura genética de las poblaciones humanas

VIVCENTE FUSTER SIEBERT

Rev. Esp. Antrop. Fís. (2010) **31**: 149-156

Dpto. Zoología y Antropología Física, Facultad de Biología, Universidad Complutense de Madrid.

Palabras clave: apellidos, isonimia, estructura población, revisión

Esta revisión se refiere a las posibles aplicaciones que pueden derivarse del conocimiento de la distribución de los apellidos en las poblaciones humanas, en relación con el análisis de su estructura genética. Se discuten los fundamentos del método de la isonimia y sus aspectos más controvertidos (el polifiletismo siendo el principal). Se presentan algunos estudios en los que se ha pretendido validarlo, así como varias de las aproximaciones de análisis que han tenido mayor seguimiento en torno a las relaciones de parentesco entre grupos y su grado de concordancia, en comparación con estimaciones derivadas de otras fuentes de información. Se concluye que, a pesar de las debilidades del método por incumplimiento de algunas de sus premisas, si el interés se dirige a la observación de los patrones de variación relativa del parentesco, los apellidos constituyen un método aceptable de estudio en comunidades históricas para las que no se dispone de información alternativa. También hacen posible, de una manera sencilla, el análisis de grandes poblaciones, pudiendo reflejar la consanguinidad remota, la existencia de cruzamientos preferenciales y la migración en el pasado.

© 2010 Sociedad Española de Antropología Física

Introducción y Fundamento del método

El uso de los apellidos en el estudio de la estructura genética de las poblaciones humanas, se fundamenta en la supuesta transmisión regular de aquellos de padres a hijos. En poblaciones en las que los hijos llevan el apellido del padre, de manera teórica su comportamiento sería equivalente al del cromosoma Y, mientras que en situaciones de transmisión materna, se asimilaría a la herencia del ADNmt. En el caso de que se cumpliera ese principio de regularidad (inexistencia de casos de ilegitimidad, adopción, cambio de inscripción, ausencia de variaciones tipográficas) así como una similar migración masculina y femenina, etc., es posible establecer una relación probabilística entre coincidencia de apellidos y parentesco biológico. Para ello, debería tenerse en cuenta la frecuencia (abundancia relativa) de cada apellido en la población.

Lo expresado anteriormente lleva implícito que el hecho de que dos personas presenten el mismo apellido, se corresponde con una probabilidad incrementada de que ambas descendan de un mismo antepasado común, incluso aunque fuera lejano en el tiempo. Esto sería cierto si cada apellido tuviera un origen monofilético, es decir que hubiera surgido tan solo una vez. En caso de polifiletismo, la identidad de apellidos no reflejará necesariamente una relación de parentesco. Realmente, es este aspecto el más controvertido a la hora de aceptar como válido este tipo de información. Rogers (1991) demostró que son necesarios supuestos más restrictivos para validar el método. Según este autor, cualquier apellido debería haber surgido en la misma generación (“strong monophyletic assumption”). Pero, incluso en poblaciones que descenden de un grupo de fundadores conocido, es improbable que se cumpla ese requisito debido al hecho de que la migración humana tiende a estar estructurada por el parentesco, de manera que el proceso de fundación puede haber implicado grupos de individuos emparentados. La débil correspondencia entre las estimas de “inbreeding” derivadas de isonimia y de genealogías (Jorde, 1989) también puede ser consecuencia del punto de origen de la información. Las genealogías se remontan a la generación

más antigua conocida, en tanto que para los apellidos se desconoce el grupo fundador, el cual, en cualquier caso será más remoto que para la mayoría de las genealogías disponibles. Sin embargo, cuando se conocen las frecuencias de apellidos en dos generaciones diferentes (por ejemplo consecutivas), sería posible obtener estimas más fiables que cuando estas se refieren a un supuesto stock inicial (fundador) desconocido (Rogers (1991).

Diferentes procedimientos de análisis

Inicialmente los apellidos se aplicaron para calcular el coeficiente de consanguinidad (“inbreeding”), partiendo de la proporcionalidad entre éste (F) y la coincidencia de un apellido (I) en cruzamientos entre individuos con diversos grados de relación ($F = I/4$). Siguiendo el algoritmo propuesto por Crow y Mange (1965) es posible descomponer dicho coeficiente en una fracción no aleatoria (F_n) y otra aleatoria (F_r), esta última determinada por la probabilidad de coincidencia por azar del mismo apellido en una pareja, la cual será tanto mayor cuanto más común sea el apellido. Por el contrario, el componente no aleatorio indicaría la desviación del azar (positiva o negativa) hacia el emparejamiento de parientes. El mismo autor perfeccionó su anterior propuesta discutiendo dos métodos que presentan algunas diferencias de cálculo (Crow, 1980). Aunque teóricamente el denominado método A sería pertinente en poblaciones muy consanguíneas, las diferencias que se obtienen de su aplicación son de poca importancia.

La primera referencia a la posibilidad de contabilizar tanto el apellido paterno como el materno en poblaciones en las que rige el sistema hispanoamericano de transmisión de apellidos se debe a Shaw (1960). Pero han sido Pinto-Cisternas et al. (1985) quienes consolidaron el procedimiento a seguir. Este sistema de transmisión de apellidos presenta la ventaja de duplicar el número de apellidos sobre los que basar la estimación del parentesco (Colantonio et al., 2003). Evidencia las relaciones vía matri o patrilocal y, si se utilizan registros maritales, incluye información sobre la generación precedente (parental) en lugar de una sola (Lasker, 1991). Cualquier diferencia en la distribución de los primeros (paterno) y segundos apellidos (materno), responde a variaciones en la estructura genética entre la última y la penúltima generación.

Dada la excepcionalidad de los cruzamientos consanguíneos, la determinación de la “isonimia marital” se restringe, en principio, a comunidades de pequeño tamaño y en cierto grado endógamas. Pero incluso en estas condiciones, la frecuencia de parejas isónimas es muy baja y, en consecuencia, las estimas del coeficiente de consanguinidad que puedan derivarse vendrán afectas por un error estándar elevado. Bajo condiciones diferentes, como ocurriría en grandes poblaciones, la posibilidad de encontrar emparejamientos isónimos está limitada. En estos casos, más informativo puede ser la contabilización de los “pares de apellidos repetidos” RP (Lasker y Kaplan, 1985) que da cuenta de los cruzamientos entre distintos linajes dentro de una población. Considerando el emparejamiento de apellidos (pertenecientes a varones y mujeres), la mayoría de las combinaciones teóricamente posibles no tienen lugar; muchas ocurren una sola vez y más raramente dos o más. El recurrente emparejamiento de algunos pares de apellidos en las parejas, en tanto que se aparta de las frecuencias esperadas por azar (RPr), permite, basándose en las mismas premisas que las aplicadas a la isonimia marital, estimar el grado de subdivisión de una población en subgrupos (efecto Wahlund). Este método puede poner en evidencia la existencia de emparejamientos preferenciales ya sea por clase social, grupo étnico o subdivisión geográfica. Chakraborty (1985) desarrolló un procedimiento que permite calcular el error estándar de RP y poner a prueba estadísticamente la hipótesis nula de que $RP = RPr$. Para comparar estos valores en diferentes poblaciones, el mismo autor propone expresar en forma de porcentaje el exceso de RP sobre lo esperado por azar, es decir $100 (RP - RPr)/RPr$. Cuanto más elevado sea este cociente, mayor será el grado de subdivisión en la población analizada. RP suele mostrar correlaciones más elevadas con variables demográficas que las que se obtienen al considerar los componentes aleatorio (F_r) y no aleatorio (F_n) de la isonimia. Esto permite atribuir una mayor robustez al método de los pares repetidos como descriptor de la estructura marital de una población (Relethford, 1992). Además, es posible apli-

carlo a linajes particulares presentes en una población (Pettener et al., 1997). De la comparación del “inbreeding” calculado por isonimia con los coeficientes determinados a partir de otra información, como son las genealogías, pronto se llegó a la conclusión de que los primeros claramente constituían una sobreestimación, con unos índices de concordancia entre los valores obtenidos por una y otra fuente, sumamente reducidos y variables (Rogers, 1987). Las diferencias fueron atribuidas al polifiletismo que afecta a muchos apellidos (Hussels, 1969; Ellis y Friedl, 1976). Estos mismos autores al igual que Gagnon y Toupance (2002) y Peña (1991) han sugerido diferentes procedimientos para minimizar el efecto indeseable del origen múltiple de ciertos apellidos, el más simple de todos consistente en restringir el análisis a los apellidos menos comunes. Se ha pretendido validar el método comparando haplogrupos del cromosoma Y en muestras aleatorias de individuos que presentaban un determinado apellido, y muestras controles de sujetos con lugares de residencia correspondientes, pero con apellidos distintos (Hill et al., 2000; Jobling, 2001; Sykes e Irvén, 2000; Zei et al., 2003). Sin descartar la existencia de polifiletismo para la mayoría de los apellidos, puede afirmarse que, en términos generales, estos estudios proporcionan fundamentos favorables a la utilización de los apellidos con un margen de fiabilidad aceptable (King et al., 2006; McEvoy y Bradley, 2006; Presciuttini et al., 2006).

La estructura genética de las poblaciones, tal como puede deducirse de la distribución de los apellidos, responde a un fenómeno jerárquico determinado por la pertenencia de los individuos a un hogar, a una vecindad, parroquia, valle o ciudad o estado. A veces, factores lingüísticos (idioma, dialecto) o étnicos, se superponen a los anteriores. Basándose en el mismo fundamento, y aceptando que el número de apellidos en una población puede considerarse un índice de aislamiento, de manera que cuantos menos apellidos existan, más consanguínea y aislada será (Lasker, 1980), ha surgido una segunda y tal vez más interesante aplicación de los apellidos. Esta aplicación consiste en determinar los modelos de variación de las frecuencias de apellidos dentro y entre poblaciones con el fin de obtener medidas de similitud, dirigidas en último término, a establecer la relación (parentesco) entre poblaciones. Si la isonimia estima el “inbreeding” de un individuo, también podrá medir el grado de relación entre sus progenitores. De esta manera, el coeficiente de relación de los padres duplicará el valor del coeficiente de “inbreeding” de su descendiente (Lasker, 1980). El coeficiente de relación puede aplicarse no solamente a parejas, sino a individuos de cualquier par de poblaciones para las que se conozca la frecuencia de los diferentes apellidos que en ellas existen. Siendo Lasker (1977) quien más insistentemente se basó en apellidos para determinar la relación biológica entre poblaciones, Morton (1973) y Morton et al. (1976) son acreedores del mérito de su propuesta inicial. También para este coeficiente ha de admitirse la posibilidad de desviaciones atribuibles a un elevado polifiletismo. Esto podría resultar en el hallazgo de coeficientes de relación entre poblaciones que excedieran de manera importante lo esperado del conocimiento de tasas de migración en el pasado. Este hecho ha sido señalado por Lasker (1978) en uno de sus estudios en Perú, al igual que en otro (Kaplan et al., 1978) donde las afinidades deducidas de la distribución de apellidos entre comunidades alpinas también superaron lo estimado por genealogías. Este comentario es extensivo a lo que sucede con los modelos derivados de las matrices de migración, ya que los patrones deducidos de este tipo de datos tienden a ser más sensibles a los factores geográficos que los apellidos (Relethford, 1986; Gradie et al. 1988), lo que se traduce, en el primer caso, en correlaciones más elevadas con las matrices de distancia geográfica entre localidades o subpoblaciones. Por el contrario, los apellidos presentan más inercia (se remontan más hacia atrás en el tiempo y suelen mostrar correlaciones más débiles). Sin embargo, si el énfasis se pone, más que en su magnitud en términos absolutos, en los valores relativos de los coeficientes de relación por isonimia, estos coeficientes han mostrado reiteradamente una correspondencia aceptable con factores geográficos (distancia, orografía), red viaria y facilidades de transporte. Así, por ejemplo, Fuster (1986) encontró para una comunidad rural durante el siglo XIX, correlaciones significativas entre los coeficientes de relación por isonimia (R_{ij}), el parentesco condicional derivado de las matrices de migración θ_r y la distancia geográfica (d) (R_{ij}/θ_r , $r=0,62$; R_{ij}/d ,

$r = -0,55$). En un exhaustivo estudio sobre los Mormones de Utah, Jorde (1989) pudo relacionar en una serie de cohortes los valores de F_{st} estimados a partir de las matrices de migración, apellidos y genealogías. Para esta población, los modelos temporales establecidos por isonimia mostraron una correspondencia mayor con los basados en genealogías que con los derivados de las matrices de migración. Respecto a la matriz de distancias geográficas entre sub-poblaciones, las tres series de valores (isonimia, genealogías y migración) fueron concordantes, lo que demuestra un importante efecto de aislamiento por distancia.

Si se acepta que la mayoría de los apellidos son selectivamente neutros (generalmente no se asocian a diferencias importantes en fertilidad o mortalidad), podrían ser analizados por procedimientos basados en los que se aplican en la genética relativa a este tipo de alelos (Zei et al., 1983). Según esta equivalencia, la aparición apellidos nuevos en una población se debería a una mutación (alteración de un apellido, la cual podría ser controlada) o más comúnmente, por la llegada de individuos con apellidos novedosos. De esta manera, podrían derivarse estimaciones de la inmigración a partir de la distribución de apellidos en la población que se esté estudiando (Degioanni y Darlu, 2001). El ajuste de la distribución de apellidos al modelo genético de alelos neutros puede probarse mediante regresión lineal entre las frecuencias relativas de los apellidos y el número de apellidos presentes en cada unidad de población considerada (Barrai et al. 1996).

Barrai et al. (1987) constataron que una transformación logarítmica de la distribución de apellidos se ajusta a una línea independientemente del tamaño de la población, lo que les llevó a la conclusión de que los apellidos se distribuyen según el modelo de Karlin y MackGregor (1967), válido para alelos neutros en una población de N individuos haploides, cada uno de ellos portando 1 de k alelos diferentes (Beretta et al., 1993). Siguiendo a Barrai et al. (1987), la diversidad genética basada en la distribución de apellidos puede estimarse mediante el de Fisher el cual es un indicador de la abundancia de apellidos, que permite identificar a partir de la frecuencia de cada uno de los apellidos en una población y de su tamaño (N), zonas con diversidad más elevadas (equivalente al número de alelos presentes en la población), a las que se les atribuye una mayor llegada de individuos aportando más apellidos diferentes. Estos mismos autores (Barrai et al. 1996; 2002) propusieron estimar, aplicando un algoritmo basado en y en N , la migración acumulada, tal como se deduciría de la distribución de apellidos, mediante el v de Karlin y McGregor (1967), a fin de diferenciar la diversidad observada en un período determinado resultante del aporte en generaciones anteriores. La migración reciente (en el momento del estudio) puede conocerse a partir de la proporción de individuos con un apellido "único", es decir con frecuencia = 1 (Rodríguez-Larralde y Barrai, 1998). Además de los anteriores procedimientos, extensamente aplicados, otros autores han optado por opciones diferentes. Así, Raspe y Lasker (1997) midieron la diversidad genética utilizando el índice G_{st} propuesto por Crow (1996). Este índice considera que los apellidos son análogos a marcadores genéticos con un número de alelos muy elevado y que la introducción de nuevos apellidos por inmigración sería comparable a una mutación que generara nuevos alelos. G_{st} mide la probabilidad de que dos apellidos tomados aleatoriamente en cada subpoblación (P_s) coincidan respecto a esa misma probabilidad para el conjunto de sub-poblaciones (población total = P_t), siendo $G_{st} = (P_s - P_t) / 4(1 - P_t)$. De esta manera, puede obtenerse un único valor que informa acerca del grado de subdivisión de la población. Caravello y Tasso (2002) optaron por un índice de aplicación en ecología basado en Bray y Curtis (1957), el cual estima la similitud asumiendo que las comunidades humanas son unidades ecológicas constituidas por los apellidos que portan sus integrantes.

Otro ejemplo de recuperación de modelos propuestos hace décadas y posteriormente aplicados a los estudios de isonimia es el análisis de autocorrelación (Sokal et al., 1992), el cual se fundamenta en Moran (1950). Este procedimiento evidencia la semejanza en las frecuencias de un apellido determinado entre pares de unidades de población situadas a intervalos de distancia arbitrarios. El coeficiente de Moran puede tomar valores comprendidos entre -1 y +1 (autocorrelación

negativa o positiva respectivamente). La representación de la autocorrelación frente a las clases de distancia recibe el nombre de correlograma. Los trabajos de Cabral et al. (2005) y Caravello y Tasso (2007) son buenos ejemplos de la aplicación de este tipo de análisis en poblaciones de naturaleza muy diferente (el archipiélago de las Azores y los Alpes italianos). Estos autores encontraron una mayor similitud dentro de clases de distancia cortas (diferentes en uno u otro estudio) coherente con un proceso de dispersión por movilidad local. De mayor complejidad es el análisis realizado por Sokal et al. (1992) en Inglaterra y Gales a partir de 100 apellidos. Aquí, además de correlogramas de una dimensión en los que la autocorrelación se calcula como una función de la distancia, pero independientemente de la dirección, se muestran también correlogramas direccionales que representan las similitudes como anillos concéntricos divididos en sectores según su orientación geográfica.

Conociendo la distribución de apellidos en cada categoría de población, la cual ha de estar definida por sus coordenadas geográficas, es posible aplicar modelos de análisis geográfico, tal como hicieron Zei et al. (1993) para reconstruir los patrones de migración en Italia. Más recientemente, Manni et al. (2004) utilizaron un procedimiento similar basándose en los algoritmos de Monmonier por medio del programa Barrier. Estos métodos permiten identificar barreras o discontinuidades entre regiones geográficas como reflejo de cambios abruptos en la variable que se esté analizando (frecuencias de apellidos).

En las últimas décadas se han ido acumulando estudios sobre grandes regiones o estados, basados en registros de votantes, censos o listados telefónicos (Barrai et al., 2000; 2001; 2002; Rodríguez-Larralde et al., 2000; 2003). La mayoría de estos trabajos han permitido establecer una buena correspondencia entre la distancia geográfica y el modelo de distribución espacial de los apellidos. Además de los aspectos hasta ahora comentados referidos al flujo de genes (migración revelada por apellidos) y la estructura genética de la población en su conjunto, otros autores se han interesado en analizar el parentesco y mezcla racial o la estructura de grupos o minorías étnicas en relación con los patrones de distribución de los apellidos (Biondi et al., 1990; Cabello et al., 1997; Caravello et al., 1999; Castro de Guerra, 1999; Colantonio et al., 2002; Hernández et al., 2000; Vienna et al., 2001), o el efecto de las fronteras (Bolsen y Lasker, 1996; Román-Busto et al., 2010). Al margen del evidente interés que la isonimia ha suscitado en relación con el análisis de la estructura genética de las poblaciones, un limitado número de estudios se ha dirigido a establecer una posible asociación entre la incidencia de ciertos desórdenes hereditarios y otras enfermedades, con determinados apellidos o linajes (De Braekeleer 1995; Holloway, 1995; Rudan, 2001; Schiliro et al., 1993).

Conclusión

En oposición a los inconvenientes derivados del polifiletismo, así como del incumplimiento de alguno de los otros supuestos antes indicados (por ej. ilegitimidad), los apellidos presentan la ventaja de permitir extender fácilmente el análisis a grandes colectivos, pudiendo reflejar la consanguinidad remota, la existencia de cruzamientos preferenciales o inferencias acerca la migración en el pasado. Aunque la principal aplicación del estudio de los apellidos inicialmente se ha centrado en la isonimia marital de pequeñas poblaciones históricas homogéneas, en las que la utilización de medidas directas de la estructura genética no es posible, en la actualidad predominan los estudios realizados a partir de censos, listados telefónicos, etc., lo que supone un cambio de escala que aporta una nueva perspectiva a la hora de interpretar los patrones de distribución de apellidos.

Bibliografía

- BARRAI I, BARBUJANI G, BERETTA G, MAESTRI I, RUSSO A, FORMICA G, Y PINTO-CISTERNAS J (1987) Surnames in Ferrara: Distribution, isonymy and levels of inbreeding. *Ann. Hum. Biol.* 14:415-423.
- BARRAI I, RODRÍGUEZ-LARRALDE A, MAMOLINI E, Manni F, SCAPOLI C. (2000) Elements of the surname structure of Austria. *Ann. Hum. Biol.* 27:607-622.

- BARRAI I, RODRIGUEZ-LARRALDE A, MAMOLINI E, MANNI F, SCAPOLI C (2001) Isonymy structure of USA population. *Am. J. Phys. Anthropol.* 114:109-123.
- BARRAI I, RODRÍGUEZ-LARRALDE A, MANNI F, SCAPOLI C (2002) Isonymy and isolation by distance in the Netherlands. *Hum. Biol.* 74:263-283.
- BARRAI I, SCAPOLI C, BERETTA M, NESTI C, MAMOLINI E, RODRIGUEZ-LARRALDE A (1996) Isonymy and the genetic structure of Switzerland. I: The distributions of surnames. *Ann. Hum. Biol.* 23:431-455.
- BERETTA M, MAMOLINI E, RAVANI A, VULLO C, SCAPOLI C, BARALE R, RODRIGUEZ-LARRALDE A, BARRAI I (1993) Comparison of structures from frequencies of genes and surnames in the population of Ferrara. *Hum. Biol.* 65:225-235.
- BIONDI G, RASPE P, PERROTTI E, LASKER GW, MASCIÉ-TAYLOR CGN (1990) Relationships estimated by isonymy among the Italo-Greco villages of Southern Italy. *Hum. Biol.* 62: 649-663.
- BOLSEN J, LASKER GW (1996) Relationship of people across an international border based on an isonymy analysis across the German-Danish frontier. *J. bios. Scie.* 28: 177-183.
- BRAY JR, CURTIS JT (1957) An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* 27: 325-349.
- CABELLO PH, HATAGIMA A, LIMA AMD, AZEVEDO ES, KRIEGER H (1997) Algunos aspectos de la estructura genética de una población del nordeste brasileño: parentesco y mezcla racial relacionados con el origen de los apellidos. *Rev. Esp. Antrop. Biol.* 18:7-18.
- CABRAL R, BRANCO CC, COSTA S, CARAVELLO G, TASSO M, PEIXOTO BR, MOTA-VIEIRA L (2005) Geography of surnames in Azores: specificity and spatial distribution analysis. *Am. J. Hum. Biol.* 17:634-645.
- CARAVELLO GU, TASSO M, LUCCHETTI E (1999) Distribution of surnames and identities in the Ladin communities of the Dolomites. *Anthrop. Anz.* 57:303-317.
- CARAVELLO G, TASSO M (2002) Use of surnames for a Demo-Ecological analysis: a study in Southwest Sardinia. *Am. J. Hum. Biol.* 14:391-397.
- CARAVELLO GU, TASSO M (2007) Surnames as alleles: Spatial distribution of surnames in a province of the Italian Alps. *J. biosoc. Scie.* 39:409-419.
- CASTRO DE GUERRA D (1999) Population structure of two black Venezuelan populations studied through their mating structure and other related variables. *Ann. Hum. Biol.* 26:141-150.
- CHAKRABORTY R (1985) A note on the calculation of random RP and its sampling variance. *Hum. Biol.* 57:713-717.
- COLANTONIO S, FUSTER V, MARCELLINO AJ (2002) Interpopulation relationship by isonymy: application to ethnosomal groups and illegitimacy. *Hum. Biol.* 74:871-878.
- COLANTONIO S, LASKER GW, KAPLAN B, FUSTER V (2003) Use of surname models in human population Biology: A review of recent developments. *Hum. Biol.* 75:785-807.
- CROW JF (1980) The estimation of inbreeding from isonymy. *Hum. Biol.* 52:1-14.
- CROW JF (1983) Discussion. *Hum. Biol.* 55:383-397.
- CROW JF (1996) Isonymy: a thirty year retrospective. *Riv. di Anthropol.* 74:25-34.
- CROW JF, MANGE P (1965) Measurement of inbreeding from the frequency of marriages between persons of the same surname. *Eugen. Quart.* 12:199-203.
- DE BRAEKELEER M (1995) Inbreeding, kinship and surnames in hereditary disorders: The experience in Saguenay Lac-Saint-Jean (Quebec). *Coll. Antrop.* 2:289-304.
- DEGIOANNI A, DARLU P (2001) A Bayesian approach to infer geographical origins of migrants through surnames. *Ann. Hum. Biol.* 28:537-545.
- ELLIS WS, FRIELD J (1976) Inbreeding as measured by isonymy and by pedigrees in Kippel, Switzerland. *Soc. Biol.* 23:158-167.
- FUSTER V. (1986) Relationship by isonymy and migration pattern in northwest Spain. *Hum. Biol.* 58:391-406.
- GAGNON A, TOUPANCE B (2002) Testing isonymy with paternal and maternal lineages in the early Quebec population: The impact of polyphyletism and demographic differentials. *Am. J. Phys. Anthropol.* 117:334-341.
- GRADIE M, JORDE L, BOUCHARD G (1988) Genetic structure of Saguenay, 1852-1911: Evidence from migration and isonymy matrices. *Am. J. Phys. Anthropol.* 77:321-333.
- HERNÁNDEZ M, GARCÍA-MORO C, MORAL P, GONZÁLEZ-MARTÍN A (2000) Population evolution in the 20th-Century Easter Island: Endogamy and admixture. *Hum. Biol.* 72:359-377.
- HILL EW, JOBLING MA, BRADLEY DG (2000) Y chromosome variation and Irish origins. *Nature* 404:351-352.
- HOLLOWAY SM (1995) Coefficients of relationship by isonymy among Scottish males with multiple primary cancers. *Disease Markers* 12:205-214.
- HUSSELS I (1969) Genetic structure of Saas, a Swiss isolate. *Hum. Biol.* 41:469-479.
- JOBLING MA (2001) In the name of the father: surnames and genetics. *Trends in Genetics* 17:353-357.
- JORDE LB (1989) Inbreeding in the Utah Mormons: an evaluation of estimates based on pedigrees, isonymy, and migration matrices. *Ann. Hum. Genet.* 53:339-355.
- KAPLAN B, LASKER GW, CHIARELLI B (1978) Community of surnames: A measure of biological interrelations among thirty-one settlements in Upper Val Varaita in the Italian Alps. *Am. J. Phys. Anthropol.* 49:251-256.
- KARLIN S, MCGREGOR J (1967) The number of mutant forms maintained in a population. *En: Proc. 5th Berkeley Symp. Math. Stat. Prob.* 4:415-438.

- KING TE, BALLEREAU SJ, SCHÜRER KE, JOBLING MA (2006) Genetic signatures of coancestry within surnames. *Current Biology* 16:384-388.
- LASKER GW (1977) A coefficient of relationship by isonymy: A method for estimating the genetic relationship between populations. *Hum. Biol.* 49:489-493.
- LASKER GW (1978) Increments through migration to the coefficient of relationship by isonymy between communities. *Hum. Biol.* 50:235-240.
- LASKER GW (1980) Surnames in the study of Human Biology. *Am. Anthropol.* 82:525-538.
- LASKER GW (1991) Revisión: Datos sobre los apellidos hispanoamericanos en los estudios de Biología Humana. *Anales de Antropología, México* 28:107-128.
- LASKER GW, KAPLAN BA (1985) Surnames and genetic structure: Repetition of the same pair of names in married couples, a measure of subdivision of the population. *Hum. Biol.* 55:431-440.
- MCEVOY B, BRADLEY DG. (2006) Y-chromosomes and the extent of patrilineal ancestry in Irish surnames. *Hum Genet.* 119:212-219.
- MANNI F, GUÉRARD E, HEYER E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by "Monmonier's algorithm". *Hum. Biol.* 76:173-190.
- MORAN PAO (1950) Notes on continuous stochastic phenomena. *Biometrika* 37:17-23.
- MORTON NE (1973) Kinship bioassay. En: Genetic structure of populations. N. Morton (ed.) University Press of Hawaii, Honolulu. pp 158-163.
- MORTON NE, SMITH C, HILL R, FRACKIEWICZ A, LAW P, YEE S (1976) Population structure of Barra (Outer Hebrides). *Ann. Hum. Genet.* 39:339-352.
- PEÑA JA (1991) Efectos del polifiletismo sobre los análisis de estructura genética a partir de listas de apellidos. En: Nuevas perspectivas en Antropología. M Botella, SA Jiménez, L Ruiz, Ph du Souich (eds.) Univ. de Granada. Granada. pp 683-691.
- PETTENER D, FIORINI S, TARAZONA-SANTOS E (1997) Within-lineage pair isonymy (Rpw) in a high altitude Quechua community in the Peruvian Central Andes, 1825-1914. *Rev. Esp. Antrop. Biol.* 18:25-37.
- PINTO-CISTERNAS J, PINEDA L, BARRAI I (1985) Estimation of inbreeding by isonymy in Ibero-American populations. An extension of the method of Crow and Mange. *Am. J. Hum. Genet.* 37: 73-385.
- PRESCIUTTINI S, TONI C, EPIRO D, SPINETTI I, MARRONI F, ROCCHI A, DOMENICI R (2006) Y-chromosome haplotypes and male isonymy: Genetic and genealogical study in a small town of Tuscany (Buti, Italy). *International Congress Series, V. 1288*, pp. 225-227.
- RASPE PD, LASKER GW (1997) G_{st}, a measure of gene diversity from surname distributions. Population subdivision in an area of England. *Rev. Esp. Antrop. Biol.* 18:19-24.
- RELETFORD JH (1986) Microdifferentiation in historical Massachusetts: A comparison of migration matrix and isonymy analyses. *Am. J. Phys. Antropol.* 71:365-375.
- RELETFORD JH (1992) Analysis of marital structure in Massachusetts using repeating pairs of surnames. *Hum. Biol.*, 64,1: 25-33.
- RODRIGUEZ-LARRALDE A, BARRAI I (1998) Estudio genético demográfico del Estado Zulia, Venezuela, a través de isonimia. *Acta Científica Venezolana* 49:134-143.
- RODRÍGUEZ-LARRALDE A, MORALES J, BARRAI I (2000) Surname frequency and isonymy structure of Venezuela. *Am. J. Hum. Biol.* 12:352-362.
- RODRIGUEZ-LARRALDE A, GONZALES-MARTIN A, SCAPOLI C, BARRAI I (2003) The names of Spain: a study of the isonymy structure of Spain. *Am. J. Phys. Anthropol.* 121:280-92.
- ROGERS LA (1987) Concordance in isonymy and pedigree measures of inbreeding: The effects of sample composition. *Hum. Biol.* 59:753-767.
- ROGERS AR (1991) Doubts about isonymy. *Hum. Biol.* 63:663-668.
- ROMÁN-BUSTO J, FUSTER V, COLANTONIO SE, ZULUAGA P, BLANCO MJ, GUARDADO-MOREIRA MJ (2010) Mate choice in Olivenza: influence of border change on Spanish-Portuguese lineages. *J. biosoc. Scie.* 42:129-140.
- RUDAN I (2001) Ancestral kinship and cancer in Lastovo Island, Croatia. *Hum Biol.* 73:871-84.
- SCHILIRO G, RODRÍGUEZ-LARRALDE A, MAMOLINI E, SCAPOLI C, BARRAI I (1993) Isonymy in haemoglobinopathies in Sicilian sample. *Hum. Hered.* 43:203-204.
- SHAW RF (1960) An index of consanguinity based on the use of the surname in Spanish-speaking countries. *J. Hered.* 51:221-230.
- SOKAL RR, HARDING RM, LASKER GW, MASCIE-TAYLOR CGN (1992) A spatial analysis of 100 surnames in England and Wales. *Ann. Hum. Biol.* 19:445-476.
- SYKES B, IRVEN C (2000) Surnames and the Y Chromosome. *Am. J. Hum. Genet.* 66:1417-1419.
- VIENNA A, PEÑA GARCIA JA, MASCIE-TAYLOR CGN, BIONDI G (2001) The ethnic minorities of Southern Italy and Sicily: Relationships through surnames. *J.biosoc. Sci.* 33:25-31.
- ZEI G, GUGLIELMINO CR, SIRI E, MORONI A, CAVALLI-SFORZA LL (1983). Surnames as neutral alleles: Observations in Sardinia. *Hum. Biol.* 55:357-365.
- ZEI G, BARBUJANI G, LISA A, FIORANI O, MENOZZI P, SIRI E, CAVALLI-SFORZA LL (1993) Barriers to gene flow estimated by surname distribution in Italy. *Ann. Hum. Genet.* 57:123-140.
- ZEI G, LISA A, FIORANI O, MAGRI C, QUINTANA-MURCI L, SEMINO O, SANTACHIARA-BENNERCETTI AS (2003) From surnames to the history of Y chromosomes: the Sardinian population as a paradigm. *European J. Hum. Genet.* 11:802-807.

Abstract

The application of information on the distribution of surnames to assess the genetic structure of human populations is described. The background of the isonymy method and its most controversial weaknesses (mainly the polyphyletism) are discussed. Examples are presented of studies which tested the validity of isonymy. Different analytical approaches are shown regarding the inter-population kinship relationships from surnames and their concordance with estimates provided by other types of data. In spite its bias from the theoretical assumptions, the isonymy method may be considered acceptable when the analysis draws inferences from the relative variation of kinship, more than from the absolute values, and when alternative data are difficult to obtain. Isonymy makes possible the study of large populations and results may reflect remote consanguinity, the existence of preferential matings and past migration.

Key words: surnames, isonymy, population structure, review

The use of surnames for the study of the genetic structure of human populations