J.biosoc.Sci, (2008) **40**, 239–246, © 2007 Cambridge University Press
doi:10.1017/S0021932007002313 First published online 22 Aug 2007

FACTORS RELATED TO INBREEDING COMPONENTS FROM ISONYMY IN AN URBAN POPULATION: ARANJUEZ (SPAIN)

SONIA E. COLANTONIO*§, VICENTE FUSTER†§, ALBERTO SANZ GIMENO‡§ AND DAVID S. REHER‡§

**Cátedra de Antropología, Fac. de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba, Argentina, †Departamento de Zoología y Antropología Física, Universidad Complutense de Madrid, Spain, ‡Departamento de Sociología II, Facultad de Ciencias Políticas y Sociología, Universidad Complutense de Madrid, Spain and §Grupo Estudios Población y Sociedad (GEPS), Universidad Complutense de Madrid, Spain*

Summary. An isonymic analysis has been carried out using a sample of 1529 reconstituted families residing during 1870–1964 in Aranjuez, an urban area situated south of Madrid, Spain. The random, non-random and total-components inbreeding coefficients from isonymy were obtained and the various combinations of surnames compared in order to infer the patri- or matrilocal pattern of residence. Throughout the period studied the random component of inbreeding (F_r) has not changed, in contrast to the non-random component (F_n), thus suggesting the latter could be responsible for the reduction of total inbreeding. Using several methodological approaches (biplot analysis, α, *v* and percentage of immigrants) the predominance of the immigration of grooms was interpreted in terms of Aranjuez as a matrilocal pattern of residence. From this study it can also be concluded that surnames provided by reconstituted families are good estimators of inbreeding and migration.

Introduction

In the later decades of the 19th century and throughout much of the 20th century, many European populations experienced large demographic changes that affected mate selection processes. Rural residents streamed into larger centres and there was also a reverse trend from suburban areas towards the countryside. At the same time, the distance travelled for work and shopping purposes and also to find a mate increased (Lasker & Mascie-Taylor, 2001). In Spain, these transformations took place in different areas with a certain delay. This is the case with Aranjuez, a medium-sized town located some 50 km south of Madrid which, prior to 1931, was one of the summer residences of the Spanish royalty thanks to its palace and gardens. At the same time, Aranjuez maintained many rural characteristics, as many of its residents continued to be engaged in agricultural activities. During the period studied (1870–1964), Aranjuez experienced a substantial increase in size (from 8154 in 1877 to 29,548 in 1970), surpassing that of many other localities outside the Madrid metropolitan area. This was because Aranjuez grew both by natural means and, even more, by immigration from surrounding areas. This mostly, but not exclusively, male immigration was greatest between 1877 and 1900 and then again between 1930 and 1940, this last period coinciding with the Spanish Civil War (1936–1939). After 1940 this scenario changed as natural growth came to predominate over immigration as a source of population growth. This was a period in which male out-migration towards Madrid took place while at the same time a certain female immigration into the town continued (Sanz-Gimeno, 2000). For these reasons, differential migration by sex may have affected local mating patterns because a deficit of males due to out-migration co-existed with a relative abundance of females who probably had difficulties in finding male partners in Aranjuez (Reher, 1997).

By means of isonymy analysis, surnames provide an alternative method for studying the genetic structure of historical as well as present populations (see Colantonio *et al*., 2003, for a review). In Spain and in many Latin American countries each individual carries two surnames (the first received from the father and the second from the mother), which enables the researcher to obtain inbreeding coefficients according to the Pinto-Cisternas *et al*. method (1985) and to differentiate according to the contribution of the four existing combinations of surnames (two for the groom and two for the bride) to total inbreeding.

The objective of this paper is to establish how the changes in mating patterns in Aranjuez are reflected in its population genetic structure by means of the analysis of surnames. This will be achieved by decomposing the inbreeding coefficients from isonymy into random, non-random and total components, and by means of comparing the different combinations of surnames according to the paternal–maternal origin in order to infer information about patri- or matrilocal patterns of residence.

Methods

Based on civil registration data and applying record linkage techniques similar to the traditional method of family reconstitution, information on surnames, place of birth and dates was derived from a sample of 1529 families thanks to the work of Sanz-Gimeno. Information based on reconstituted families differs from most studies of isonymy, which preferably use marriage registers. Since using all registered marriages would have meant including couples emigrating after marriage as contributors to the local gene pool, only the reconstituted families residing in the population under study throughout the wife's complete reproductive period were used for this analysis.

Despite reconstituted families constituting relatively modest samples of all available marriages celebrated, they provide a feasible representation of couples contributing to the gene pool.

Families were selected if they had at least one child born alive and the delivery of the first child occurred within the 1870–1964 period. Only women for whom there were relatively complete reproductive histories were considered for this paper, based on the criteria of their presence in Aranjuez after the age of 45. The date assigned to each marriage corresponds to the date of birth of the first child. Two periods were considered for analysis: 1870–1900 and 1901–1964 (634 and 895 cases, respectively), which approximately correspond to a pre-transitional population and to one immersed in the demographic transition in Aranjuez.

Birth places were numerically codified in order to facilitate the selection of marriages according to their endogamous (both mates born in Aranjuez) or exogamous character (one of the mates not born in Aranjuez). The four surnames per couple appearing in the original data base were listed alphabetically ($N=1239$) and spelling mistakes were corrected. Variations were maintained when they corresponded to known surnames (for example 'Sanz', 'Sainz'). Finally, distinct surnames were coded numerically, with values between 1 and 1107 being assigned to the different surnames kept after correction.

The frequency for each surname was obtained separately for each of the four surnames of all couples. Surnames were designated according to their position as: 1–2 (groom surnames) and 3–4 (bride surnames). Inbreeding coefficients were calculated using all surnames following the Pinto-Cisternas method (1985) and distinguishing the F_r (random), F_n (non-random) and F_t (total) components of inbreeding.

The different combinations of surnames corresponding to the partners (1–3, 1–4, 2–3 and 2–4) or to the parental generation (1–2 and 3–4) can also be used to ascertain some residential patterns by plotting pairs of F_r coefficients (Eizaguirre, 1994). For example, in a biplot representing on the horizontal axis F_r of the 2–4 combination (husband's mother and wife's mother surnames) and the 1–3 combination (husband's father and wife's father surnames) on a vertical axis, identical inbreeding coefficients between mothers and between fathers will lie on the diagonal. F_r values in the 1–3 combination (surnames coming from both fathers) higher than the 2–4 combination indicate more elevated inbreeding in males, and the corresponding co-ordinates on the graph will be situated above the diagonal in the area corresponding to a 'patrilocal' pattern of migration, while the area below the diagonal (with higher F_r in the 2–4 combination, that is, surnames given by mothers) would represent the 'matrilocal' zone.

Knowledge about the diversity of surnames α (Fisher, 1943), which may be interpreted as the effective allele number within a genetic pool, was inferred by applying the formula (Barrai *et al.*, 1992):

$$\alpha = 1/(\Sigma\,(p_{ik})^2 - 1/N_i),$$

where p_{ik} represents the relative frequency of surname k in population i, summed over all surnames and N_i is the total sample size of the corresponding population. $\Sigma(p_{ik})^2$ is equivalent to random isonymy (I_r).

As a measure of immigration into the population under study (Karlin & McGregor, 1967; Barrai *et al.*, 1987), the v coefficient was calculated as:

$$v = \alpha/(\alpha + N_i).$$

Another estimator of migration taking into account the proportion of unique surnames (A) provided values for recent movement (Rodrí guez-Larralde & Barrai, 1998).

Table 1. Random (F_r), non-random (F_n) and total (F_t) components of inbreeding from isonymy, by period, combination of surnames (1: husband's first surname, 2: second; 3: wife's first surname, 4: second), and endogamy–exogamy, according to the Pinto-Cisternas method

	Period 1			Period 2		
Index	F_r	F_n	F_t	F_r	F_n	F_t
Crow						
1–3	0·00210533	−0·00013486	0·00197076	0·00214382	0·00121859	0·00335979
1–4	0·00212959	0·00182921	0·00395490	0·00205393	0·00074549	0·00279789
2–3	0·00225489	0·00136317	0·00361499	0·00228040	0·00080647	0·00308502
2–4	0·00248153	0·00194482	0·00442153	0·00244404	0·00120718	0·00364827
Pinto-Cisternas	0·00220578	0·00129007	0·00349300	0·00222520	0·00102354	0·00324645
Endog.	0·00252764	0·00090024	0·00342561	0·00266933	0·00122664	0·00389270
Exog.	0·00204285	0·00103992	0·00308065	0·00217245	0·00033935	0·00251106

Results

The coefficients of inbreeding from isonymy were obtained separately for each surname combination (1–3, 1–4, 2–3, 2–4) and together for all of them, controlling for endogamous and exogamous couples according to the Pinto-Cisternas *et al.* (1985) method (Table 1). All coefficients are very low, generally corresponding to a kinship more remote than fourth cousins. This result is not unexpected since at the time of the study Aranjuez was a relatively large town strongly affected by migration. In this sense, the percentage of grooms and brides not born in Aranjuez was larger than 20% in both cases.

Considering the four-surname combinations taken together, the random component of inbreeding (F_r) presents values almost equal in the two periods studied and higher than the preferential or non-random inbreeding (F_n). Thus, the surname frequency distribution does not reflect change over time in the population structure with regard to random inbreeding, despite the fact that, at least theoretically, population growth in Aranjuez should have decreased this component. The initially low preferential consanguinity (F_n) decreased after 1900, consequently contributing to the reduction of total inbreeding (F_t). With regard to the proportion of grooms and brides not born in Aranjuez (exogamy) F_r is lower among exogamous matings, while F_n shows the higher values for exogamous partners before 1900 and for endogamous partners afterwards. In the first period high F_n values for exogamous couples could reflect the existence of immigration, probably coming from surrounding areas, which brought into Aranjuez surnames identical to those already existing there. According to Clegg (2001) a concentrated distribution of surnames in some areas would indicate that those living in a given neighbourhood are expected to be more frequently isonymous by having more frequently the same surnames. Surnames of

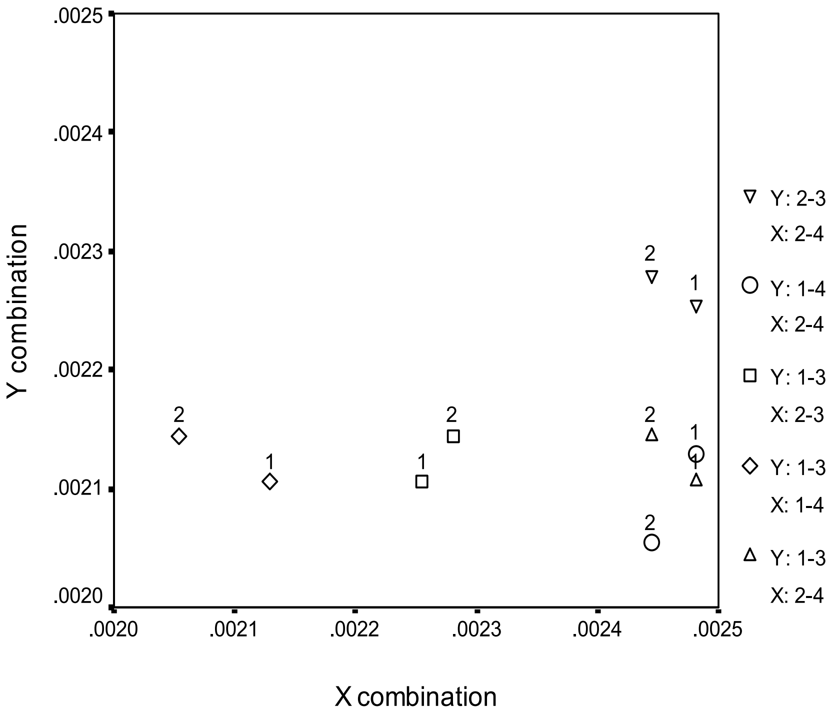

Fig. 1. Biplot drawing of F_r surname combinations for the parental generation. 1: Period 1. 2: Period 2.

immigrants established in Aranjuez throughout the first period would be counted in the second period as corresponding to the endogamous category (both mates born in Aranjuez).

A more plausible explanation for the greater preferential consanguinity among exogamous marriages could be related to 'kin-structured migration', meaning that immigrants into Aranjuez would promote the arrival of other relatives taking the same surnames. Rather than being a random sample of the donor gene pool, a kin-structured migrant group is formed by a set of biologically related kin who share genes (Fix, 1994) or surnames in studies of isonymy. This is an example of how social behaviour, governing the formation of immigrant groups, can affect the biology of a population, a result also reported by González-Martín & Toja (2002).

Regarding each surname combination, in the case of 2–4 (each mate's mother), in both periods F_t and F_r are larger than for the other combinations; F_n is one of the highest, indicating that women were more inbred than men and that migrants to Aranjuez might have been mainly males.

Although the range of variation of F_r is small, graphical representations are provided for an easier interpretation of values indicated in Tables 1 and 2. Figure 1 represents the F_r for the different pairs of surname combinations involving mates' mothers (2–4) and fathers (1–3) in a co-ordinate system. With the exception of 1–3 against 1–4 (whose values are close to the diagonal) all points have higher weights in

https://doi.org/10.1017/S0021932007002313 Published online by Cambridge University Press

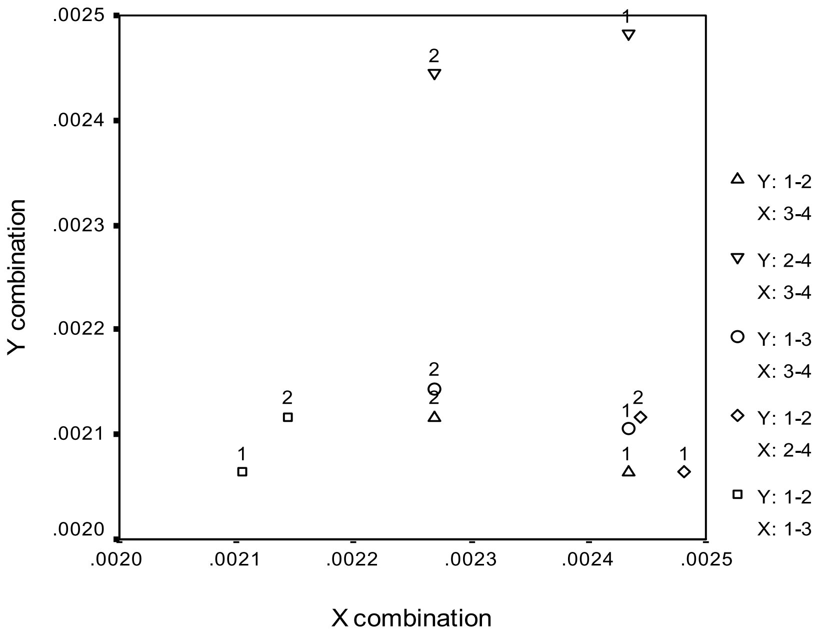

Fig. 2. Biplot drawing of F_r surname combinations for the mates and parental generations. 1: Period 1. 2: Period 2.

the *X*-axis, corresponding to the 'area of matrilocal residence', even more for the period before 1900 (Period 1).

Comparisons of F_r corresponding to the groom's surnames (1–2) appear clearly situated under the diagonal, that is in the area of matrilocality (Fig. 2), as well as for brides (3–4) compared with the groom (1–2) or with both mates' fathers (1–3) combinations. F_r obtained from the bride's surnames (3–4) compared with that of the mothers of both mates (2–4) has a higher value on the axis representing 2–4 (*Y*) than in the *X*-axis (1–2), indicating that women of the preceding generation had an even higher tendency to matrilocality. This could be related to higher immigration of males, who after marriage would became residents in their wife's place of residence, in this case, Aranjuez, especially between 1870 and 1900.

With respect to temporal variation, apart from comparisons including only males (1–2 with 1–3) and only women (2–4 with 3–4), the three remaining F_r combining both sexes (1–2 with 3–4, 1–3 with 3–4, 1–2 with 2–4), all of which correspond to the female combinations, present larger values on the *X*-axis in Period 1. The above would confirm a greater tendency to matrilocality and more frequent masculine immigration during that period.

Table 2 shows indicators of migration calculated from isonymy, compared with the percentage of in-migrants, as derived from place of origin of parents that appear on the birth certificates of the children of these reconstituted families. Random isonymy (I_r) and a have also been compared with the Aranjuez values appearing in the data base for all Spanish localities (Rodríguez-Larralde *et al.*, 2003:), the latter shown in parentheses in Table 2.

Table 2. Isonymy and migration estimates by sex and period

Sex	Period	N	I_r	a	v	A	% mig.
Men	1	634	0·008421	146·11	18·730	23·445	25·21
			0·008575	134·13			
Men	2	893	(0·00735)	(136·1)	13·059	18·233	25·28
Women	1	634	0·009926	119·78	15·890	20·695	24·02
			0·009776	115·52			
Women	2	893	(0·00856)	(116·8)	11·454	18·548	23·07

N, sample size; I_r, random isonymy; a:, surname diversity; v, immigration; A, recent immigration. Percentage of immigrants derived from birth places. In parentheses, Aranjuez values from http://web.unife.it/progetti/genetica/pdata.htm#iso

In both periods the diversity of surnames (a) is greater among men than among women; the indicators of migration (v) are also larger for men than for women within each period, confirming that immigration was predominantly masculine, in agreement with the matrilolocality pattern commented above. This result tends to support the notion that throughout Spanish history matrilocal residence patterns have tended to predominate in most regions (Pérez Moreda, 1986).

The fact that both a and v are more elevated during the first period is also concordant with a matrilocal tendency more evident before 1901, though still present afterwards.

Comparing recent immigration from isonymy (A) with the percentage of in-migrants according to birth place, the values of the first are lower, which could be interpreted in terms of the hypothesis commented before of the existence of kin-structured immigration. This applies mainly to the second period, which explains why the differences between both indicators are larger.

The values of the random or expected isonymy (I_r) obtained in the present analysis are rather high in comparison with the Rodríguez-Larralde *et al*. (2003) figures of 0·00735 for males and 0·00856 for females. This may be explained by the fact that the data used by these authors are based on telephone users in Aranjuez in the year 2000, while our data go back to 1901. In all cases female I_r values exceed those of males, indicating a higher proportion of surnames shared by women than by men. This, in addition to a lower surname diversity (a) among women in the present study and also in Rodríguez-Larralde (2003), supports the existence of a matrilocal model of residence, which even today continues to be the case in many Spanish regions (Iglesias de Ussel, 1987).

From this analysis it can be concluded that in Aranjuez surname data provided by reconstituted families provide reliable values of inbreeding and migration as indicated by the basic agreement of the different indicators used in this paper. The random component of inbreeding (F_r) shows temporal stability while the non-random component (F_n) changed with time, suggesting that the latter is responsible for the reduction of total inbreeding. Various types of data analysis (biplot analysis, a, v and percentage of immigrants) all point to the existence of predominantly male immigration and matrilocal patterns of residence in Aranjuez during the period under study.

Acknowledgments

This research received funding from the Fundación Banco Bilbao Vizcaya (BBVA) [BBVA02–11562] and from the Comunidad Autónoma de Madrid (CAM) [06/0105/97, 06/0070/98, 06/0046/2000, 06/0180/2001, 06/0033/2002].

References

Barrai, I., Barbujani, G., Beretta, M., Maestri, I., Russo, A., Formica, G. & Pinto-Cisternas, J. (1987) Surnames in Ferrara: distribution, isonymy and levels of inbreeding. *Annals of Human Biology* **14**, 415–423.

Barrai, I., Formica, G., Scapoli, C., Beretta, M., Mamolini, E., Volinia, S., Barale, R., Ambrosio, P. & Fontana, F. (1992) Microevolution in Ferrara: Isonymy 1890–1990. *Annals of Human Biology* **19**, 371–385.

Clegg, E. J. (2001) Some factors affecting marital isonymy in three Outer Hebrian parishes. *Annals of Human Biology* **28**, 455–466.

Colantonio, S. E., Lasker, G., Kaplan, B. & Fuster, V. (2003) Use of surname models in human populations: A review of recent developments. *Human Biology* **75**, 785–807.

Eizaguirre, M. (1994) Down to the river: Marital movement and genetic structure on the Hispanic–Portuguese Border. Masters Thesis, University of Durham, Department of Anthropology, p. 136.

Fisher, R. A. (1943) The relation between the number of species and the number of individuals in a random sample of animal population. *Journal of Animal Ecology* **12**, 42–58.

Fix, A. G. (1994) Detecting clinal and balanced selection using spatial autocorrelation analysis under kin-structured migration. *American Journal of Physical Anthropology* **95**, 385–397.

González-Martín, A. & Toja, D. I. (2002) Inbreeding, isonymy, and kin-structured migration in the Principality of Andorra. *Human Biology* **74**, 587–600.

Iglesias de Ussel J. (1987) *Sociología del noviazgo en España.* Caja General de Ahorros y Monte de Piedad, Granada.

Karlin, S. & McGregor, J. (1967). The number of mutant forms maintained in a population. In *Proceedings of the Fifth Berkeley Symposium on Mathematics, Statistics and Probability*, vol. 4, pp. 415–438.

Lasker, G. W. & Mascie-Taylor, C. G. N. (2001) The genetic structure of English villages: surname diversity changes between 1976 and 1997. *Annals of Human Biology* **28**, 546–553.

Pérez Moreda, V. (1986) Matrimonio y familia. Algunas consideraciones sobre el modelo matrimonial español en la Edad Moderna. *Boletín de la Asociación de demografía histórica* **IV**(1), 3–51.

Pinto-Cisternas, J., Pineda, L. & Barrai, I. (1985) Estimation of inbreeding by isonymy in Iberoamerican populations: an extension of the method of Crow and Mange. *American Journal of Human Genetics* **37**, 373–385.

Reher, D. S. (1997) *Perspectives on the Family in Spain Past and Present.* Clarendon Press, Oxford.

Rodríguez-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., Gonzales-Martin, A., Scapoli, C. & Barrai, I. (2003) The names of Spain: A study of the isonymy structure of Spain. *American Journal of Physical Anthropology* **121**, 280–292.

Sanz-Gimeno, A. (2000) *Aranjuez, 1870–1970. Cien años de historia demográfica. Aranjuez Studia* **3**.