

SIMILARITIES IN THE SURNAMENES OF ISLAND AND CONTINENTAL POPULATIONS OF THE NORTH-WESTERN MEDITERRANEAN AREA

E. LUCCHETTI*, M. TASSO†, P. PIZZETTI*, S. DE IASIO* AND
G. U. CARAVELLO†

**Dipartimento di Biologia evolutiva e funzionale, University of Parma, Italy and*
†*Dipartimento di Medicina Ambientale e Sanità Pubblica-sede di Igiene, University of Padua, Italy*

Summary. This paper compares the structures of the surnames of 75 municipal populations living in six north-western Mediterranean regions. Its purpose is to unravel the relations between the local populations in Corsica and Sardinia and the links between these populations and those living in the Italian and French continental territory. On the basis of the matrix of similarity of surnames, some topological representations have been drafted showing the above-mentioned relations between populations under the light of their geographical position, their recent history and studies of genetic analysis. Corsica has an heterogeneous surname structure and evident similarity of the north with Tuscany and some centres of continental France. When only the populations of Sardinia were taken into consideration, it emerged that they differ among each other in relation to their geographical position and their history; when, instead, they were considered in relation to other populations outside the island, it was possible to observe that they form a highly different cluster. This study also identified many differences in the analysed geographical areas of Sardinia. In the minor islands – Elba, Giglio, Capraia – the structure of the surnames has a Tuscan origin as well as some similarity with other geographically distant areas, as in the case of the island of Giglio, if compared with some communities of Liguria.

Introduction

Surnames allow identification of a person. They date back to the late Middle Ages and Renaissance, when the first economic, administrative, legal and religious institutions were established in Western European society: it was then when surnames became crystallized and began to pass hereditarily from father to son (De Felice, 1978). Surnames reflect the very culture and traditions of the place in which they were created, being at the same time symbols that identify a family and join together all its members.

Surnames are inherited through the male line and can be considered neutral alleles at a locus on the Y chromosome (Zei *et al.*, 1983a, 1983b, 1986). As a consequence of this, they satisfy the expectations of the neutral theory of evolution, which is entirely described by random genetic drift, mutation and migration (Kimura, 1983). Thanks to their high polymorphism surnames allow a great differentiation between both big populations, although geographically and historically far apart; and small, neighbouring populations, when biological markers cannot identify any significant difference (Lucchetti & Soliani, 1989). Surnames are linked to cultural factors. They are inherited biologically and they spread over an area depending on the displacements of their bearers. They are useful in order to study populations as they comprise the sociocultural dimension, the family and genetic dimension, the territorial and geographical dimension and the temporal dimension (Brunet *et al.*, 2001).

In the last decades, surnames have been increasingly used by biodemographers, anthropologists and geneticists (Crow, 1996; Colantonio *et al.*, 2003), which implies a dramatic rise in the comprehensive range of possible topics to be analysed and the creation of new and in-depth methods of analysis (Brunet *et al.*, 2001; Lucchetti & Soliani, 2004).

Telephone directories provide data about surnames and their frequency of distribution, giving useful information that overcomes the above-mentioned hindrances, assuming an equal telephone diffusion. Surnames taken from official telephone directories have been used to carry out much research, some of which relates to the whole national territory of a country. This is the case for studies that have calculated migration rates in Italian provinces and compared them with those taken from official demographic sources (Piazza *et al.*, 1987). These studies identified five main clusters in which more than one hundred Italian towns are divided (Barrai *et al.*, 1999) and demonstrated the effects of internal migration in Italy, in particular from southern to northern regions, taking into consideration geographical, historical social and economic factors (Lucchetti *et al.*, 1996).

This paper seeks to compare the structures of the surnames of 75 municipal populations living in the north-western Mediterranean area (Figs 1 and 2; Table 1), in order to find out the relations among them and how they are influenced by the nation factor, the island factor, their geographical position and their different history. In particular, the paper will underline the following relations: the one between Corsican and Sardinian populations and the one between these insular communities and those living in Italy and southern France. Since they live on islands, Corsican and Sardinian populations show some peculiarities, which many in-depth studies have demonstrated. Sardinia is anomalous with respect to Italy and the remaining parts of the Mediterranean area (Cavalli-Sforza *et al.*, 1994; Vona, 1997), whereas Corsica is more genetically heterogeneous (Varesi *et al.*, 1996), although it shares some genetic features with the surrounding regions (Calafell *et al.*, 1996), particularly Sardinia (Calò *et al.*, 1998; Morelli *et al.*, 1999; Varesi *et al.*, 2000).

Methods

This paper considers 75 municipal populations living in France and Italy, who therefore speak different languages and differ in history and traditions. One group is

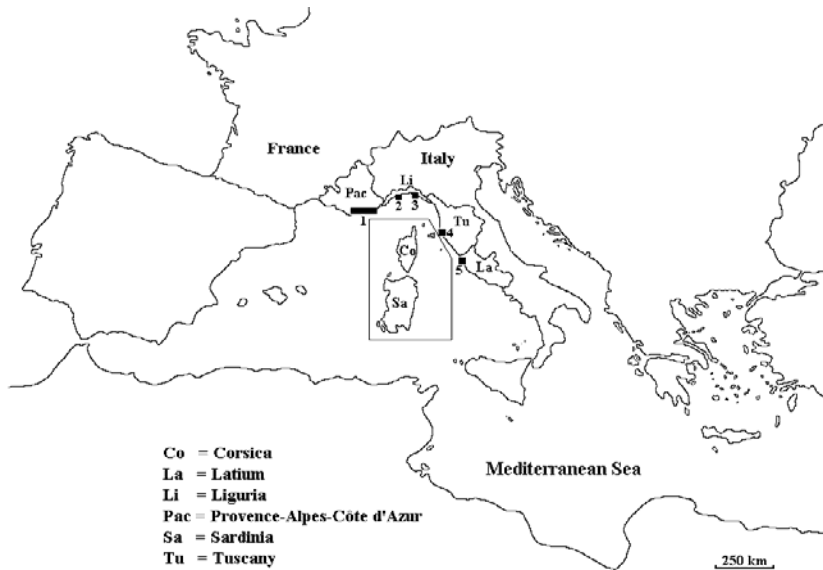


Fig. 1. The geographical positions of Corsica, Sardinia and the Mediterranean area. The area comprising the islands that were studied is shown in bigger proportions in Fig. 2. The numbers 1 to 5 refer to the areas comprising the coastal continental Italian and French municipalities listed in Table 1.

formed by 25 Corsican municipalities, eighteen belonging to central–southern Corsica; another consists of 25 Sardinian municipalities, eighteen in northern Sardinia; and the last group consists of 25 other populations living in Italy, France and some minor islands historically or geographically linked to Corsica and Sardinia. These are basically coastal regions whose peculiarities have not been altered by the tourism factor. They include the Liguria region, Tuscany, Latium and Capraia, Elba, the Giglio islands (Tuscany) in Italy, and the Provence-Alpes-Côte d’Azur region in France (Figs 1 and 2; Table 1).

The distribution of surnames for each local population was obtained from official telephone directories. Among the subscribers, only physical persons were considered, in order to provide an overview of family surnames spread over the study area. Due to national differences in bureaucracy, organization and contents, there have been few analyses that have gathered of surnames from different countries.

The surnames taken from telephone directories were used to calculate similarities between municipal populations using the R_{ij} index (Chen & Cavalli-Sforza, 1983), or index of ‘standardized isonymy’, by the formula:

$$R_{ij} = \frac{\sum p_{ik} p_{jk}}{\sqrt{\sum p_{ik}^2 \sum p_{jk}^2}}$$

where R_{ij} is the coefficient of similarity between two populations i and j ; p_{ik} and p_{jk} are the frequencies of the k^{th} surname in these populations; and the total is extended

Table 1. List of all the abbreviations of all the municipalities considered

Abbr.	Municipality	Subscribers	Surnames	Region	Nation
Ca	Aleria	666	416	Corsica	France
Cb	Aullene	158	83	Corsica	France
Cc	Bonifacio	1166	744	Corsica	France
Cd	Calvi	1863	1285	Corsica	France
Ce	Carbini	83	39	Corsica	France
Cf	Cargese	592	385	Corsica	France
Cg	Conca	306	151	Corsica	France
Ch	Figari	386	219	Corsica	France
Ci	L'Île Rousse	1217	709	Corsica	France
Cj	Levie	396	194	Corsica	France
Ck	Monacia d'Aullene	247	112	Corsica	France
Cl	Murato	279	118	Corsica	France
Cm	Oletta	394	271	Corsica	France
Cn	Patrimonio	299	197	Corsica	France
Co	Petretto-Bicchisano	307	146	Corsica	France
Cp	Pianottoli-Caldarelo	344	225	Corsica	France
Cq	Porto Vecchio	3936	2291	Corsica	France
Cr	Propriano	1084	688	Corsica	France
Cs	Quenza	244	118	Corsica	France
Ct	Saint Florent	718	484	Corsica	France
Cu	Sainte-Lucie-de-Tallano	301	141	Corsica	France
Cv	Santo Pietro di Tenda	186	94	Corsica	France
Cw	Sartene	1360	656	Corsica	France
Cx	Sotta	421	188	Corsica	France
Cy	Zonza	254	154	Corsica	France
Fa	Agay	1220	1089	Provence-Alpes-Côte d'Azur ¹	France
Fb	Rayol Canadel sur Mer (Le)	809	749	Provence-Alpes-Côte d'Azur ¹	France
Fc	Saint Jean Cap Ferrat	1034	873	Provence-Alpes-Côte d'Azur ¹	France
Fd	Saint Mandrier sur Mer	2413	2050	Provence-Alpes-Côte d'Azur ¹	France
Fe	Théoule sur Mer	1471	1350	Provence-Alpes-Côte d'Azur ¹	France
La	Arenzano	6143	3637	Liguria ²	Italy
Lb	Bogliasco	2170	1495	Liguria ³	Italy
Lc	Cogoleto	4046	2280	Liguria ²	Italy
Ld	Pieve Ligure	1225	903	Liguria ³	Italy
Le	Sori	2105	1244	Liguria ³	Italy
Sa	Aggius	498	212	Sardinia	Italy
Sb	Aglientu	673	462	Sardinia	Italy
Sc	Arzachena	2378	969	Sardinia	Italy
Sd	Badesi	527	194	Sardinia	Italy

Table 1. Continued

Abbr.	Municipality	Subscribers	Surnames	Region	Nation
Se	Bortigiadas	274	95	Sardinia	Italy
Se	Bortigiadas	274	95	Sardinia	Italy
Sf	Burcei	728	136	Sardinia	Italy
Sg	Calangianus	1149	386	Sardinia	Italy
Sh	Carbonia	9664	2383	Sardinia	Italy
Si	Carloforte	2645	776	Sardinia	Italy
Sj	Castelsardo	1565	565	Sardinia	Italy
Sk	La Maddalena	4244	2185	Sardinia	Italy
Sl	Luogosanto	540	214	Sardinia	Italy
Sm	Luras	805	246	Sardinia	Italy
Sn	Muravera	1634	724	Sardinia	Italy
So	Palau	1499	979	Sardinia	Italy
Sp	Porto Torres	5683	2196	Sardinia	Italy
Sq	San Vito	1121	292	Sardinia	Italy
Sr	Santa Maria Coghinas	385	178	Sardinia	Italy
Ss	Santa Teresa Gallura	1805	1015	Sardinia	Italy
St	Tempio Pausania	3844	1256	Sardinia	Italy
Su	Trinita d'Agultu e Vignola	915	510	Sardinia	Italy
Sv	Valledoria	978	434	Sardinia	Italy
Sw	Viddalba	451	180	Sardinia	Italy
Sx	Villaputzu	1305	353	Sardinia	Italy
Sy	Villasimius	1152	583	Sardinia	Italy
Ta	Campiglia Marittima	4281	1879	Tuscany ⁴	Italy
Tb	Capraia Isola	174	152	Tuscany	Italy
Tc	Castagneto Carducci	3642	1855	Tuscany ⁴	Italy
Td	Castiglione della Pescaia	4353	2557	Tuscany ⁴	Italy
Te	Elba Capoliveri	1330	869	Tuscany	Italy
Tf	Elba Marciana	1349	844	Tuscany	Italy
Tg	Elba Marciana Marina	1083	651	Tuscany	Italy
Th	Elba Marina di Campo	1269	763	Tuscany	Italy
Ti	Elba Rio Marina	565	390	Tuscany	Italy
Tj	Follonica	9300	4316	Tuscany ⁴	Italy
Tk	Isola del Giglio	782	454	Tuscany	Italy
Tl	San Vincenzo	3304	1856	Tuscany ⁴	Italy
Tm	Scarlinò	1100	671	Tuscany ⁴	Italy
Za	Montalto di Castro	3277	1806	Latium ⁵	Italy
Zb	Tarquinià	6667	3129	Latium ⁵	Italy

Each municipality is matched to the total number of telephone subscribers and surnames found in telephone directories. Also given are the names of the regions and nations each municipality belongs to. The exponents on the Provence-Alpes-Côte d'Azur region¹, Liguria region^{2,3}, Tuscany⁴ and Latium⁵ refer to the areas shown in Fig. 1.

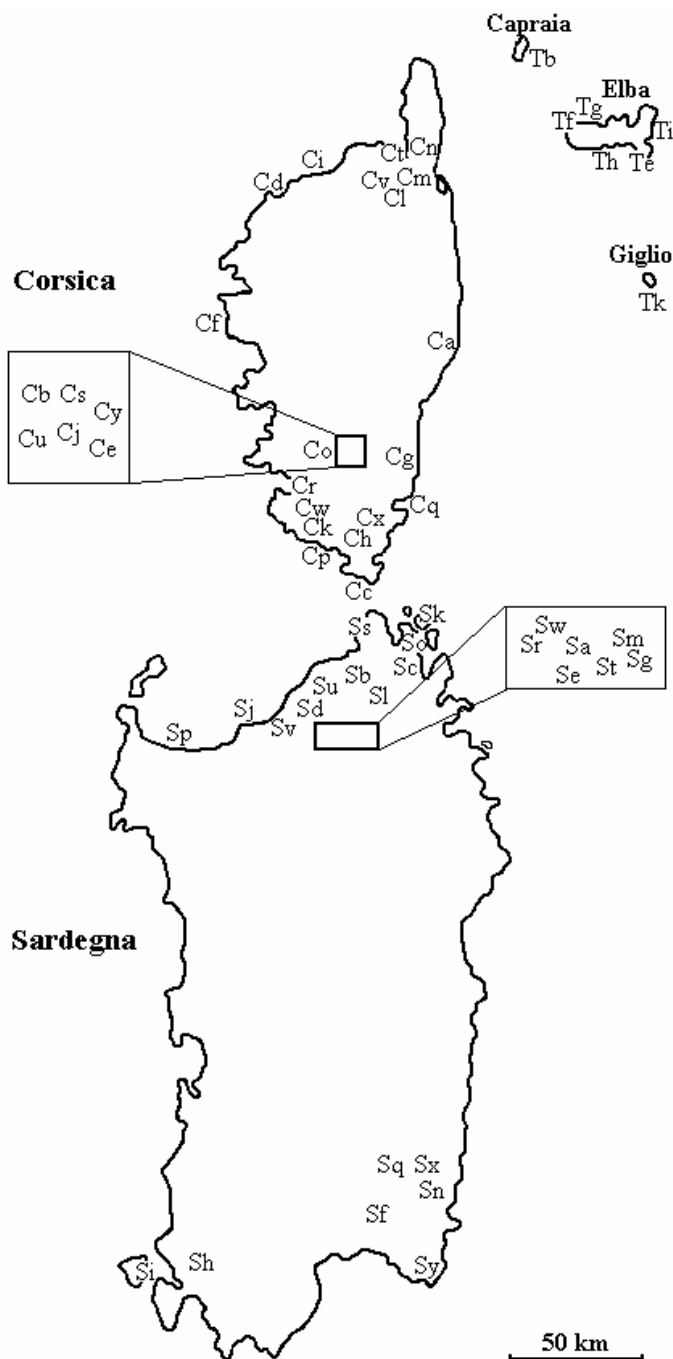


Fig. 2. Enlargement of the area represented in Fig. 1 comprising the islands considered in this paper. The corresponding abbreviations can be found in Table 1.

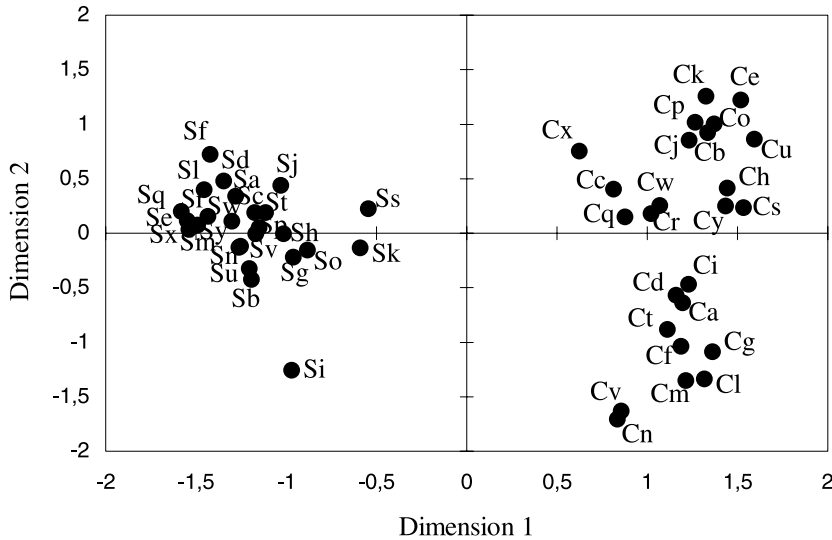


Fig. 3. Topological representation showing the relations between Corsican and Sardinian municipalities (Stress=0.20251; $R^2=0.83970$). Each municipality has an abbreviation (Table 1) and its geographical position is shown in Fig. 2.

to all surnames. Comparing all pairs of populations it is possible to create a matrix of similarity and, by applying to it ‘non-metric multidimensional scaling’ (MDS) (Schiffman *et al.*, 1981; Davidson, 1983), it is possible to create a topological representation that can provide graphical evidence for the relations existing among the above-mentioned populations. This representation is easy to interpret and respects the degree of similarity that exists between one population and all the others, thus reducing to a minimum the distortion that is necessarily introduced when expressing the phenomenon on a small number of dimensions. The accuracy of the MDS solution is evaluated by two goodness-of-fit measures: the S-Stress level (stress) and the R^2 test (RSQ). Stress is a fit measure that measures the difference between interpoint distances in computed MDS space and the corresponding actual input distances; RSQ is simply the squared correlation of the input distances with the scaled distances using MDS coordinates (Kruskal & Wish, 1978; Cox & Cox, 2001).

This analysis is based on topological representations that take into consideration the relations among different geographical regions. The purpose is to identify the factors that influence the relations among populations, such as the nation factor, the island factor, geographical position and historical events.

Results

Figure 3 compares Corsican and Sardinian populations: the representation shows a deep gap between Corsica and Sardinia, clear and without exceptions, on the first dimension. Sardinian populations make up a close cluster, with little dispersion on either the first or the second dimension. The only exceptions are Santa Teresa Gallura

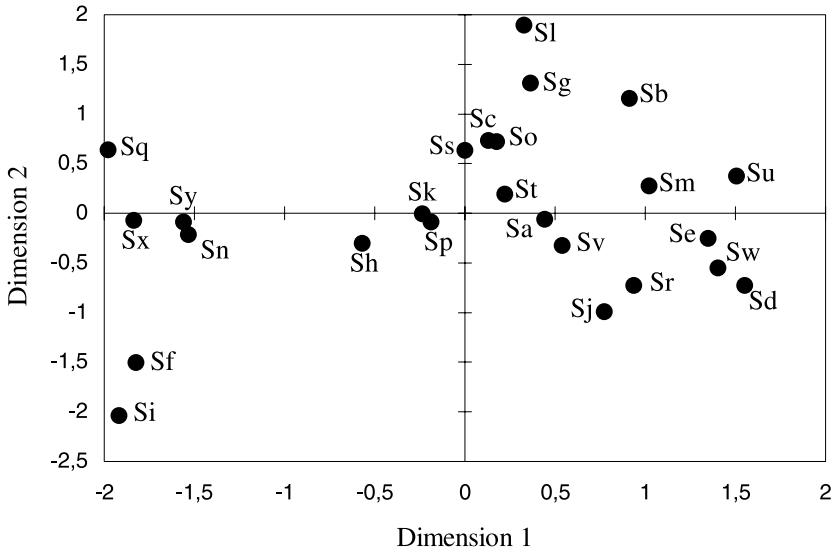


Fig. 4. Topological representation showing the relations among Sardinian municipalities (Stress=0.18022; $R^2=0.85406$). Each municipality has an abbreviation (Table 1) and its geographical position is shown in Fig. 2.

and La Maddalena, which are shifted slightly towards the centre of the chart, and Carloforte, with a bigger shift on the second dimension. The second dimension shows that there is greater variability among Corsican municipalities. Two main clusters can be identified and linked to their geographical position: southern municipalities at the top of the chart and northern ones at the bottom. Conca is the only exception: it is in southern Corsica but belongs to the northern populations in the figure.

Figure 3 shows that Sardinia makes up a close cluster without dispersions, but if only Sardinian municipalities are studied, there is variability among them (Fig. 4). They disperse on the first dimension: southern municipalities are on the left of the chart whereas northern ones are on the right. There is a wide gap between them and Carbonia is the only exception: it is in the middle of the chart although it belongs geographically to southern Sardinia. As far as the second dimension is concerned, Carloforte appears to be still separated from the other municipalities and a slight differentiation can be seen between western and eastern municipalities in the north.

Sardinian peculiarity can be observed both in the comparison with Corsican populations and with the populations living in the continental Italian regions and in the minor islands (Fig. 5). Considering the first dimension, Sardinian populations are on the right part of the chart and form a new cluster; however, Carloforte is not included in it if the second dimension is considered. All the other Italian populations are on the left part of the representation, with more dispersion on the second dimension. The Ligurian populations and the island of Giglio in Tuscany, pull away from the main cluster on the second dimension at the bottom of the chart. The populations of Latium and Tuscany are at the top of the figure and the highest values

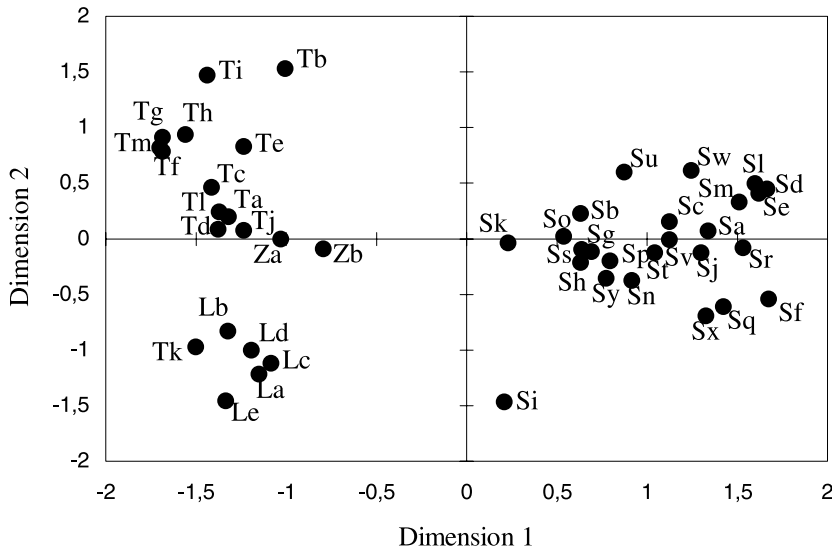


Fig. 5. Topological representation showing the relations between Sardinian municipalities and the Italian regions of Liguria, Tuscany and Latium (Stress=0.16535; $R^2=0.88763$). Each municipality has an abbreviation (Table 1) and its geographical position is shown in Figs 1 and 2.

are represented by the islands of Elba and Capraia (Tuscany), although they are not very different from those of the other Tuscan municipalities.

From the comparison of French municipalities alone (Fig. 6) it is clear that the island factor does not affect Corsican populations. These municipalities are dispersed on the first dimension and two main clusters can be identified: the first one on the right, made up of southern populations, and a second one on the left, represented by northern populations. The five French continental municipalities have more or less the same values as those of northern Corsican populations as far as the first dimension is concerned, whereas they differ from them on the second dimension. Conca is again far from the southern Corsican sub-cluster.

Sardinian populations make up a close cluster on the left part of the representation, whereas Italian and Corsican municipalities are on the right (Fig. 7). Corsican and Italian continental populations share more or less the same values on the first dimension. The nation factor, which marks the difference between Corsican and Italian municipalities not belonging to Sardinia, is evident on the second dimension, although some values overlap.

Similar results can be obtained from the comparison of Sardinian, continental French and Corsican municipalities (Fig. 8). Sardinia makes up a very close cluster but there are a few exceptions: Santa Teresa Gallura and La Maddalena to the right, and Carloforte at the bottom. French municipalities are on the right part of the chart. They are quite dispersed on the second dimension: continental municipalities to the bottom of the chart, northern Corsican municipalities in the middle, and southern Corsican municipalities at the top. The island factor appears on the second

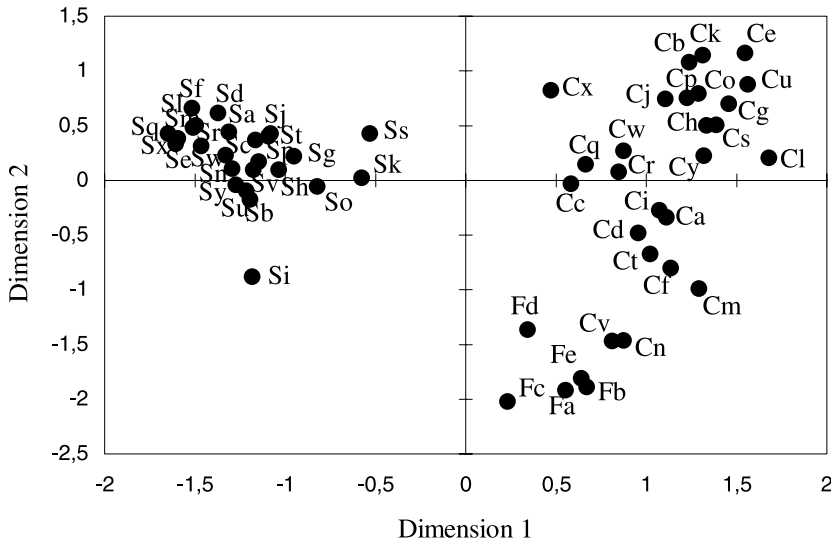


Fig. 8. Topological representation showing the relations among Corsican, Sardinian and continental French municipalities (Stress=0.21148; $R^2=0.81445$). Each municipality has an abbreviation (Table 1) and its geographical position is shown in Figs 1 and 2.

dimension: it is more pronounced for southern Corsican populations, and less pronounced for northern ones, some of which are included in the cluster made up of French continental municipalities.

The final topological representation comprising all the 75 municipalities (Fig. 9) provides evidence once again for the peculiarity of Sardinian populations, which make up a very close cluster with the only exceptions being Santa Teresa Gallura, La Maddalena and Carloforte. The first dimension shows the wide gap between Sardinian municipalities and all the other populations considered; the second dimension shows the gap between Corsican and French continental populations. The island factor is not very pronounced for northern Corsica, whose values are quite close to those of Italy and continental France. Moreover, Saint Mandrier sur Mer, in continental France, is close to southern Corsican municipalities in the chart. The nationality factor does not appear clearly either on the first or on the second dimension, although the values for French populations are generally higher on the first dimension.

Discussion

This paper analyses the distribution of surnames taking into consideration 75 municipalities of six regions in the north-western Mediterranean area. Of the populations analysed, the peculiarity of Sardinia is clear since it makes up a close cluster (Figs 3, 5, 7, 8 and 9). However, inside Sardinian territory there is some variability in the distribution of surnames, with a gap between the north and the south (Fig. 4). The

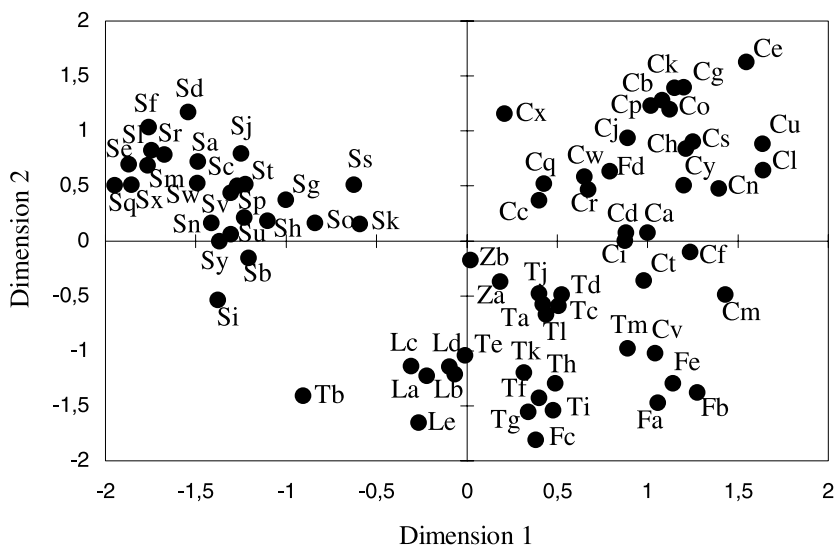


Fig. 9. Topological representation showing the relations among Corsican, Sardinian, continental French and Italian (Liguria region, Tuscany and Latium) municipalities (Stress=0.23243; $R^2=0.73578$). Each municipality has an abbreviation (Table 1) and its geographical position is shown in Figs 1 and 2.

difference between Sardinian and non-Sardinian populations is always visible on the first dimension, without considering the nationality factor or the geographical position. Corsica is more heterogeneous (Figs 3, 6, 7, 8 and 9): despite the island factor, the municipalities are more dispersed and there is a gap between the north and the south that varies depending on the populations considered. There is no evidence of relations between northern Sardinian and southern Corsica, with the exception of some seaports like Bonifacio and Porto Vecchio in Corsica, and La Maddalena and Santa Teresa Gallura in Sardinia. In the topological representation they are placed at the edge of their own cluster, slightly oriented towards the cluster of the opposite island. From the global analysis (Fig. 9) and from the comparison with France (Fig. 6) and Italy (Fig. 7), it is clear that northern Corsican populations are more similar to continental French and Italian municipalities than southern Corsican ones. This can be explained by geographical features, although it is probable that it is due to historical reasons, as well as, to a certain degree, the isolation of southern populations. The nation factor has less importance than the very peculiarity of Sardinia and the island factor.

The structure of the surnames changes if the composition of the populations changes. In Sardinia there is a large predominance of native surnames that reflects the Sardinian linguistic and onomastics tradition (De Felice, 1980; Pittau, 1990), as well as a pronounced demographic isolation (Francalacci *et al.*, 2003), which is one of the main reasons for the peculiarity of the Sardinian surname structure. On the other hand, in Corsica more than 50,000 inhabitants out of a total of 260,000 Corsicans were originally from continental France, Sardinia, Portugal and North Africa, in

particular Morocco and Tunisia (Maxia, 2002). After the end of the War of Algeria (1962) a lot of families from Algeria, most of which were of Corsican origin, went back to Corsica (Caratini, 1995). Internal migrations, however, are limited and typical of certain areas, as a study on distribution and similarity of frequencies of surnames showed, which considered the main lines of communication in the island (Caravello & Tasso, 2002a). This is confirmed by Conca municipality which, although belonging to southern Corsica (Fig. 2) in the topological representations, is included in the cluster of northern populations (Figs 3 and 6), probably thanks to easy displacements along the eastern coast, without hills or rough inlets (M. Maxia, personal communication).

As already demonstrated, Sardinian populations are more homogeneous than Corsican ones. In fact, Sardinian populations make up a close cluster if compared with others; whereas, if the 25 Sardinian populations are analysed alone they show some internal differences depending on their geographical position, with a wide gap between the north and the south, as already confirmed by linguistic and genetic studies as well as analyses on surnames taking into consideration some Sardinian areas (Cavalli-Sforza *et al.*, 1994). However, there are some exceptions. In all the topological representations of Sardinia, Carloforte is not included in the main Sardinian cluster since its inhabitants have Ligurian origins and they settled in Sardinia in the 18th century, coming from the island of Tabarka in north-western Tunisia (Vallebona, 1988). The structure of surnames in Carloforte is still rather peculiar, and very different from that of the other Sardinian municipalities (Vona *et al.*, 1996; Caravello & Tasso, 2002b). Carbonia, another Sardinian municipality not far from Carloforte (Fig. 2), is included in the group of northern Sardinian municipalities in the chart (Fig. 4), although it belongs geographically to the south. This is due to the recent origins of Carbonia: it was created in 1938 as a district where coal could be exploited. A lot of settlers, mainly from Sardinia (Delogu, 2003), migrated to Carbonia in order to find a job. As a consequence of this, the structure of surnames in this area is very heterogeneous, being composed of surnames typical of various parts of Sardinia. Last, but not least, Santa Teresa Gallura and La Maddalena, on the Sardinian coast (Fig. 2), are slightly separated from the main Sardinian cluster (Figs 3, 7, 8 and 9): this is probably due to the influence of Corsican surnames, since Corsican people migrated to Sardinia contributing to the foundation of the above-mentioned municipalities (Maxia, 2002). Moreover, the foreign population of La Maddalena has increased since the 1970s thanks both to tourism and the presence of a United States naval base (Racheli, 1982).

The island factor is very important also for the minor islands belonging to the Italian region of Tuscany. The five municipalities of the island of Elba make up a cluster (Figs 7 and 9), although not very close in spite of the small size of the island. This may be explained historically: at the end of the 18th century the island was divided among three important high-class families. Moreover, since the 1960s tourism has been one of the main activities of the island (Lambertini, 2004) and as a consequence of this, a lot of non-native surnames were recorded in the official telephone directories. Giglio and Capraia are small islands and each of them makes up a single municipality: they belong to different clusters depending on the topological representation considered. They share some features with the cluster of the island of

Elba and Tuscany, as well as with that composed of Ligurian municipalities (Figs 5, 7 and 9), probably due to the small size of their populations and their past history. The island of Giglio was attacked by the Saracens and the Barbarians, who almost depopulated the whole island. It was then re-populated by Tuscan, Ligurian and Neapolitan families, who brought with them surnames typical of their native regions, still existing today among the inhabitants (Lambertini, 2002). Most people have Tuscan features (Figs 7 and 9), although there are some analogies with Ligurian populations (Fig. 5). Capraia is closer to Corsica than Italy (Figs 1 and 2) and it was a penal settlement for many years: this slowed the economic and social development of the island, which caused constant migrations in search of job opportunities abroad (Barsotti & Lambertini, 1999). All this probably brought about a peculiar structure of surnames that characterizes Capraia, making it different from the other Italian municipalities.

Studies carried out through the use of genetic markers confirm the peculiarity of Sardinia. An average number of 88 genes of 26 European populations were analysed in order to find the genetic distances among them: the study revealed that Sardinia differs greatly from the other populations, ranking second following the Laplanders (Cavalli-Sforza *et al.*, 1994). Moreover, the very research carried out to find the differences between Italian regions confirm Sardinia is different from the other regions, both if surnames (Lucchetti *et al.*, 1996) and genetic information (Soliani *et al.*, 1981, 1982) are considered. There is also a linguistic boundary separating Sardinia from the rest of continental Italy (Barbujani & Sokal, 1990; Sokal *et al.*, 1990). The main reason for the genetic difference between the Sardinians and the other Europeans lies in the genetic drift: in fact, the frequency of most genes in Sardinia is very different from the average frequency of the other European regions, although in the past Sardinia was invaded several times and the colonizers did not leave many genetic traces (Cavalli-Sforza *et al.*, 1994).

Genetic studies can also explain the heterogeneity of Corsican populations, providing evidence for the differences between the north, the centre and the south (Varesi *et al.*, 1996; Vona *et al.*, 2003). The explanation for the internal genetic variability might be found in the marriage structure, which characterized even the recent past of the Corsican communities: the rates defining exogamy between the different micro-regions have consistently low values, particularly for the two most differentiated areas (Vona *et al.*, 2002). The above-mentioned variability might also be explained by the low density of the population and genetic drift.

This paper clearly demonstrates that there is no similarity between northern Sardinia and southern Corsica, although studies carried out through genetic markers give conflicting results. On the one hand, although Sardinia and Corsica are geographically very close together, their populations are very different from the ethnic, historical (Cavalli-Sforza *et al.*, 1994) and linguistic point of view, due to the presence of a linguistic boundary between them (Barbujani & Sokal, 1990; Sokal *et al.*, 1990). On the other hand, with respect to the other Mediterranean communities, Sardinian populations seem to have the greatest genetic affinity with the Corsicans, thereby reaffirming the observations of previous work (Vona *et al.*, 1995; Varesi *et al.*, 1996; Memmi *et al.*, 1998) that had suggested a certain level of genetic similarity between the populations of the two islands; it is a similarity resulting from

a common initial peopling and the constant exchange of culture and people. Indeed, another study (Tofanelli *et al.*, 2001) pointed out in an STR (Short Tandem Repeat: a class of DNA polymorphisms) analysis that the north of Corsica was closer to Tuscany compared with the south of Corsica, which was, on the contrary, genetically closer to Sardinia. This is also confirmed by the analysis of surnames, showing a great similarity between northern Corsica and some Tuscan municipalities as well as the Provence-Alpes-Côte d'Azur region. On the contrary, there are no similarities between southern Corsica and northern Sardinia (Figs 7, 8 and 9). The current structure of Sardinian surnames reflects the linguistic and the local onomastics tradition of Sardinia (De Felice, 1980; Pittau, 1990), which is shown by the clusters composed of Sardinian municipalities. The structure of Corsican surnames has, instead, altered over the centuries, being now quite heterogeneous. Only the surnames of small mountain villages of the interior have kept the same structure as that identified a few centuries ago (Maxia, 2002). The variability of surnames is more pronounced in coastal Corsican areas than in isolated regions where there might have been genetic drift phenomena (Morelli *et al.*, 2002).

The internal variability of surnames and the relations among Corsica, Sardinia and continental regions can be explained through the history of Corsican peopling. The geographical position of Corsica might have favoured the arrival of invaders and immigrants from various parts of the Mediterranean area, which might possibly have contributed to the Corsican gene pool (Calafell *et al.*, 1996). However, although Corsica had contacts with many other populations, their contribution to the Corsican gene pool was probably limited (Memmi *et al.*, 1998). Some say that the Corsicans were born in the 11th century, when some Italians migrated to Corsica and so peopling it; the original Corsican inhabitants had died out due to epidemics, Saracen and Barbarian raids, invasions by Vandals and medieval fights (Caratini, 1995). From the 11th to the 13th century Corsica was dominated by the Pisan dynasty; then, from the 13th to the 18th century by the Genovese dynasty. Some historical and archival research shows that there were migrations between southern Corsica and northern Sardinia, mainly between the 14th and the 16th century (Maxia, 2002). From the 18th to the first half the 20th century, most immigrants came from the Tuscan Apennine region and were directed to Corsica, where they were used to do humble jobs, and in consequence some Tuscan people decided to settle in Corsica (Arrigoni, 2002). Today there are still surnames of Italian etymology (Beaucarnot, 1988; Fordant, 1999) that have remained unchanged because Corsica was annexed to France only in 1789, when surnames had already been fixed. Yet, there are some analogies between continental French populations and some Corsican municipalities. This is basically due to the presence of Corsican surnames in the Provence-Alpes-Côte d'Azur region from the end of the 19th century (D'Orsetti *et al.*, 2001), which confirms there were migrations from Corsica to continental France, rather than the opposite. Saint Mandrier sur Mer, for instance, is geographically near the port of Toulon, in continental France, but in the chart it appears among the municipalities of southern Corsica (Fig. 9). Moreover, at the end of the 19th century lots of Corsican people emigrated in order to improve their social and economic situation, which caused a sharp fall in the size of the local population. So, foreign farmers and workers made the re-peopling of the island possible between the First and the Second World War:

they came from Italy, Portugal and Spain first, and later from North Africa and the French colonies (Caratini, 1995). Another fact confirming that the structure of Corsican surnames changed is that there are no similarities between Corsican and Ligurian municipalities (Figs 7 and 9), although Corsica was dominated by the Genovese dynasty for five centuries. This is true also for Bonifacio and Calvi, the two Corsican coastal municipalities most linked to Genoa. Until the 16th century, for example, the inhabitants of Calvi were mainly from Genoa, but afterwards its demographic situation changed thanks to the arrival of newcomers from other parts of Corsica (Caratini, 1995), because Calvi became an important port-of-call. Changes in the Corsican demographic situation affected the structure of the local population anyway from the 19th century until very recent times. Therefore, genetic similarities between Corsica and Sardinia are the consequence of a remote gene flow (Vona *et al.*, 1995; Moral *et al.*, 1996), assuming some reciprocal exchanges during the late Neolithic period and the subsequent Bronze and Iron Ages (Falchi *et al.*, 2004), so these events affected the gene pool of the local inhabitants. Yet, this cannot be explained through studies on surnames, because they started to spread all over the island only at the end of the 16th century, when parish registrations became compulsory.

Surnames are useful to describe the relations among populations. The origins of surnames, although polyphyletic, can always identify the cultural area in which they originated and how they spread over a territory depending on displacements of people. Thanks to their high polymorphism and the possibility of identifying surnames quite easily, it is possible to carry out in-depth and detailed analyses that could not be done with only biological markers. At the same time, however, they reflect the temporal limit they are linked to: surnames spread throughout the people at the end of the 16th century, when strict rules of registration and transmission were established. Therefore surnames cannot explain remote events, but they are very useful to describe and explain recent micro-evolutionary phenomena.

Acknowledgments

The authors would like to thank Dr Mauro Maxia for his useful collaboration on historical and onomastics studies about Corsica and Sardinia; Silvo Tasso for his helpful bibliographical researches on Tuscan islands; Dr Ornella Foffano for the translation of some parts taken from French books; and Dr Alessandra Maggiora for the translation of this paper from Italian into English.

References

- Arrigoni, T. (2002) *Uomini dei boschi e della natura. Emigrazione stagionale dall'Appennino toscano alla Corsica (XVIII–XX secolo)*. Pacini, Ospedaletto (PI).
- Barbujani, G. & Sokal, R. R. (1990) Zones of sharp genetic change in Europe are also linguistic boundaries. *Proceedings of the National Academy of Sciences of the USA* **87**, 1816–1819.
- Barrai, I., Rodriguez-Larralde, A., Mamolini, E. & Scapoli, C. (1999) Isonymy and isolation by distance in Italy. *Human Biology* **71**, 947–961.
- Barsotti, G. & Lambertini, M. (1999) *Isola di Capraia*. Pacini, Ospedaletto (PI).

- Beaucarnot, J.-L.** (1988) *Les Noms de Famille et Leurs Secrets*. Laffont, Paris.
- Brunet, G., Darlu, P. & Zei, G.** (2001) *Le Patronym Histoire, Anthropologie, Société*. CNRS Editions, Paris.
- Calafell, F., Bertranpetit, J., Rendine, S., Cappello, N., Mercier, P., Amoros, J.-P. & Piazza, A.** (1996) Population history of Corsica: a linguistic and genetic analysis. *Annals of Human Biology* **23**, 237–251.
- Calò, C. M., Autuori, L., Di Gaetano, C., Latini, V., Mameli, G. E., Memmi, M., Varesi, L. & Vona, G.** (1998) The polymorphism of the APOB 3' VNTR in the populations of the three largest islands of the Western Mediterranean. *Anthropologischer Anzeiger* **56**, 227–238.
- Caratini, R.** (1995) *Histoire du Peuple Corse*. Critérion, Paris.
- Caravello, G. U. & Tasso M.** (2002a) Noms de famille comme allèles: analyse des distributions spatiales des noms en Corse. *Biométrie Humaine et Anthropologie* **20**, 247–259.
- Caravello, G. U. & Tasso, M.** (2002b) Use of surnames for a demo-ecological analysis: a study in southwest Sardinia. *American Journal of Human Biology* **14**, 391–397.
- Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A.** (1994) *The History and Geography of Human Genes*. Princeton University Press, Princeton, NJ.
- Chen, K.-H. & Cavalli-Sforza, L. L.** (1983) Surnames in Taiwan: interpretations based on geography and history. *Human Biology* **55**, 367–374.
- Colantonio, S. E., Lasker, G. W., Kaplan, B. A. & Fuster, V.** (2003) Use of surnames models in human populations biology. *Human Biology* **75**, 785–807.
- Cox, M. F. & Cox, M. A. A.** (2001) *Multidimensional Scaling*. Chapman and Hall, London.
- Crow, J. F.** (1996) Isonymy: a thirty year retrospective. *Rivista di Antropologia* **74**, 25–34.
- Davidson, M.** (1983) *Multidimensional Scaling*. Wiley, New York.
- De Felice, E.** (1978) *Dizionario dei Cognomi Italiani*. Mondadori, Milan.
- De Felice, E.** (1980) *I Cognomi Italiani*. Il Mulino, Bologna.
- Delogu, I.** (2003) *Carbonia. Storia di una città*. Tema, Cagliari.
- D'Orsetti, S., Paoli, C., Palfrey, V., Mergnac, M.-O., Soulatges, A., Belser, C., Mitrevski, A. & Robin, F.** (2001) *Les noms de famille en Provence-Alpes-Côte d'Azur*. Archives & Culture, Paris.
- Falchi, A., Vacca, L., Lopez Alomar, A., Esteban, E., Memmi, M., Varesi, L., Moral, P. & Vona, G.** (2004) Population variability in some genes involving the haemostatic system: data on the general population of Corsica (France), Sardinia and Sicily (Italy). *Genetics and Molecular Biology* **27**, 139–146.
- Fordant, L.** (1999) *Atlas des Noms de Famille en France*. Archives & Culture, Paris.
- Francalacci, P., Morelli, L., Underhill, P. A., Lillie, A. S., Passarino, G., Useli, A. et al.** (2003) Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. *American Journal of Physical Anthropology* **121**, 270–279.
- Kimura, M.** (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
- Kruskal, J. B. & Wish, M.** (1978) *Multidimensional Scaling*. Sage University Paper Series on Quantitative Applications in the Social Sciences, Beverly Hills, CA.
- Lambertini, M.** (2002) *Isola del Giglio*. Pacini, Ospedaletto (PI).
- Lambertini, M.** (2004) *Isola d'Elba*. Pacini, Ospedaletto (PI).
- Lucchetti, E., Siri, E. & Soliani, L.** (1996) Internal migrations and relationships between Italian regional populations: an analysis based on surname distributions. *Rivista di Antropologia* **74**, 171–85.
- Lucchetti, E. & Soliani, L.** (1989) Similarità tra popolazioni esaminate mediante i cognomi. *Rivista di Antropologia* **67**, 181–198.

- Lucchetti, E. & Soliani, L.** (2004) Génétique et démographie. In Caselli, G., Vallin, J. & Wunsch, G. (eds) *Démographie: Analyse et Synthèse*. Vol. VI *Population et Société*. Editions de l'Institut National d'Etudes Démographiques, Paris.
- Maxia, M.** (2002) *Dizionario dei Cognomi Sardo-corsi. Frequenze, Fonti, Etimologia*. Condaghes, Cagliari.
- Memmi, M., Moral, P., Calò, C. M., Autuori, L., Mameli, G. E., Succa, V., Varesi, L. & Vona, G.** (1998) Genetic structure of Southwestern Corsica (France). *American Journal of Human Biology* **10**, 567–577.
- Moral, P., Memmi, M., Varesi, L., Mameli, G. E., Succa, V., Gutierrez, B., Lutken, N. & Vona, G.** (1996) Study on the variability of seven genetic serum protein markers in Corsica (France). *Anthropologischer Anzeiger* **54**, 97–107.
- Morelli, L., Paoli, G. & Francalacci, P.** (2002) Surname analysis of the Corsican population reveals an agreement with geographical and linguistic structure. *Journal of Biosocial Science* **34**, 289–301.
- Morelli, L., Vona, G., Varesi, L., Memmi, M., Autuori, L. & Calò, C. M.** (1999) Finger dermatoglyphics in the Corsican population (France). *Anthropologischer Anzeiger* **57**, 339–347.
- Piazza, A., Rendine, S., Zei, G., Moroni, A. & Cavalli-Sforza, L. L.** (1987) Migration rates of human populations from surname distributions. *Nature* **329**, 714–716.
- Pittau, M.** (1990) *I Cognomi della Sardegna. Significato e Origine di 5000 Cognomi Indigeni*. Delfino, Sassari.
- Racheli, G.** (1982) *L'arcipelago de La Maddalena Nella Storia*. Vert Sardegna, Calasetta (CA).
- Schiffman, S., Reynolds, M. & Young, F.** (1981) *Introduction to Multidimensional Scaling*. Academic Press, New York.
- Sokal, R. R., Oden, N. L., Legendre, P., Fortin, M.-J., Kim, J., Thomson, B. A., Vaudor, A., Harding, R. M. & Barbujani, G.** (1990) Genetics and language in European populations. *American Naturalist* **135**, 157–175.
- Soliani, L., De Iasio, S., Lucchetti, E. & Conterio, F.** (1981) Frequenze alleliche nelle popolazioni regionali italiane. I. Le distribuzioni regionali dei sistemi AB0, Rh, MN, Kell e delle aptoglobine. *Antropologia Contemporanea* **4**, 283–297.
- Soliani, L., De Iasio, S., Lucchetti, E. & Conterio, F.** (1982) Frequenze alleliche nelle popolazioni regionali italiane. II. Distanze genetiche. *Antropologia Contemporanea* **5**, 207–217.
- Tofanelli, S., Taglioli, L., Varesi, L. & Paoli, G.** (2001) The STR-based genetic profile of the population from Corsica island (France). *Forensic Science International* **123**, 33–38.
- Vallebona, G.** (1988) *Carloforte. Storia di una Colonizzazione*. Edizioni Della Torre, Cagliari.
- Varesi, L., Memmi, M., Cristofari, M.-C., Mameli, G. E., Calò, C. M. & Vona, G.** (2000) Mitochondrial control-region sequence variation in Corsican population, France. *American Journal of Human Biology* **12**, 339–351.
- Varesi, L., Memmi, M., Moral, P., Mameli, G. E., Succa, V. & Vona, G.** (1996) La distribution de quatorze marqueurs génétiques dans la population de l'île de Corse (France). *Bulletins et Mémoires de la Société d'Anthropologie de Paris* **1–2**, 5–14.
- Vona, G.** (1997) The peopling of Sardinia (Italy): history and effects. *International Journal of Anthropology* **12**, 71–87.
- Vona, G., Calò, C. M., Lucia, G., Mameli, G. E., Succa, V., Esteban, E. & Moral, P.** (1996) Genetics, geography, and culture: the population of S. Pietro island (Sardinia, Italy). *American Journal of Physical Anthropology* **100**, 461–471.
- Vona, G., Calò, C. M., Memmi, M., Vacca, L., Falchi, A., Ghiani, M. E., Moral, P. & Varesi, L.** (2002) Corsica and its genetic structure: analysis through classical genetic markers and

mitochondrial DNA. In *Environnement et Identité en Méditerranée – Actes du Congrès*. Université de Corse Pascal-Paoli, pp. 63–78.

- Vona, G., Memmi, M. R., Varesi, L., Mameli, G. E. & Succa, V.** (1995) A study of several genetic markers in the Corsican population (France). *Anthropologischer Anzeiger* **53**, 125–132.
- Vona, G., Moral, P., Memmi, M., Ghiani, M. E. & Varesi, L.** (2003) Genetic structure and affinities of the Corsican population (France): classical genetic markers analysis. *American Journal of Human Biology* **15**, 151–163.
- Zei, G., Guglielmino, C. R., Siri, E., Moroni, A. & Cavalli-Sforza, L. L.** (1983a) Surnames as neutral alleles: observations in Sardinia. *Human Biology* **55**, 357–365.
- Zei, G., Guglielmino, C. R., Siri, E., Moroni, A. & Cavalli-Sforza, L. L.** (1983b) Surnames in Sardinia. I. Fit of frequency distributions for neutral alleles and genetic population structure. *Annals of Human Genetics* **47**, 329–352.
- Zei, G., Piazza, A., Moroni, A. & Cavalli-Sforza, L. L.** (1986) Surnames in Sardinia. III. The spatial distribution of surnames for testing neutrality of genes. *Annals of Human Genetics* **50**, 169–180.