

## Linguistic, geographic and genetic isolation: a collaborative study of Italian populations

**Marco Capocasa<sup>2,3</sup>, Paolo Anagnostou<sup>1,2</sup>, Valeria Bachis<sup>4</sup>, Cinzia Battaglia<sup>1</sup>,  
Stefania Bertoncini<sup>5</sup>, Gianfranco Biondi<sup>6</sup>, Alessio Boattini<sup>7</sup>, Iliaria Boschi<sup>8</sup>,  
Francesca Brisighelli<sup>1,9§</sup>, Carla Maria Calò<sup>4</sup>, Marilisa Carta<sup>7</sup>, Valentina  
Coia<sup>10,11§</sup>, Laura Corrias<sup>4</sup>, Federica Crivellaro<sup>12</sup>, Sara De Fanti<sup>7</sup>, Valentina  
Dominici<sup>1,2</sup>, Gianmarco Ferri<sup>13</sup>, Paolo Francalacci<sup>14</sup>, Zeldia Alice Franceschi<sup>15</sup>,  
Donata Luiselli<sup>7</sup>, Laura Morelli<sup>14</sup>, Giorgio Paoli<sup>5</sup>, Olga Rickards<sup>16</sup>, Renato  
Robledo<sup>17</sup>, Daria Sanna<sup>14</sup>, Emanuele Sanna<sup>4</sup>, Stefania Sarno<sup>7</sup>, Luca Sineo<sup>18</sup>,  
Luca Taglioli<sup>5</sup>, Giuseppe Tagarelli<sup>19</sup>, Sergio Tofanelli<sup>5</sup>, Giuseppe Vona<sup>4</sup>,  
Davide Pettener<sup>7</sup> & Giovanni Destro Bisol<sup>1,2</sup>**

1) *Sapienza Università di Roma, Dipartimento di Biologia Ambientale, Roma, Italy*

2) *Istituto Italiano di Antropologia, Roma, Italy*

3) *Sapienza Università di Roma, Dipartimento di Biologia e Biotecnologie “Charles Darwin”, Roma, Italy*

4) *Università di Cagliari, Dipartimento di Scienze della Vita e dell’Ambiente, Cagliari, Italy*

5) *Università di Pisa, Dipartimento di Biologia, Pisa, Italy*

6) *Università dell’Aquila, Dipartimento di Scienze Ambientali, L’Aquila, Italy*

7) *Università di Bologna, Dipartimento di Scienze Biologiche, Geologiche e Ambientali, Bologna, Italy*

8) *Università Cattolica del Sacro Cuore, Istituto di Medicina Legale, Roma, Italy*

9) *Universidade de Santiago de Compostela, Instituto de Ciencias Forenses, Santiago de Compostela, Spain*

10) *Università di Trento, Dipartimento di Filosofia, Storia e Beni Culturali, Trento, Italy*

11) *Accademia Europea di Bolzano, Istituto per le Mummie e l’Iceman, Bolzano, Italy*

12) *Sezione di Antropologia, Museo Nazionale Preistorico Etnografico “Luigi Pigorini”, Roma, Italy*

13) *Università di Modena e Reggio Emilia, Dipartimento di Medicina Diagnostica, Clinica e di Sanità  
Pubblica, Modena, Italy*

14) *Università di Sassari, Dipartimento di Scienze della Natura e del Territorio, Sassari, Italy*

15) *Università di Bologna, Dipartimento di Discipline Storiche, Antropologiche e Geografiche, Bologna, Italy*

16) *Università degli Studi di Roma “Tor Vergata”, Centro di Antropologia Molecolare per lo studio del  
DNA Antico, Dipartimento di Biologia, Roma, Italy*

17) *Università di Cagliari, Dipartimento di Scienze Biomediche, Cagliari, Italy*

18) *Università di Palermo, Dipartimento di Biologia Ambientale e Biodiversità, Palermo, Italy*

19) *Istituto di Scienze Neurologiche, CNR, Mangone, Cosenza, Italy*

§ *Present affiliation*

e-mail: marco.capocasa@uniroma1.it; davide.pettener@unibo.it; destrobisol@uniroma1.it

**Summary** - *The animal and plant biodiversity of the Italian territory is known to be one of the richest in the Mediterranean basin and Europe as a whole, but does the genetic diversity of extant human populations show a comparable pattern? According to a number of studies, the genetic structure of Italian populations retains the signatures of complex peopling processes which took place from the Paleolithic to modern era. Although the observed patterns highlight a remarkable degree of genetic heterogeneity, they do not, however, take into account an important source of variation. In fact, Italy is home to numerous ethno-linguistic minorities which have yet to be studied systematically. Due to their difference in geographical origin and demographic history, such groups not only signal the cultural and social diversity of our country, but they are also potential contributors to its bio-anthropological heterogeneity. To fill this gap, research groups from four Italian Universities (Bologna, Cagliari, Pisa and Roma Sapienza) started a collaborative study in 2007, which was funded by the Italian Ministry of Education, University and Research and received partial support by the Istituto Italiano di Antropologia. In this paper, we present an account of the results obtained in the course of this initiative. Four case-studies relative to linguistic minorities from the Eastern Alps, Sardinia, Apennines and Southern Italy are first described and discussed, focusing on their micro-evolutionary and anthropological implications. Thereafter, we present the results of a systematic analysis of the relations between linguistic, geographic and genetic isolation. Integrating the data obtained in the course of the long-term study with literature and unpublished results on Italian populations, we show that a combination of linguistic and geographic factors is probably responsible for the presence of the most robust signatures of genetic isolation. Finally, we evaluate the magnitude of the diversity of Italian populations in the European context. The human genetic diversity of our country was found to be greater than observed throughout the continent at short (0-200 km) and intermediate (700-800km) distances, and accounted for most of the highest values of genetic distances observed at all geographic ranges. Interestingly, an important contribution to this pattern comes from the “linguistic islands” (e.g. German speaking groups of Sappada and Luserna from the Eastern Italian Alps), further proof of the importance of considering social and cultural factors when studying human genetic variation.*

**Keywords** – *Genetic structure, Linguistic diversity, Minority languages, Linguistic islands.*

## Introduction

The plant and animal biodiversity found on Italian territory is among the richest in the Mediterranean basin and in Europe as a whole. This depends on the presence of different biomes, with Alpine tundra and the Mediterranean arid zones at the extremes, whose variety may be seen as a consequence of the natural role of Italy as a bridge connecting central European and North African environments (Blasi *et al.*, 2005). As has happened throughout the continent, biodiversity has been influenced by human activities, such as cattle breeding and the agricultural exploitation of the lands since the Neolithic (Goudie, 2013). However, the remarkable ethno-cultural

diversity of human populations which have settled on Italian territory since prehistory has led to different ways of managing natural resources (Padovani *et al.*, 2009). This has probably reduced the loss of biodiversity produced by man-driven modifications of natural ecosystems to some extent. The latter points introduce us to another important aspect, which is more directly concerned with the anthropological dimension, i.e. biodiversity of human populations.

Looking at biodiversity from a holistic perspective, it is worth raising the question: does the great variety observed in Italy for plants and animals hold for human populations? Since prehistory, numerous peopling events have occurred on Italian territory. During the Last

Glacial Maximum, most of the country provided a refugium not only for animal and plant species (e.g. see Taberlet *et al.*, 1998; Petit *et al.*, 2003; Grassi *et al.*, 2009), but also for human groups (Banks *et al.*, 2008). Migrations during the Late Paleolithic, the Neolithic and, even more so, the Metal Ages characterized the complex peopling process of pre-historic Italy, leaving more (e.g. Greeks; see Scheidel, 2003) or less (e.g. Etruscans; see Barker & Rasmussen, 1998; Beekes, 2003) clear signatures. Under Roman rule, there were heavy demographic reshuffles caused by warfare and slavery and this had a significant impact on population composition (Hin, 2013). In more recent times, further contribution to the population heterogeneity came from invasions of the Barbarians (Lombards and Normans among others; see Jörg, 2002; Donald, 2008) and Arabs (Aghlabids; see Metcalfe, 2009). All these events explain at least part of the genetic structure of extant Italian populations, structured in two main geographic clusters (North-West and South-East) for Y-chromosome markers but more homogeneous for mtDNA polymorphisms (Boattini *et al.*, 2013). However, the investigations carried out so far have yet to provide an exhaustive picture of the genetic diversity of Italian populations, since they could not take into account another important component of the human biodiversity of our country. We are referring to the historic ethno-linguistic minorities which total about 5% of the population today. They differ according to area of origin (central and southeastern Europe) and demographic history, and the majority of them settled in their present locations in relatively recent times (from the Middle Ages to the 18th century). The linguistic isolation from neighboring populations and, in many cases, the settlement as small communities in secluded areas make such groups very important when evaluating the entire spectrum of biodiversity of Italian populations. At present, twelve ethno-linguistic minorities are officially safeguarded by Italian legislation: Albanian, Catalan, Croatian, French, Franco-provençal, Friulian, German, Greek, Ladin, Occitan, Slovene and Sardinian (Toso, 2008).

Furthermore, other minor linguistic groups are recognized at regional level (e.g. Tabarkian in Sardinia) and ancient linguistic substrates can be still recognized in marginal areas.

The interest in ethnic minorities and local communities is considerable in the present-day international debate concerning the preservation of cultural diversity and its influence on the biodiversity in the globalized world. Since 1999, the United Nations Environment Programme (UNEP) and the United Nations Educational, Scientific and Cultural Organization (UNESCO) have turned their attention to the “indigenous” peoples for the assessment of global biodiversity. UNEP focused on the relationships between biodiversity and cultural diversity with the publication of the report *Cultural and Spiritual Values of Biodiversity* (Addison Posey, 1999). After having defined *Cultural Diversity* as the “common heritage of humanity” (UNESCO, 2001), UNESCO has recently pointed out to the risk of extinction of 2500 languages, among which most of the minority languages spoken in Italy (Moseley, 2010). The message emerging from these initiatives is that paying attention to ethno-linguistic minorities is important not only for their contribution to the overall human biodiversity, but also for their role in maintaining cultural traditions which have an impact on biodiversity. This seems to expand the view, shared by many evolutionary Anthropologists, that biological and cultural diversity should be studied together to understand how the past has shaped the present of human species: “*The subjects are (on the one hand) data, and (on the other) the cultural history surrounding the collection and interpretation of those data*” (Marks, 1995).

In order to achieve a more complete characterization of the genetic structure of Italian populations, four research units of four Italian Universities (Bologna, Cagliari, Pisa and Roma Sapienza) initiated a collaborative research on linguistic minorities in 2007. During the realization of the project, the working group started further collaborations with researchers from other Italian Universities. A first part of work was done in the framework of the project “Isolating the isolates:

geographic and cultural factors of human genetic variation” supported by the Italian Ministry of Education, University and Research (PRIN project 2007 and 2009). In this initiative, our aim was to understand the relationships between cultural and geographical factors and the genetic structure of the Italian isolates (Destro Bisol *et al.*, 2008). The research work continued under the umbrella of the project “Human biodiversity in Italy: micro-evolutionary patterns” (PRIN project 2009-2011). In this second step, we also focused on the role of social structures in shaping the human biodiversity of isolated groups<sup>1</sup>. Further support to the whole initiative was provided by the *Istituto Italiano di Antropologia* (project “Bio-cultural Atlas of Italian populations”).

As genetic tools, we used mtDNA and Y chromosome polymorphisms, a choice based on their power to detect microevolutionary events on a wider timescale than bi-parentally transmitted polymorphisms, their relatively low costs, abundance of comparative data both at micro- and macro-geographic scale and the possibility to estimate gender-biased processes. Notably, the outputs of the study are not limited to scientific papers (e.g. Destro Bisol *et al.*, 2008; Robledo *et al.*, 2009; Boattini *et al.*, 2010; Bertoni *et al.*, 2012; Capocasa *et al.*, 2013b; see Supplementary File S1 for a complete list of papers and communications to congress), but also include the return of scientific results to the investigated communities (e.g. Capocasa *et al.*, 2012, 2013a) and a dedicated genetic online database (<http://www.isita-org.com/Anthro-Digit/italianisolates/index.html>).

In this paper, we present an account of the results obtained by the research units (RUs) involved in the collaborative study. To this aim, four case studies spanning from the eastern Alps to southern Italy are described and discussed, with a focus on their micro-evolutionary and anthropological implications. Thereafter, we present the results of a systematic analysis of the relations between linguistic, geographic and genetic

isolation combining the data obtained in the course of the long-term study with literature and unpublished results. More in particular, we: (i) analyze the genetic structure of numerous Italian populations in order to evaluate the association of linguistic and geographic factors with genetic isolation; (ii) compare the extent of genetic diversity observed in Italy and Europe taking into account both open populations and geographically and/or linguistically isolated groups.

## Materials and methods

### *The dataset*

Our original dataset includes data from 873 individuals for mtDNA and 795 individuals for Y chromosome from 19 populations (see Table 1) studied in the course of the PRIN projects. It was integrated with other published and unpublished results, reaching a total of 2875 and 1811 entries for mtDNA and Y chromosome polymorphisms, respectively (see Appendix 1; Congiu *et al.*, 2012).

Buccal swabs were collected in apparently healthy donors (see Tab. 1). A detailed description of laboratory analyses for DNA extraction, amplification, purification and genotyping is reported in Supplementary File S2.

### *Statistical analyses*

Haplotype diversity (HD) and its standard errors were calculated according to Nei 1987. Pairwise differences among all the populations of the datasets were calculated using the genetic distance measure  $F_{st}$  (Reynolds *et al.*, 1983; Slatkin, 1995). Analyses of molecular variance (AMOVA) were performed in order to test genetic differences among Sardinian populations. Demographic descriptive indices (Tajima's  $D$ , Fu's  $F_s$  and Harpending's raggedness) were calculated to test for the presence of signs of demographic expansion (Tajima, 1989; Harpending, 1994; Fu, 1997). An analysis of haplotype sharing was performed in order to evaluate the specific contribution of private haplotypes to the observed patterns of genetic differentiation among populations.

<sup>1</sup> As a rule, by the term “isolated populations” we mean linguistic and/or geographic isolates.

**Tab. 1. Populations analyzed in the course of the collaborative study. Abbreviations: ALT, altitude; FS, founder surnames; G, geographic isolate; GL, geographic/linguistic isolate; GP, grandparents; L, linguistic isolate; O, open population.**

POPULATION	ABB.	STATUS	SAMPLING STRATEGY	CENSUS SIZE*	ALT.°	MTDNA		Y CHROMOSOME	
						N	REFERENCE	N	REFERENCE
Arbereshe (Calabria)	ARC	GL	FS	28034	535	87	This study	87	Boattini et al., 2011; Pettener, pers. com.
Arbereshe (Sicily)	ARS	L	FS	7875	645	40	This study	42	Pettener, pers. com.
Benetutti	BEN	G	FS	2010	406	50	This study	45	Calò et al., 2013
Cadore	CAD	O	GP	10797	984	32	Caglià, pers. com.	52	Caglià, pers. com.
Carloforte	CFT	GL	FS/GP	6420	10	51	Calò et al., 2012	41	Robledo et al., 2012
Cimbrians (Lessinia)	LES	GL	GP	13455	758	40	Capocasa et al., 2013	29	Coia et al., 2013
Cimbrians (Luserna)	LUS	GL	GP	286	1333	21	Coia et al., 2012	25	Coia et al., 2013
Circello	CIR	G	FS	2501	700	27	This study	34	Tofanelli, pers. com.
Cosenza	COS	O	GP	69131	238	42	Pettener, pers. com.	28	Pettener, pers. com.
Fiemme Valley	FIE	G	GP	18990	1033	41	Coia et al., 2012	41	Coia et al., 2013
Ladins (Fassa Valley)	LVF	GL	GP	9894	1345	47	Coia et al., 2012	47	Coia et al., 2013
Lucca plain	PIL	O	GP	154928	81	50	This study	50	This study
North Sardinia	NSA	O	GP	67253	440	40	This study	47	Calò et al., 2013
Sappada	SAP	GL	GP	1307	1217	59	Capocasa et al., 2013b	36	Coia et al., 2013
Sauris	SAU	GL	GP	429	1212	48	Capocasa et al., 2013b	29	Coia et al., 2013
Sulcis-Iglesiente	SGL	O	GP	128614	96	50	Robledo et al., 2012	46	Robledo et al., 2012
Timau	TIM	GL	GP	500	830	46	Capocasa et al., 2013b	22	Coia et al., 2013
Trapani-Enna	TEN	O	GP	96834	-	80	Pettener, pers. com.	71	Pettener, pers. com.
Vagli	VAG	G	FS	995	575	22	This study	23	Tofanelli pers. com.

\* Source: ISTAT (2011) <http://demo.istat.it>.

° meters above sea level.

All parameters of intra- and inter-population genetic diversity were calculated using the Arlequin 3.5 software (Excoffier & Lischer, 2010). Multidimensional scaling (MDS) was applied to genetic distance matrices to visualize genetic differentiation among populations using the SPSS software (release 16.0.1 for Windows, S.P.S.S. Inc.). In order to evaluate the effect of census size on the intra- and inter-population diversity, we performed a stepwise multiple regression analysis using the current census size and altitude as covariates. Census size was log-transformed in order to linearize its relation to HD and average Fst. Unless otherwise stated, analyses were performed using mtDNA HVR-1 region (from 16033 to 16365 n.p.) and 15 Y chromosome STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 and GATA H4.1).

Genetic relationships among Arbereshe, putative source populations (Albania, Greece) and sink populations (Calabria, Sicily) were analysed as follows: populations were first grouped in clusters using a non-hierarchical EM algorithm (mclust R package; see Fraley & Raftery, 2002, 2006), then posterior membership probabilities (for each population and for each cluster) were calculated using Discriminant Analysis of Principal Components (DAPC, adegenet R package; see Jombart *et al.*, 2010) and graphically represented with a bar-plot.

To assess which interval of stepwise mutational differences between the male specific region of Y chromosome (MSY) haplotypes is the most suitable to represent the *Ligures* contribution to Samnium, we used the Equation 31 described by Walsh (2001) implemented in the software ASHes 1.1 (ashes.codeplex.com, Tofaneli *et al.*, 2011). The Bayesian posterior distribution of expected pairwise mutation differences was calculated for haplotypes separated by 78 generations (2,190 years ago using 28 years per generation), assuming a strict stepwise mutational model, a mutation rate of  $3.09 \times 10^{-3}$  per locus per generation (averaged values from observed single-locus germ-line mutations following Ballantyne *et al.*, 2010) and a lambda value of 0.0002.

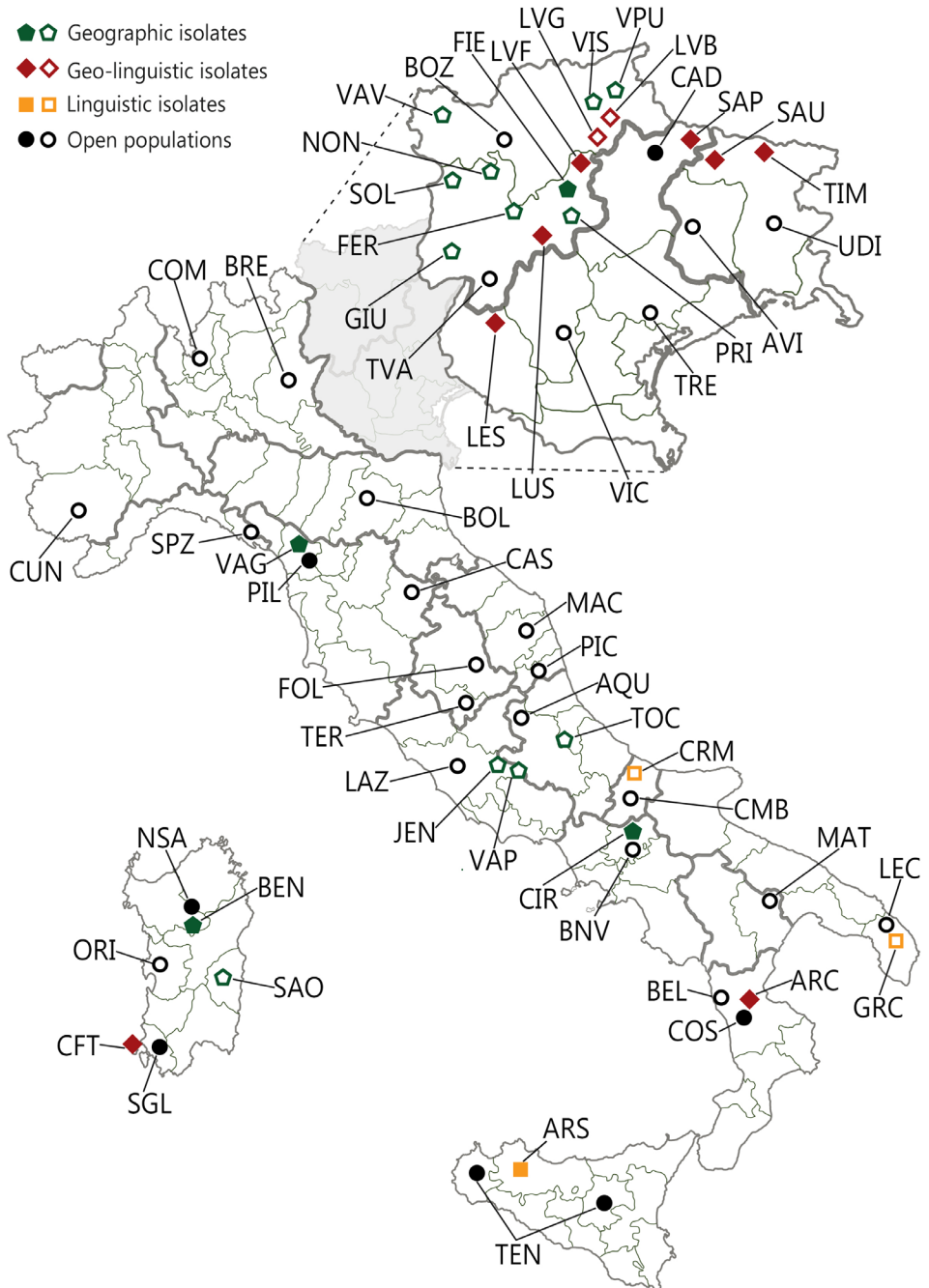
## Genes, geography and language: a tale of four stories

The collaborative project covered a substantial part of the linguistic and/or geographic isolates settled on Italian territory. What follows is a selection of the results obtained by each of the four RUs, which illustrates the variety of factors to be taken into account when studying human isolated populations and the diversity of approaches needed to carry out these investigations.

### *The Arbereshe: between Italy and the Balkans*

The Arbereshe are one of the largest linguistic minorities settled in Italy (~100,000 individuals; see Fig. 1). They originate from migratory waves from Albania between the end of the 15th and beginning of the 16th century in response to the invasion of the Balkans by the Ottoman Empire. Historians agree that most of the immigrants came from the south of Albania (Toskeria), often passing along the Peloponnese peninsula (Zangari, 1941). This hypothesis is supported by evidence that Arberisht, the language spoken by the Arbereshe, is part of the Tosk dialect group originally spoken in Toskeria. At present, there are 50 Arbereshe communities in Italy. They are located in Southern Italy, where they are separated from each other and interspersed among Italian villages. As such, they constitute an interesting case study allowing us to explore the effects of cultural isolation on genetic variability, as well as their relationships with the source (Southern Balkans) and sink populations (Southern Italy).

The Arbereshe of Calabria are probably the most important and interesting group, due to the high number of villages (30, with a total of ~60,000 individuals) and an exceptional preservation of their cultural identity. The latter feature is evident not only in the language (Arberisht), but also by the Greek Orthodox religion and a common mytho-history (Fiorini *et al.*, 2007). The investigated populations are scattered around the Pollino massif, an area that was historically characterised by strong geographic isolation and an economy based on sheep breeding and crop cultivation. During



**Fig. 1 - Geographic location of the populations analysed in this study. Filled and empty symbols indicate populations studied in the PRIN projects and literature data, respectively.**

the second half of the 20th century, the whole area was heavily struck by emigration which led to a partial loss of original cultural traits (break-down of isolates) (Fiorini *et al.*, 2007).

In Sicily, the Arbereshe are currently found in only three municipalities of the Palermo province (Contessa Entellina, Piana degli Albanesi and Santa Cristina Gela). Characterized by a relatively low population size (~15,000 individuals), they are settled in less geographically isolated areas than Calabrians. Their migration history is particularly complex, involving various intermediate steps (both in the Balkans and in Italy) before their definitive settlement. Furthermore, there is documented evidence of partial re-peopling events from Greece, the most well-known case being that of approximately 100 families coming from the island of Andros in 1520 AD (Giunta, 2003).

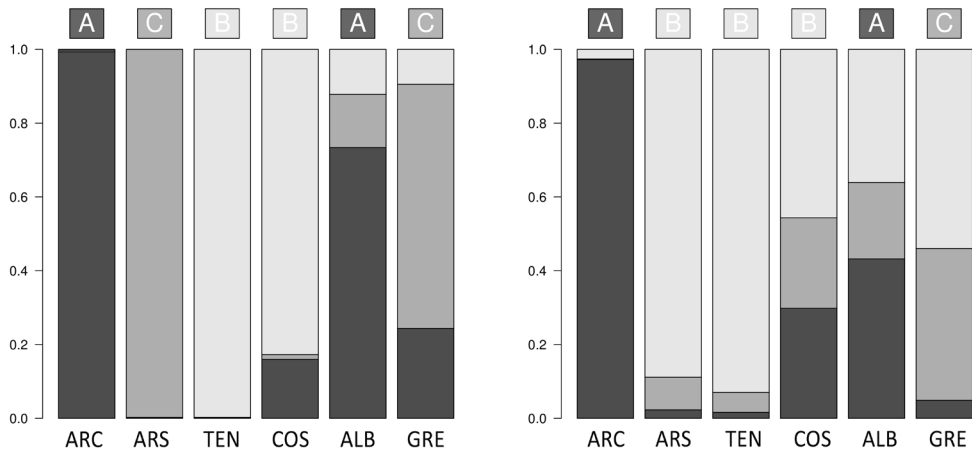
Calabrian Arbereshe were sampled in 13 different villages (Acquaformosa, Cerzeto, Civita, Firmo, Frascineto, Lungro, Plataci, S. Basile, S. Cosmo Albanese, S. Demetrio Corone, S. Giorgio Albanese, S. Sofia d'Epiro and Spezzano Albanese). Sicilian Arbereshe were sampled in two villages (Contessa Entellina, Piana degli Albanesi). In both cases, founder surnames were used as inclusion criterion (Boattini *et al.*, 2011) and "control" samples were collected from open neighbouring populations. All the sampled individuals have been unrelated for at least three generations. Data for Albanian and Greek populations were obtained from literature (Bosch *et al.*, 2006).

We first detected the genetic signatures of ethno-linguistic affiliation of the Arbereshe from Calabria by means of biodemographic analyses (Fiorini *et al.*, 2007; Tagarelli *et al.*, 2007). We showed that Arbereshe populations are clearly distinguishable from their Italian neighbours, and observed a correlation between geographic- and surname-based distances between the Arbereshe groups. In addition, Arbereshe groups from the Pollino area showed considerably higher marital isonymy levels than in Italians (0.080 and 0.050, respectively). All these lines of evidence suggest that Arbereshe of Calabria are a group of populations that have undergone cultural and

geographical isolation. Importantly, diachronic analyses revealed that such isolation features declined progressively during the last two centuries, and more rapidly in the last 50 years. A Y-chromosome study based on a first sampling of Calabrian Arbereshe (a total of 40 individuals) confirmed and extended these findings (Boattini *et al.*, 2011). Arbereshe were found to provide a signal of discontinuity in the Italian genetic landscape, while showing a strong affinity with some modern Balkan populations (in particular Albanians and Kosovars).

Further insights are provided by taking into account unpublished results for Y-chromosomal and mitochondrial polymorphisms from a larger sampling of Calabrian Arbereshe (104 and 101 individuals, respectively) together with new data on Sicilian Arbereshe (44 and 57 individuals). Haplotype diversity values observed for male and female lineages in Calabrian Arbereshe (0.976 and 0.995, respectively) and Sicilian Arbereshe (0.979 and 0.992) are very close to those of Italian neighbour populations of Cosenza (0.994 and 1.000) and Trapani-Enna (0.985 and 0.999). These results are in line with our previous conclusions that Arbereshe have "conserved much of [their] ancestral genetic diversity along with [their] founder surnames and cultural features" (Fiorini *et al.*, 2007; Boattini *et al.*, 2011) and that their Y-chromosomal gene pool may mirror "the genetic structure of the migrants that came to Italy from the Southern Balkans (Albania) five centuries ago" (Boattini *et al.*, 2011). As shown in Fig. 2, Calabrian Arbereshe cluster (with high membership probabilities) with Albanians both for male and female lineages, as predicted by the isolation pattern inferred from surname analysis and their similarity with the source population. Intriguingly, our results suggest a different and more complex scenario for the Arbereshe from Sicily. Their Y-chromosome clustering pattern supports a significant admixture with Greeks, which agrees with historical information. On the other hand, mtDNA results highlight a substantial similarity to populations from Southern Italy, a likely consequence of admixture processes (and/or sexually imbalanced migration patterns).





**Fig. 2 - Admixture-like barplots for Y-chromosome (left) and mtDNA (right).** The barplots represent DAPC-based posterior membership probabilities for each of the considered populations and for each inferred cluster (mclust algorithm). The affiliation of each population to a given cluster and its corresponding colour code are represented by letters (within coloured squares) on the top of each bar. Labels: Arbereshe from Calabria (ARC), Arbereshe from Sicily (ARS), Trapani-Enna (TEN), Cosenza (COS), Albania (ALB), Greece (GRE).

These findings suggest a noticeable difference in the ethnogenesis of Calabrian and Sicilian Arbereshe, with a more evident continuity with source populations for the former and lower impact of isolation for the latter. The study of Albanian-speaking minorities of Southern Italy illustrates, therefore, how the genetic structure of populations sharing the same ethno-linguistic label might have been shaped by diverse micro-evolutionary histories.

#### *An analysis of isolated populations in Sardinia*

Populations from Sardinia provide a paradigmatic example of genetic differentiation in an insular context. The combined effect of small population size, endogamy, and the associated consanguinity, resulting from geographical and cultural barriers, has increased among-group diversity, mainly via long-term genetic drift. A further source of genetic heterogeneity was provided by the different incidence of malaria morbidity. Thus, the unique genetic structure and high level of isolation of some Sardinian villages create opportunities to map genes involved in multifactorial diseases (Zavattari *et al.*, 2000,

Angius *et al.*, 2001). However, the interplay between microevolutionary, demographic and cultural factors makes Sardinian populations of great interest also for anthropological studies.

During the collaborative project, the RU of the University of Cagliari focused on the isolates of Carloforte (linguistic) and Benetutti (geographic), in order to assess the degree of isolation and the relative contribution of cultural and geographical barriers (see Fig. 1). Carloforte (6420 inhabitants), located in the small island of San Pietro, off the Southwestern coast of Sardinia, is an alloglot community founded in 1738 AD by Ligurian migrants coming from Tabarka, a Tunisian island (Vallebona, 1974). As a historical imprint, the present-day Carloforte inhabitants speak the ancient dialect of Pegli (Liguria), also referred to as “tabarkino” (a minority language officially recognized by Sardinian regional legislation). Benetutti, a small village of 2010 inhabitants, is located in an area of Northern Sardinia, Goceano, which is characterized by strong geographic isolation. The first historical-demographic documents regarding Benetutti may be found in the “*Quinque Libri*”, dating back to 1618 AD.

By studying Y chromosome and mtDNA polymorphisms, we were able to detect signatures of genetic isolation and of a constant demographic state in both populations (Robledo *et al.*, 2012; Calò *et al.*, 2013; present study). In fact, haplotype diversity observed in Carloforte (0.975 and 0.921 for Y chromosome and mtDNA, respectively) and Benetutti (0.975 and 0.918 for Y chromosome and mtDNA, respectively) are lower than the ones observed in open Sardinian populations (Robledo *et al.*, 2012). In fact, using both founder surnames and grandparent rule as a selection criteria, we were able to select individuals descending from village founders and capture a large extent of within-population genetic diversity at the same time. A similar conclusion was reached in a different study on Y chromosome lineages in isolates located in the Italian Western Alps (Boattini *et al.*, 2010). In any case, the lack of a substantial drop in within-population diversity is consistent with historical records which do not point to any dramatic change in population size. This is also supported by demographic inferences based on mtDNA analysis. In fact, the negative but statistically insignificant values of Tajima's D and Fu's index together with values of Harpending's raggedness suggest that these populations have kept their sizes small and constant over time.

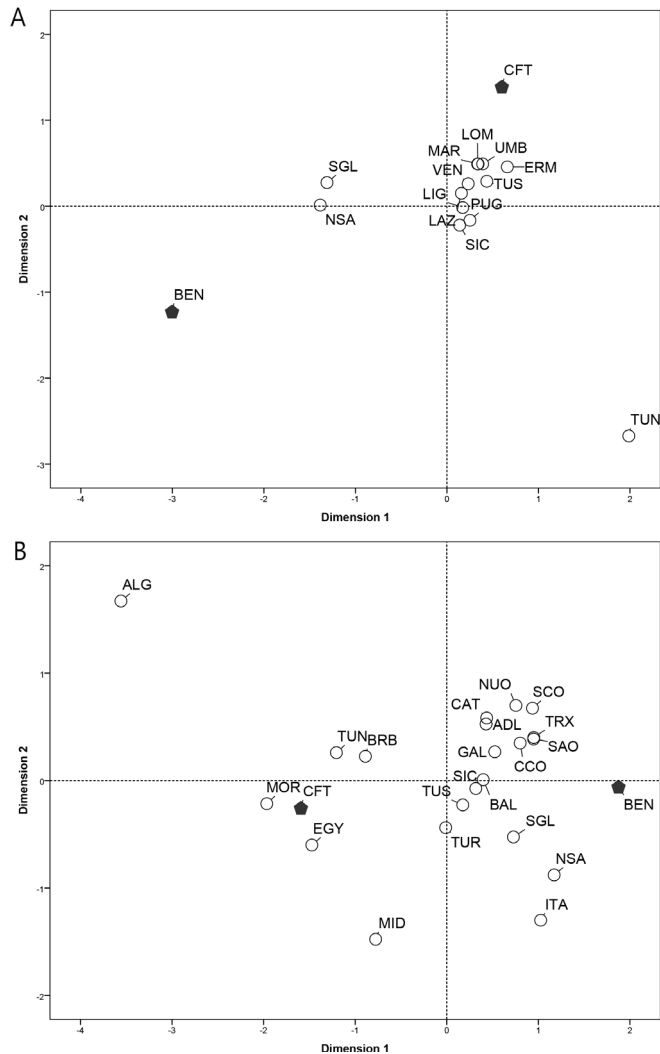
The analysis of inter-population variation proved useful to shed light on other aspects of the genetic history of the two populations under study. Effects of genetic isolation were detected with both uniparental markers, since genetic distances of Benetutti and Carloforte from neighbouring open populations were found to be statistically significant ( $p < 0.05$ ). In the multi-dimensional scaling (MDS) plot based on Y chromosome data (Fig. 3a), Carloforte lies in a rather isolated position, although not very distant from other Italian populations. Genetic differentiation between the sampled Ligurian population and the founders of Carloforte may account for the former evidence. Benetutti is rather separated from other groups, including Sardinians. The mtDNA based plot (Fig. 3b) confirms both the peculiar behaviour of the Carloforte, which is the only

Sardinian population located in the x-positive values of the plot, and the genetic differentiation of Benetutti from other Sardinians. However, differently from what we observed using paternal lineages, we found that Carloforte lies close to African populations, particularly Tunisia. Although no support comes from historical sources, a hypothesis of admixture of Carloforte males with Tunisian females is worth considering due to the occurrence at appreciable frequency (8%) of the mtDNA haplogroup M, characteristic of Northern African and Western Asian populations but so far undetected in Sardinia and Italy (Gonzalez *et al.*, 2007; Fraumene *et al.*, 2003, 2006; Morelli *et al.*, 2000).

An analysis of molecular variance among Sardinian populations was carried out to evaluate the apportionment of genetic diversity. Consistent with previous research work in Sardinia, only a small portion of variation could be attributed to differences between populations (3.78% for mtDNA and 3.31% for Y chromosome;  $p < 0.05$ ). Interestingly, data on preferential male mobility in the Carloforte (Calò & Vona, 1994) and, more in general, the historically documented practice of matrilocal behaviour of some areas of Sardinia (Murru Corrigan, 1995) might explain the slightly larger variance observed for maternal lineages.

#### *Genetic and cultural isolation in Northern Apennines*

Recent studies suggest an ethno-linguistic continuity between the human groups currently inhabiting the inner valleys of North-West Tuscany (Lunigiana, Garfagnana and orographically linked areas) and populations settled in this area in pre-Roman times. The genetic evidence of this connection was formerly reported by Piazza *et al.* (1988), who associated one of the main principal components obtained from blood protein alleles to the inheritance of those ancient tribes named *Ligures* according to the Latin etymology. This possible link is supported by recent studies of variation in the male-specific portion of the Y chromosome, which have identified North Western Tuscany as an area where lineages



**Fig. 3 - Multi-dimensional scaling plots of *F<sub>st</sub>* genetic distances among Sardinian and Mediterranean populations based on (A) 7 Y chromosome STRs (*DYS19*, *DYS389I*, *DYS389II*, *DYS390*, *DYS391*, *DYS392*, and *DYS393*; stress value=0.046) and (B) mtDNA HVR-1 sequences (stress value=0.130). Abbreviations and references: (a) Benetutti, BEN; North Sardinia, NSA (Calò et al., 2013). Carloforte, CFT; Sulcis-Iglesiente, SGL (Robledo et al., 2012). Emilia Romagna, ERM (Ferri et al., 2009). Lazio, LAZ; Liguria, LIG; Puglia, PUG; Toscana, TUS; Umbria, UMB (Presciuttini et al., 2001). Lombardia, LOM (Cerri et al., 2005). Marche, MAR (Onofri et al., 2007). Sicilia, SIC (Ghiani et al., 2004); Tunisia, TUN (Frigi et al., 2006). Veneto, VEN (Turrina et al., 2006). (b) Algeria, ALG; Catalogna, CAT (Cortez-Real et al., 1996). Andalusia, ADL; Balearic Islands, BAL; Gallura, GAL; Morocco, MOR; Nuorese, NUO; South Corsica, SCO; Trexenta, TRX (Falchi et al., 2006). Benetutti, BEN, North Sardinia, NSA (this study). Carloforte, CFT (Calò et al., 2012). Central Corsica, CCO (Varesi et al., 2000). Egypt, EGY (Krings et al., 1999). Middle East, MID (Di Rienzo & Wilson, 1991). North Italy, ITA (Stenico et al., 1996). Ogliastra, SAO (Fraumene et al., 2006). Sicilia, SIC (Vona et al., 2001). Sulcis-Iglesiente, SGL (Robledo et al., 2012). Toscana, TUS (Francalacci et al., 1996). Tunisia, TUN (Plaza et al., 2003). Tunisian Berbers, BRB (Fadhlaoui-Zid et al., 2004). Turchia, TUR (Comas et al., 1996).**

with a putative post-Neolithic date of origin persist at relatively high frequencies (Busby *et al.*, 2012, Boattini *et al.*, 2013).

Modern historical sources consider *Ligures Apuani* to be not entirely eradicated by the Roman military campaigns of the II century BC and the following mass deportations to *Ager Taurasinum* (Samnium) of at least 47,000 people (Titus Livius, In: Storia Romana XXXVIII). Rather, they depict the fate of natives in the form of confederated tribes surviving in hilly or mountainous areas, until they were eventually absorbed in the demo-territorial institutes later imposed by the Roman authorities (i.e. *res publicae, arcifinalis, excepta*), often maintaining a certain degree of autonomy and ownership (Dilke, 1971; Gagliardi, 2006; Marcuccetti, 2012).

From a linguistic point of view, many hamlets remained in the Apuan Alps and the Northern Apennines show similar forms of resistance to the “*tuscanization*” of the spoken language, maintaining a sort of bilingualism exerted through the use of lexemes and phonemes which are traceable to a pre-Latin matrix (Ambrosi, 1956; Giacomelli, 1975; ALT-Web). Moreover, the distribution of many toponyms with pre-Latin roots, in particular those referring to general orographic names such as rivers, mountain reliefs, open places (i.e. *nava, asc, var, dur, mat, penn, aus, kar*), is limited over Europe to conservative areas and have a peak of density in the territory once inhabited by *Ligures* tribes (Marcuccetti, 2008).

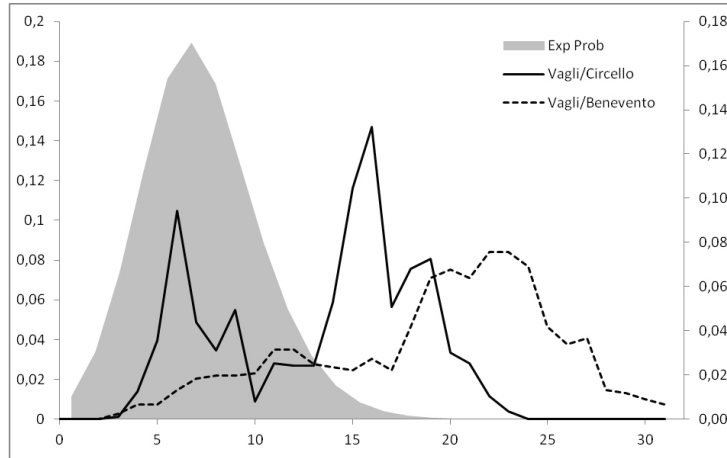
The hypothesis of a genetic legacy between *Ligures Apuani* and present Apuan and Samnite isolates (Ligures Legacy Model, LLM) was tested by the RU of the University of Pisa through the genetic characterization of uni-linear markers in a sample of unrelated donors from the communities of Vagli (Province of Lucca) and Circello (Province of Benevento) selected according to founder surnames. Vagli is located in the core of the area which has many archeological records linked to *Ligures Apuani*. Its elder inhabitants still speak a language characterized by a number of linguistic relicts (Ambrosi, 1956; Guazzelli, 2001). Circello lays in close proximity to the remains of Macchia, the town in the Samnium

around which the deported *Ligures* (*L. Baebiani*) were forced to settle in 180 BC (Patterson, 2009).

A slight reduction of HD at both mtDNA and Y Chromosome has been observed for Vagli (0.948 and 0.984) and Circello (0.960 and 0.975) relative to the neighbouring populations of Piana di Lucca (0.983 and 0.999) and Benevento (0.989 and 1.000). As expected for communities of Indo-European ancestry, usually practicing prevalence of female *vs* male mobility (patrilocal), genetic distances based on mtDNA are weaker discriminators than distances based on MSY haplotypes (see Supplementary Tab. S1). Preliminary comparative assessments of MSY profiles suggest that the diversity of Apuans might be due to an excess of R-U152 haplotypes, whose diffusion in Italy is thought to coincide with the diffusion of *Ligures* cultural features in the Middle-Late Bronze Age (Bertoncini *et al.*, 2012).

As a whole, the two communities under study (Vagli and Circello) showed a genetic pattern which is compatible with a long history of isolation but also with quite diverging micro-evolutionary histories after contacts implied by the LLM. As a more direct test of a genetic continuum with *Ligures* tribes (Fig. 4), we assessed whether an enrichment of matches compatible with the Titus Livius deportation hypothesis is detectable when comparing any MSY haplotype of the local population (Vagli) with haplotypes of both, the putative displaced (Circello) and the open Samnite population (Benevento). The enrichment of LLM-compatible matches in the Vagli-Circello curve totaled about 80% and was extremely statistically significant (Fisher exact test,  $p < 0.0001$ ).

The hypothesis of direct descent of the resident males in the Apuan and in the Circello area from members of *Ligures* tribes, who escaped deportation, cannot be ruled out. Further data from wider samples and haplogroup diagnostic markers, as well as more extensive simulation analyses will help achieve more robust inferences. Nonetheless, our case study shows that even mild geographical and cultural isolation may lead to the preservation of a long genetic thread connecting present populations to ancient layers of



**Fig. 4 - Distribution of 17-locus Y-STR haplotype pairwise mutational differences between Vagli and Circello (solid line) and between Vagli and Benevento (dotted line). The shaded area is the Bayesian posterior distribution of pairwise mutation differences compatible with the LLM hypothesis.**

pre-Roman Italy. As a corollary, it suggests that many isolated Italian communities other than Northern Apennine ones may escape simplistic classification schemes (i.e. linguistic *vs* geographic isolate), owing to the gradual fringing and recent oblivion of a common ancient cultural identity. Finally, our study highlights the usefulness of accurate non random genetic samplings to uncover genetic layers obscured by recent reshufflings within and among human populations.

#### *Genetic and cultural isolation in the Eastern Italian Alps*

Despite the presence of several linguistic islands, the Eastern Alps had not been thoroughly investigated from a bio-anthropological perspective before the start of the collaborative project. The RU from the University of Rome “La Sapienza” focused on six populations: Ladins from the Fassa Valley, Cimbrians from Lessinia and Luserna and the ethno-linguistic “islands” of Sappada, Sauris and Timau (see Fig. 1). Dolomitic Ladins probably originated from a founder group of pre-Indoeuropean speakers that underwent a process of “Latinization” under the pressure of the Roman Empire. At the beginning of the Middle Ages, they fragmented as a

result of the arrival of German people. The Ladin community of the Fassa Valley consists of approximately 6,000 individuals settled in seven villages (Toso, 2008). The other populations under study are German-speaking groups thought to be in continuity with peoples migrated from Carinthia, Bavaria and Tyrol in the high and late Middle Ages (Denison, 1971; Rizzi, 1993). After their initial settlement, these ethno-linguistic isolates maintained close social and cultural links (De Concini, 1997; Navarra, 2002). At present, their common background is evident due to their sharing of cultural aspects, such as language and traditions, which persist despite a certain degree of cultural exchanges with the surrounding neo-latin groups. The two Cimbrian settlements have different social and environmental characteristics. Luserna is a small town located over 1,300 meters a.s.l. in the Trentino region, and inhabited by approximately 300 people. Conversely, Lessinia is populated by over 13,000 inhabitants who are distributed in a wider and less “hostile” mountainous territory in the province of Verona in the Veneto region. Sappada is located at an altitude of 1,245 m.a.s.l. in the Veneto region (province of Belluno; 1,307 inhabitants). Sauris and Timau are two small

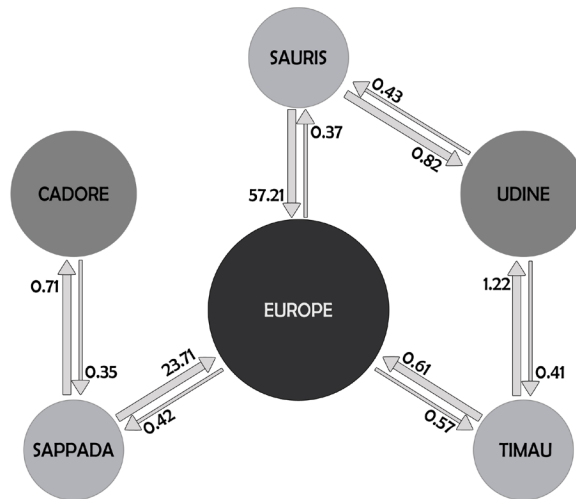
villages (429 and 500 inhabitants, respectively) located in the Carnic Alps. The former is placed at 1,212 m.a.s.l. within the province of Udine, in the Friuli Venezia Giulia region. Timau is situated at 830 m.a.s.l. in the But valley.

The analysis of genetic variation suggests a certain degree of genetic isolation from surrounding populations for Luserna, Sappada, Sauris and Timau. In fact, in accordance with the James Neel's statement that genetic isolates are "derived from a relatively small population sample, which then slowly expand, with very little recruitment from outside the group" (Neel, 1992), we observed low haplotype diversity values and lack of any signal of demographic expansion (Fu's  $F_s$  and Harpending's raggedness) (Coia *et al.*, 2012, 2013; Capocasa *et al.*, 2013b). Analyses of genetic distances highlight the differentiation among the four communities and their diversity from other European populations. Furthermore, genetic differentiation among Alpine populations was detected even at individual level using a Bayesian method to cluster multilocus genotypes based on 15 autosomal microsatellites (Montinaro *et al.*, 2012). A different pattern was detected for Ladins from Fassa and Cimbrians from Lessinia, who showed neither an evident reduction of mtDNA and Y chromosome intra-population diversity nor a significant departure from the European genetic background.

Although patterns of genetic variation and the lack of signature of demographic expansion observed in Sappada, Sauris and Timau are compatible with what is to be expected in human isolates, it is possible that these results may simply reflect their small initial effective size. As a more direct test of genetic isolation, we performed an analysis of gene flow using a Bayesian approach, applying an Isolation with Migration model (Hey & Nielsen, 2007). Through this approach, we inferred a reduced incoming gene flow for Sappada, Sauris and Timau, both from a wide Central Western European and a neighbouring open Italian-speaking population (Fig. 5). Therefore, our results support the hypothesis that the peculiar linguistic and geographical factors acting on the communities under study

might have determined a substantial degree of genetic isolation, so shaping their genetic structure (Capocasa *et al.*, 2013b).

Another aspect which we noticed was the extreme genetic differentiation among Sappada, Sauris and Timau, an unexpected finding considering their relative geographic proximity and closeness in terms of language and traditions (De Concini, 1997; Navarra, 2002). In order to evaluate the magnitude of the genetic differentiation among the three linguistic islands, we first compared them with other European language minorities (Cimbrians and Ladins from Eastern Alps, Aromuns from Albania and Macedonia) whose members are settled in geographically close but distinct locations. Interestingly, Sappada, Sauris and Timau showed inter-population distance values which were two (Ychromosome) or three (mtDNA) times higher than in other groups (Capocasa *et al.*, 2013b). We then wondered whether cultural factors might help explain this pattern. This seems to be worth testing since members of the Eastern Alps linguistic islands do not self-identify as belonging to the same community despite their cultural homogeneity (Steinicke *et al.*, 2011). Such behaviour seems to be in sharp contrast with the sense of belonging of Cimbrians, Ladins and Aromuns and their respective ethnic groups. Therefore, we went on by hypothesizing that this "local ethnicity" may have played a role in marriage strategies, decreasing the genetic exchange among the three linguistic islands. Carrying out coalescent simulations, we inferred that the combined effect of a small initial effective size, as indicated by historical sources (Petris, 1980; Brunettin, 1998; Peratoner, 2002), and a substantial reduction of gene flow, as implied by a local ethnicity model, can explain the pattern observed for Sappada, Sauris and Timau communities (Capocasa *et al.*, 2013b). More in general, our case study shows that complementing classical measures of genetic diversity with Bayesian estimates of gene flow and simulations of micro-evolutionary models may help us better understand patterns of genetic isolation and its relations with demographic and cultural factors in human populations.



**Fig. 5 - Estimates of gene flow among German-speaking Eastern Alps linguistic islands, a neighbouring population and an European metapopulation based on mtDNA HVR-1 region (values reported in effective number of migrating haplotypes).**

### Linguistic, geographic and genetic isolation in Italian populations

Studies of cultural and geographical factors of genetic isolation have been so far mainly concerned with single or a limited number of groups, focusing on specific populational contexts (e.g. Jeran *et al.*, 2009; van Oven *et al.*, 2011; Veeramah *et al.*, 2011). Combining data relative to numerous isolated groups, produced in the course of this collaborative project, with literature and unpublished results gives us the opportunity to draw inferences on the relations between culture, geography and genetics on a larger geographic scale. To this aim, we built large mtDNA and Y Chromosome databases (see Tab. 1 and Appendix 1), including 57 populations (29 isolates and 28 open groups) and 46 populations (22 isolates and 24 open), respectively. In order to reach the maximum coverage of the Italian territory/populations, we decided to include geographic and/or linguistic isolates studied with different sampling schemes.

As a preliminary step, we compared the genetic structure of isolated and open populations,

without considering any further subdivision. We decided to focus on HD and  $F_{st}$  values, since these parameters are little influenced by differences in sample size among those commonly used to detect genetic isolation (Anagnostou *et al.*, in preparation). The HD of isolated populations was found to be significantly lower (Mann-Whitney test  $p < 0.001$ ) than those of open groups, both for mtDNA (from 0.778 to 0.990 vs 0.903-0.995) and Y Chromosome (from 0.590 to 0.998 vs 0.993-1.000) (see Supplementary Tab. S2 and S3; Supplementary Fig. S1). Most of the highest average genetic distances from open populations were observed in geographic and geographic/linguistic isolates, which is another expected effect of genetic isolation (Fig. 6; Supplementary Fig. S2). The difference between open and isolated population groups was slightly more evident for Y chromosome (0.011-0.322 for isolates vs 0.004-0.046 for open populations) than mtDNA (0.000-0.085 vs 0.000-0.013), but highly statistically significant for both polymorphisms (Mann-Whitney test  $p < 0.001$ ). Private haplotypes do not contribute in any significant way to

these patterns (Supplementary Fig. S3), a finding which is consistent with the relatively recent age of most isolated populations under study.

On the whole, these results suggest that populations subject to geographic and linguistic isolation are more likely to show lower HD and higher average  $F_{st}$  than open groups. However, they do not clarify whether these differences may be actually regarded as an effect of genetic isolation and the relative weight of geographic and linguistic barriers. To explore these issues, we carried out further analyses in three steps.

First, we used the departure from the pattern of open populations as a criterion to distinguish between simple fluctuations of values and signatures of genetic isolation. To do this, we used the inter-quartile method to identify weak and strong outliers for HD and average  $F_{st}$  distributions. To be considered as bearing a signature of genetic isolation, a given population should display a departure falling in the “weak outlier” interval for both parameters. Furthermore, they were labelled as strong only when both reduction in HD and increase in average  $F_{st}$  exceeded the threshold delimiting the strong outlier area. Using these criteria, we identified 11 populations (out of 29 geographic and/or linguistic isolates) bearing signatures of genetic isolation (10 strong and 1 weak) for mtDNA. Results for Y chromosome do not markedly differ for the number of groups showing signatures of genetic isolation (8 out of 22), but, rather, for the proportion between strong (3) and weak (5) signals (see Fig. 7).

Second, we took into account the possible confounding effect of census sizes in order to validate the obtained inferences. This parameter was preferred to population density, for which no statistically significant correlation with either HD or average  $F_{st}$  was observed (data not shown). To this purpose, we carried out a stepwise multiple regression analysis on all our populations subject to isolation factors using census as a dependent variable. In these calculations, altitude was used as a control variable. This choice was based on two reasons: (i) it may be regarded as one of the ways in which geography may determine isolation between populations, rather than a confounding factor; (ii) we

observed a significant correlation between altitude and census size in our dataset (Pearson's  $\rho=-0.495$ ,  $p=9.201*10^{-5}$ ; Spearman's  $\rho=-0.546$ ,  $p=1.566*10^{-5}$ ), as already reported in previous studies (Cavalli Sforza & Bodmer, 1971; for the Italian context see Franceschi & Paoli, 1994; Morelli *et al.*, 2002; Capocasa *et al.*, in press).

We found a significant correlation between census size and average  $F_{st}$  for both polymorphisms. However, they account only for a minor portion of the total variance (10.5% for mtDNA and 31.8% for Y chromosome). Census size and HD were found to be significantly correlated only for Y chromosome, but again with a weak effect (26.5% of variance explained) (see Supplementary Tab. S4). Therefore, the effect of census size, although detectable, does not seem so marked to undermine our inferences. Altitude was found to have a low and significant, but yet negligible effect only on Y chromosome  $F_{st}$  distribution (10.5% of total variance), implying that other factors are responsible for the observed signals of genetic isolation.

Third and finally, we tried to understand whether the conditions of geographic isolation alone, or combined with linguistic isolation, may have different outcomes on the genetic structure of populations under study. Unfortunately, the presence of only 3 linguistic isolates do not make it possible to perform any test of the effect of linguistic barriers alone. In the mtDNA plot (Fig. 7), signatures of genetic isolation were observed more frequently in geographic/linguistic (6 out of 10; 60%) than in geographic isolates (5 out of 16; 31%). A comparable disproportion was shown by Y chromosome polymorphisms, with 2 geographic (out of 9; 22%) and 5 geographic/linguistic isolates (out of 10; 50%) showing the signal. Limiting the comparison to strong signatures (defined as above), the ratio of positives between geographic/linguistic and geographic isolates does not change substantially either for mtDNA (5 out of 10 *vs* 5 out of 16, respectively) or Y chromosome (3 out of 10 *vs* 0 out of 9). A further indication of the effect of linguistic isolation on the genetic structure of populations is provided by the demographic inferences based on mtDNA data.





In fact, four of the 6 groups which showed a lack of signals of demographic expansion (non significant  $F_u$ 's and high raggedness values) belong to the geographic /linguistic isolates (Cimbrians from Luserna, Sappada, Sauris, Timau), whereas only 2 are geographic isolates (Vallepietra and Val Pusteria) (Supplementary Tab. S2).

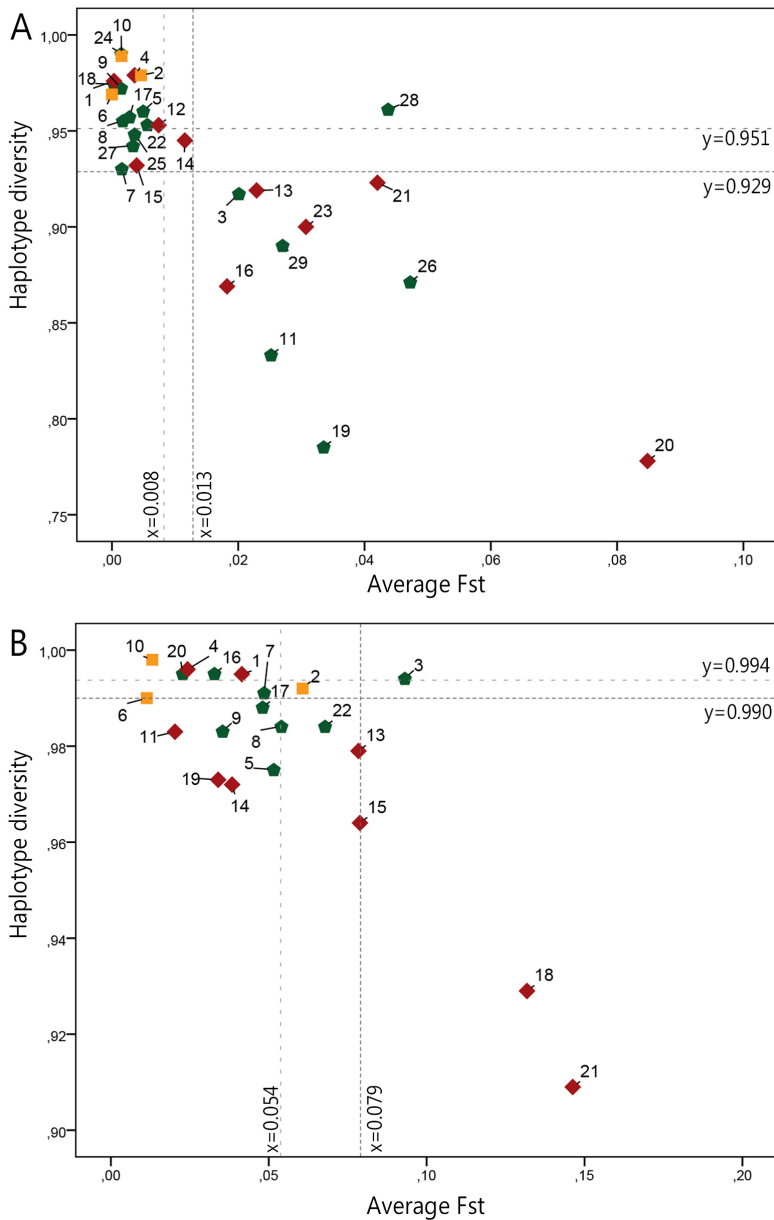
What do all these findings tell us? We admit that no simple cause-effect relationship between linguistic barriers and genetic structure can be inferred when studying populations with different ethnogenesis and demographic histories. Nonetheless, it seems clearly evident that the combination of linguistic and geographic isolation factors is frequently associated with the presence of robust signatures of genetic isolation and a condition of slow, if any, demographic expansion, fitting the classical definition of genetic isolate given by James Neel in 1992. This is not trivial since, to date, evidence of a joint effect of geographic and linguistic isolation factors in Italian populations has been provided by biodemographic investigations only (e.g. Fiorini *et al.*, 2007; Boattini *et al.*, 2011). It is also worth noting that the extreme reduction of genetic heterogeneity observed in some geographic isolates in which language has acted as an additional barrier to gene flow makes such groups of particular usefulness for genetic studies of complex diseases (see Heutink & Oostra, 2002).

### Re-assessing genetic diversity of Italian populations

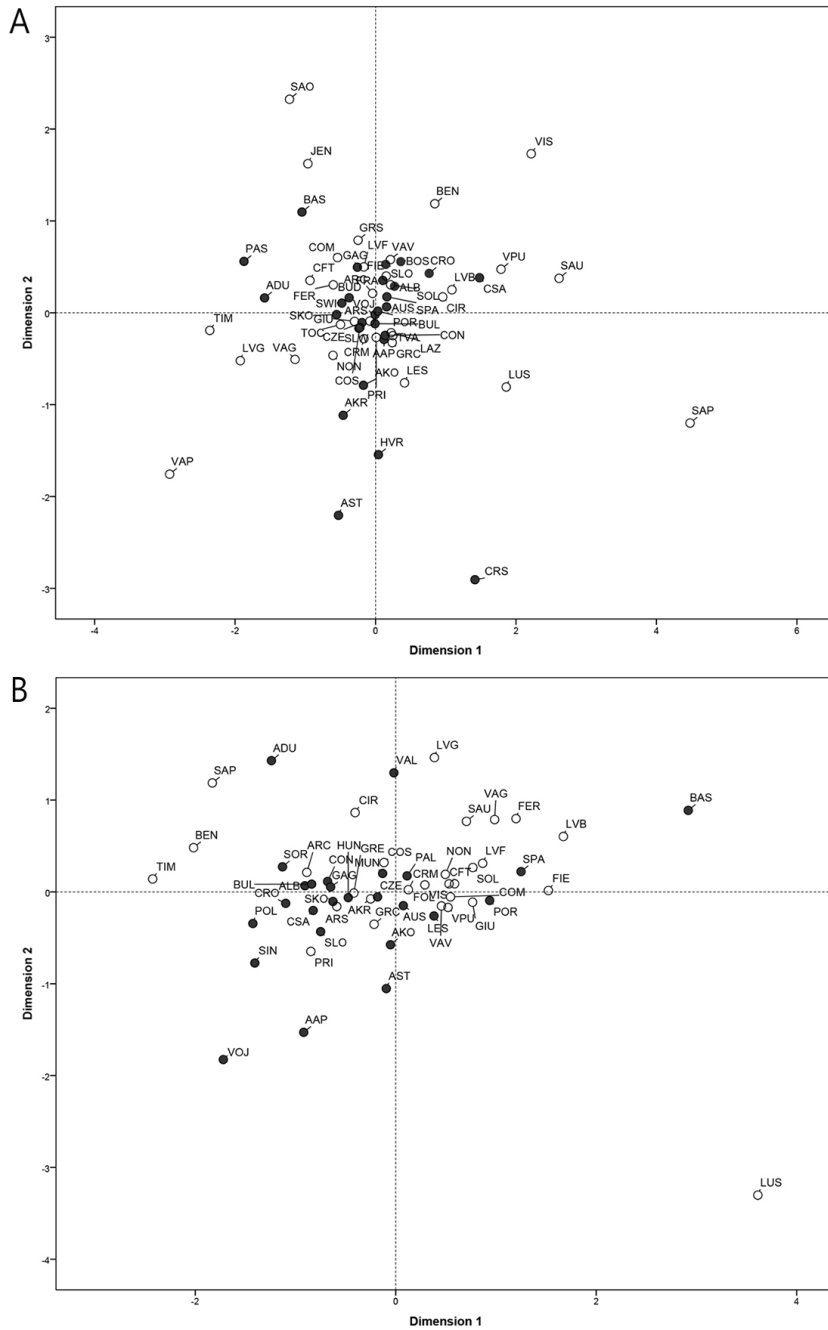
Another opportunity offered by our collaborative study was to produce a more complete picture of the diversity of Italian populations, combining data from large populations, which are presumably exempt from isolation factors, with the results obtained for ethno-linguistic minorities and geographical isolates. A contextualization into the European background seemed to us to represent a convenient approach in order to evaluate the magnitude of genetic heterogeneity, a necessary step to understand whether human genetic variation parallels plant and animal biodiversity (see introduction). Also in this case,

we started by assembling *ad hoc* mtDNA and Y chromosome datasets, in which we included populations settled throughout the continent (Supplementary Tab. S5 and S6).

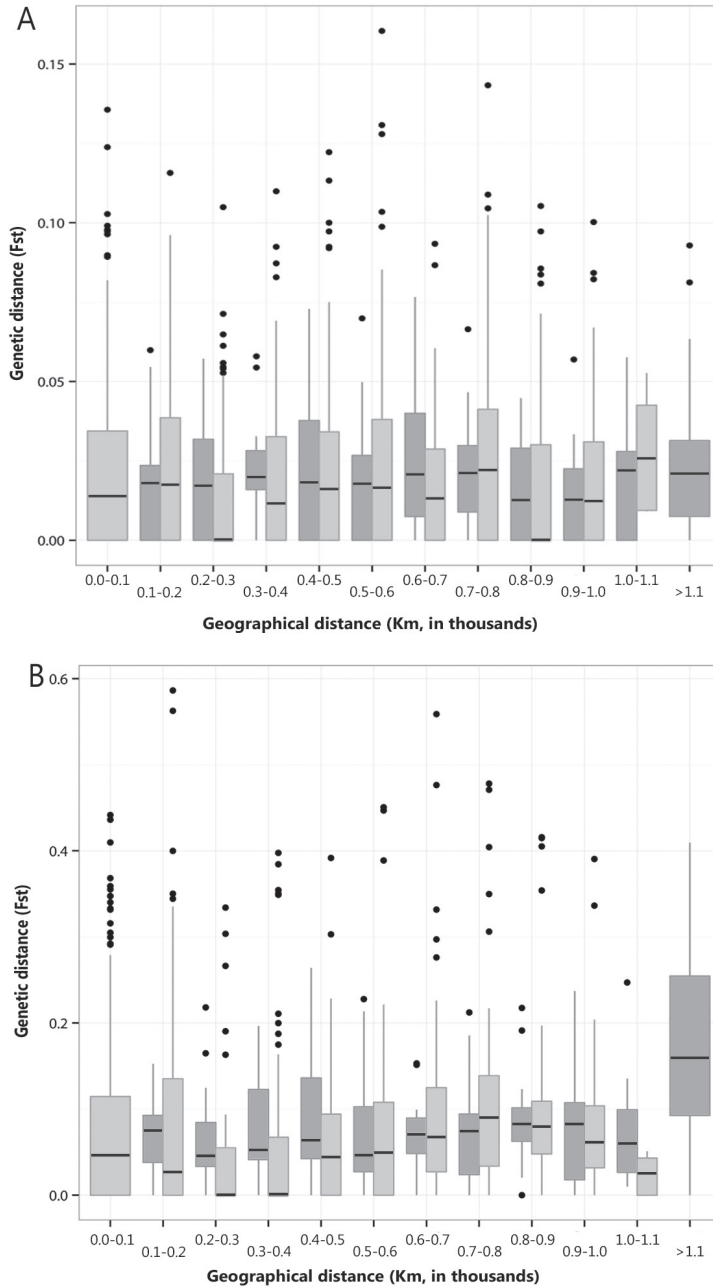
Both Italian and European open populations are found in the middle of the MDS plots of genetic distances, with a more tight clustering observed for mtDNA (Fig. 8). Although strong signals of genetic differentiation for both maternal and paternal lineages were detected for some European isolates (e.g. Aromuns from Stip and Basques), the most evident outliers are represented by some Italian isolates (Sappada and Luserna). Looking at mtDNA only, the genetic distinctiveness of the Isarco Valley, Ogliastra and Vallepietra is also noticeable. To obtain a quantitative assessment of the diversity among Italian and European populations, we carried out an AMOVA using an extended dataset (see appendices 2 and 3). Including geographically and/or isolated groups in the analyses, we observed increased estimates of diversity among Italians (mtDNA from 0.38% to 1.89%; Y chromosome from 3.19 to 6.89%) which were higher than for Europeans (mtDNA from 0.33% to 1.52%; Y chromosome 8.95% from to 10.60%). It seems, therefore, that inclusion of isolates in the dataset makes the extent of diversity of Italian populations slightly higher than Europeans for mtDNA and reduces the negative difference observed for Y chromosome. However, given the substantial difference in the ratio between open and isolated groups and spatial dispersion between Italy and Europe, we thought it useful to compare genetic and geographic distances using box plots of quartile distributions (Fig. 9). In the case of mtDNA, genetic distances of Italian isolates between each other and with Italian open populations are characterized by wider distributions and lower median values than Europeans for 7 and 8 ranges of geographic distances (out of 10), respectively. However, a high level of differentiation among Italian populations is signalled by the outlying genetic distances observed at practically all geographic ranges. Although this finding may be again related to the availability of a larger number of data on Italian compared to European isolates, it should be noted that large genetic distances



**Fig. 7 - Scatter plot of haplotype diversity and average Fst of isolated populations for (A) mtDNA and (B) Y chromosome (LUS not shown HD=0.590; average Fst= 0.322). Dotted lines represent the boundaries defined by the analysis of outliers (interquartile method). Symbols and colors as in Fig. 6. MtDNA labels: 1=ARC, 2=ARS, 3=BEN, 4=CFT, 5=CIR, 6=CRM, 7=FER, 8=FIE, 9=GIU, 10=GRC, 11=JEN, 12=LES, 13=LUS, 14=LVB, 15=LVF, 16=LVG, 17=NON, 18=PRI, 19=SAO, 20=SAP, 21=SAU, 22=SOL, 23=TIM, 24=TOC, 25=VAG, 26=VAP, 27=VAV, 28=VIS, 29=VPU. Y chromosome labels: 1=ARC, 2=ARS, 3=BEN, 4=CFT, 5=CIR, 6=CRM, 7=FER, 8=FIE, 9=GIU, 10=GRC, 11=LES, 12=LUS, 13=LVB, 14=LVF, 15=LVG, 16=NON, 17=PRI, 18=SAP, 19=SAU, 20=SOL, 21=TIM, 22=VAG.**



**Fig. 8 - Multi-Dimensional Scaling plot of  $F_{st}$  genetic distances among Italian and European populations based on (A) mtDNA HVR-1 sequences (stress value=0.226) and (B) 5 Y chromosome STRs (stress value=0.183). Filled and empty dots represent European and Italian populations, respectively. Population abbreviations as in Supplementary Table S3 and S4.**



**Fig. 9 - Box plot analyses performed on  $F_{st}$  values within different ranges of geographic distances (expressed as thousands of kilometers) for mtDNA (A) and Y chromosome (B). The light gray boxes describe the distribution of all pairwise genetic distances between Italian isolates and those between Italian isolates and open Italian populations; the dark gray boxes indicate the distribution of pairwise genetic distances between European isolates and those between European isolates and open European populations.**

between Italian groups are observed even at short geographic distances (<200km). As predicted by previous analyses (Fig. 7), an important contribution to this finding comes from ethno-linguistic minorities of the Eastern Alps. Another range of geographic distances in which the signal of heterogeneity of Italian populations is more marked is between 700 and 800 km. Here, a substantial effect comes from the genetic distances of Sardinian isolates (Benetutti, Carloforte and Ogliastra). Results obtained using Y chromosome data point to a closer correlation between genetic and geographical distances ( $r^2$  value of 0.226,  $p < 0.001$ , for European populations), which is in line with previous observations at a wider geographic scale (Rosser *et al.*, 2000). As observed for mtDNA, a greater differentiation is shown by Italian populations in the ranges between 0-200 and 700-800 km, with a higher frequency of extreme values for all intervals of geographic distances.

### Concluding remarks

Our study moves from the assumption that human biological variation is an integral part of what we mean by the term biodiversity. This concept may seem self-evident. However, it is contradicted by many examples in which biodiversity is thought to be necessarily linked to conservation, rather than a value *per se*, while human factors are considered only in relation to anthropic effects or educational aspects (e.g. Johns, 2009; Marton-Lefèvre, 2010). Not less importantly, the study of human biodiversity is often perceived as a difficult, if not risky, task, given the long tradition of misuse of this concept which still survives in bio-medical literature (e.g. Burchard *et al.*, 2003).

Once shortcomings and prejudices have been removed, the question of whether human genetic diversity observed in a certain territory shows similarities with other layers of biodiversity becomes worth answering. Certainly, comparing observations on a large number of species studied at various ecological scales using different approaches (e.g. morphological and genetic) with

results from populations of a single species could appear to be simplistic. Obviously, our survey is intended to be just a first step towards further studies driven by explicit hypotheses regarding the impact of environmental and demographic factors on different layers of biodiversity. We also maintain that studying human populations in the framework of biodiversity has an intrinsic value, since it may help us gain insights into the impact of socio-cultural factors on human genetic variation and on biological diversity as a whole.

The results of this collaborative study may be summarized in two main points. We could shed light on the genetic structure of most of the Italian ethno-linguistic minorities, showing that a combination of linguistic and geographic factors is probably responsible for the presence of the most robust signatures of genetic isolation. Furthermore, drawing a picture of Italian populations which also includes linguistic and/or geographic isolates allowed us to better appreciate the noticeable extent of the genetic diversity of our country. It turned out to be slightly greater than that of Europeans for maternal lineages on the whole, and at specific ranges of geographic distances for both genetic markers. Therefore, our survey provides a first affirmative response to the question of whether the genetic diversity of human populations shows a pattern which is comparable to what has been observed in our country for plant and (the rest of) animal biodiversity. The longitudinal extension and the nature of the natural bridge between Central Europe and the Mediterranean seems to provide an explanation also to the high level of inter-population variation. In fact, some of the isolates which determine the noticeable diversity we observed in our study are located along the boundaries of Italian territory. Furthermore, the small demographic size of some of these groups has probably increased their departure from the surrounding genetic background. From an anthropological point of view, the finding that an important contribution to the genetic diversity of Italy comes from the so called “linguistic islands” (e.g. German speaking groups of Sappada and Luserna from the Eastern Italian Alps) is of

### Info on the web

[http://www.isita-org.com/Anthro-Digit/Italian\\_isolates\\_data/index.html](http://www.isita-org.com/Anthro-Digit/Italian_isolates_data/index.html)

*An online repository of genetic data regarding linguistic and/or geographic isolates from the Italian territory.*

[http://www.isita-org.com/Anthro-Digit/Italian\\_isolates\\_pictures/index.html](http://www.isita-org.com/Anthro-Digit/Italian_isolates_pictures/index.html)

*A photogallery of places and peoples.*

[http://ec.europa.eu/languages/languages-of-europe/regional-and-minority-languages\\_en.htm](http://ec.europa.eu/languages/languages-of-europe/regional-and-minority-languages_en.htm)

*Information about projects regarding European minority languages supported by the European Commission.*

[http://it.wikipedia.org/wiki/Minoranze\\_linguistiche\\_d'Italia](http://it.wikipedia.org/wiki/Minoranze_linguistiche_d'Italia)

*A useful starting point for a web search on Italian minority languages (in Italian only).*

<http://www.isolelinguistiche.it/home.page>

*Web site of the unitary committee of the historical German linguistic islands (in Italian and German, with a good link page).*

particular significance. In fact, this is a further proof of how taking into account social and cultural factors may help us understand the structure of human genetic variation more in depth.

Obviously, our study should be seen as a first step towards a more complete assessment. Extending the sampling to the few linguistic isolates that have yet to be included in the survey (e.g. Walser, Occitan and Franco-Provençal linguistic minorities from the Western Alps) could offer an even more detailed picture. Not less importantly, some limitations of unilinearly transmitted polymorphisms (e.g. their nature of single loci and the possible confounding effect of small sample sizes) might be overcome by increasing their resolution or, even better, exploiting the power of large panels of SNPs. Our research network has already planned to deepen the analysis of the genetic structure of isolated populations using a genome-wide approach implemented in the new genotyping array “GenoChip” (Elhaik et al., 2013). Enforcing the interdisciplinary collaborations already initiated in the course of this study (e.g. Fiorini et al., 2007; Robledo et al. 2012; Coia et al., 2013) will be a key-step to really take advantage of these new genomic data.

### Author Contributions

Designed the research: MCP CMC VC GDB DP GP GV ST.

Collected the samples: PA AB CB GB SB CMC LC MCP VC FC GDB ZAF DL LM GP OR RR ES LS SS GT ST GV.

Performed the experiments: MA CB FB IB SB VB VC MCP MCR SDF GF PF SS.

Analyzed the data: AB PA CMC MCP VD FM LT ST.

Wrote the paper: PA AB CMC MCP GDB DP ST GV.

Read and approved the manuscript: All authors.

### Acknowledgements

*This paper is dedicated to our friend and colleague Laura Morelli. We would like to thank: the medical and nursing staff of the Azienda Provinciale per i Servizi Sanitari, Trento; the volunteers of the Associazione Volontari Italiani Sangue del Trentino (AVIS), Biblioteca Intercomunale di Primiero, Istituto Cultural Ladin and Kulturinstitut Lusern; all the people who kindly contributed to the collection of*

*samples in Trentino* (<http://laboratoriobagolini.it>); Marcella Benedetti (Municipality of Sappada), Nino Pacilè and Lucia Protto (Municipality of Sauris), Vito Massalongo (Giazza), Ottaviano Matiz and Velia Plözner (Timaù) for their warm assistance in sample collection and hospitality; all the people from Benetutti, Circello, Carloforte and Vagli who agreed to participate to this study; Silvano Zaccone (Consorzio “Il Cigno”), Sauro Mattei, Alfonso Tatavitto and Massimo Del Grosso for their fundamental help in collecting samples in Samnite and Apuan areas; Lorenzo Marcuccetti for his commentary to Liguers’ history and linguistics; Michele Musacchia (Azienda Sanitaria Locale 4, Corleone) and Giorgio Ales (MD, Piana degli Albanesi) for their help in collecting Sicilian Arbereshe.

*This research was made possible by grants from the Italian Ministry of Education, University and Research (PRIN projects 2007 and 2009, Principal Investigator DP). Research work in the Eastern Alps was further supported by the Istituto Italiano di Antropologia (project “Atlante bioculturale Italiano”; funds to GDB), the University of Rome “La Sapienza” (project “L’isolamento genetico in popolazioni europee”, funds to GDB) and the Provincia Autonoma di Trento (Project “Biostre”; funds to VC).*

## References

- Addison Posey D. 1999. *Cultural and Spiritual Values of Biodiversity*. Intermediate Technology Publications, London.
- Ambrosi A.C. 1956. Gli attuali limiti dell’area fonetica cacuminale nelle Alpi Apuane. *Giornale Storico della Lunigiana e del territorio lucense*, 7: 5-25.
- Angius A., Melis P.M., Morelli L., Petretto E., Casu G., Maestrale G.B., Fraumene C., Bebbere D., Forabosco P. & Pirastu M. 2001. Archival, demographic and genetic studies define a Sardinian sub-isolate as a suitable model for mapping complex traits. *Hum. Genet.*, 109: 198-209.
- Ballantyne K.N., Goedbloed M., Fang R., Schaap O., Lao O., Wollstein A., Choi Y., van Duijn K., Vermeulen M., Brauer S., Decorte R., Poetsch M., von Wurmb-Schwark N., de Knijff P., Labuda D., Vézina H., Knoblauch H., Lessig R., Roewer L., Ploski R., Dobosz T., Henke L., Henke J., Furtado M.R. & Kayser M. 2010. Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *Am. J. Hum. Genet.*, 87: 341-353.
- Banks W.E., d’Errico F., Peterson A.T., Vanhaeren M., Kageyama M., Sepulchre P., Ramstein G., Jost A. & Lunt D. 2008. Human ecological niches and ranges during the LGM in Europe derived from an application of eco-cultural niche modeling. *J. Archaeol. Sci.*, 35: 481-491.
- Barker G. & Rasmussen T. 1998. *The Etruscans*. Blackwell, Oxford.
- Beekes R.S.P. 2003. *The origins of Etruscans*. KNAW, Amsterdam.
- Bertoncini S., Ferri G., Busby G., Taglioli L., Alù M., Capelli C., Paoli G. & Tofanelli S. 2012. A Y Variant Which Traces the Genetic Heritage of Liguers Tribes. *J. Biol. Res.*, 84: 143-146.
- Blasi C., Filibeck G. & Taglianti A.V. 2005. Biodiversità e biogeografia. In C. Blasi, L. Boitani, S. La Posta, F. Manes & M. Marchetti (eds.): *Stato sulla biodiversità in Italia*. Contributo alla strategia nazionale per la biodiversità, pp. 40-56. Palombi Editori, Roma.
- Boattini A., Martinez-Cruz B., Sarno S., Harmant C., Useli A., Sanz P., Yang-Yao D., Manry J., Ciani G., Luiselli D., Quintana-Murci L., Comas D., Pettener D. & Genographic Consortium 2013. Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata. *PLoS One*, 8: e65441.
- Boattini A., Luiselli D., Sazzini M., Useli A., Tagarelli G & Pettener D. 2011. Linking Italy and the Balkans. A Y-chromosome perspective from the Arbereshe of Calabria. *Ann. Hum. Biol.*, 38: 59-68.
- Boattini A., Pedrosi M.E., Luiselli D. & Pettener D. 2010. Dissecting a human isolate: Novel sampling criteria for analysis of the genetic structure of the Val di Scalve (Italian Pre-Alps). *Ann. Hum. Biol.*, 37: 604-609.
- Bosch E., Calafell F., Gonzalez-Neira A., Flaiz C., Mateu E., Scheil H.G., Huckenbeck W., Efremovska L., Mikerezi I., Xirotiris N., Grasa C., Schmidt H. & Comas D. 2006. Paternal



- and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Ann. Hum. Genet.*, 70: 459–487.
- Brunettin G. 1998. L'insediamento di Sauris tra storiografia e rappresentazione di un'origine. In D. Cozzi, D. Isabella, E. Navarra (eds.): *Sauris, Zahre, una comunità delle Alpi carniche*, pp. 43-61. Forum Editrice Universitaria Udinese, Udine.
- Burchard E.G., Ziv E., Coyle N., Gomez S.L., Tang H., Karter A.J., Mountain J.L., Pérez-Stable E.J., Sheppard D. & Risch N. 2003. The importance of race and ethnic background in biomedical research and clinical practice. *N. Engl. J. Med.*, 348: 1170–1175.
- Busby G.B., Brisighelli F., Sánchez-Diz P., Ramos-Luis E., Martínez-Cadenas C., Thomas M.G., Bradley D.G., Gusmão L., Winney B., Bodmer W., Vennemann M., Coia V., Scarnicci F., Tofanelli S., Vona G., Ploski R., Vecchiotti C., Zemunik T., Rudan I., Karachanak S., Toncheva D., Anagnostou P., Ferri G., Rapone C., Hervig T., Moen T., Wilson J.F. & Capelli C. 2012. The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. *Proc. Roy. Soc. B. - Biol. Sci.*, 279: 884-892.
- Calò C.M., Corrias L., Vona G., Bachis V. & Robledo R. 2012. Sampling strategies in a linguistic isolate: results from mtDNA analysis. *Am. J. Hum. Biol.*, 24: 192-194.
- Calò C.M., Corrias L., Bachis V., Vona G., Brandas A., Scudiero C.M., Di Fede C., Mameli A. & Robledo R. 2013. Analisi di due isolati della Sardegna (Italia) attraverso lo studio dei polimorfismi del cromosoma Y. *Antropo*, 29: 1-7.
- Calò C.M., Vona G. 1994. Carloforte: evoluzione della struttura matrimoniale. *Rendiconti Seminario Facoltà Università Cagliari*, 64: 73-87.
- Capocasa M., Anagnostou P., Battaglia C. & Destro Bisol G. 2012. Il patrimonio genetico dei Cimbri della Lessinia, uno studio interdisciplinare. *Cimbri/Tzimbar*, 47: 99-107.
- Capocasa M., Anagnostou P., Battaglia C. & Destro Bisol G. 2013a. Omogeneità culturale e peculiarità genetiche della comunità alpina di Timau. *Asou Geats*, 72: 6-7.
- Capocasa M., Battaglia C., Anagnostou P., Montinaro F., Boschi I., Ferri G., Alu M., Coia V., Crivellaro F. & Destro Bisol G. 2013b. Detecting genetic isolation in human populations: a study of European language minorities. *PLoS One*, 8: e56371.
- Capocasa M., Taglioli L., Anagnostou P., Paoli G. & Danubio M.E. Determinants of marital behaviour in five Apennine communities of Central Italy inferred by surname analysis, repeated pairs and kinship estimates. *Homo* (in press), doi: 10.1016/j.jchb.2013.08.001.
- Cavalli Sforza L.L. & Bodmer W.F. 1971. *The Genetics of Human Populations*. Freeman, San Francisco.
- Cerri N., Verzeletti A., Bandera B. & De Ferrari F. 2005. Population data for 12 Y-chromosome STRs in a sample from Brescia (northern Italy). *Forensic. Sci. Int.*, 152: 83-87.
- Coia V., Boschi I., Trombetta F., Cavulli F., Montinaro F., Destro Bisol G., Grimaldi S. & Pedrotti A. 2012. Evidence of high genetic variation among linguistically diverse populations on a micro-geographic scale: a case study of the Italian Alps. *J. Hum. Genet.*, 57: 254-260.
- Coia V., Capocasa M., Anagnostou P., Pascali V., Scarnicci F., Boschi I., Battaglia C., Crivellaro F., Ferri G., Alu M., Brisighelli F., Busby G.B.J., Capelli C., Maixner F., Cipollini G., Viazzo P.P., Zink A. & Destro Bisol G. 2013. Demographic histories, isolation and social factors as determinants of the genetic structure of Alpine linguistic groups. *PLoS One*, 8: e81704.
- Comas D., Calafell F., Mateu E., Perez-Lezaun A. & Bertranpetit J. 1996. Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationship to the European populations. *Mol. Biol. Evol.*, 13: 1067-1077.
- Congiu A., Anagnostou P., Milia N., Capocasa M., Montinaro F. & Destro Bisol G. 2012. Online databases for mtDNA and Y chromosome polymorphisms in human populations. *J. Anthropol. Sci.*, 90: 201-215.
- Côrte-Real H.B., Macaulay V.A., Richards M.B., Hariti G., Issad M.S., Cambon-Thomsen A.,

- Papiha S., Bertranpetit J. & Sykes B.C. 1996. Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann. Hum. Genet.*, 60: 331-350.
- De Concini W. 1997. *Gli altri delle Alpi. Minoranze linguistiche dell'arco alpino italiano*. Comune di Pergine Valsugana, Trento.
- Denison N. 1971. Some observations on language variety and plurilingualism. In E. Ardener (ed.): *Social anthropology and language*, pp. 157-183. Tavistock, London.
- Destro Bisol G., Anagnostou P., Batini C., Battaglia C., Bertoncini S., Bottini A., Caciagli L., Calò C.M., Capelli C., Capocasa M., Castri L., Ciani G., Coia V., Corrias L., Crivellaro F., Ghiani M.E., Luiselli D., Mela C., Melis A., Montano V., Paoli G., Sanna E., Rufo F., Sazzini M., Taglioli L., Tofanelli S., Useli A., Vona G. & Pettener D. 2008. Italian isolates today: geographic and linguistic factors shaping human biodiversity. *J. Anthropol. Sci.*, 86: 179-188.
- Di Rienzo A. & Wilson A.C. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*, 88: 1597-1601.
- Dilke O.A.W. 1971. *The Roman Land Survivors: an introduction to the agrimensores*. David & Charles, Newton Abbott.
- Donald M. 2008. *I Normanni in Italia*. Laterza, Roma-Bari.
- Elhaik E., Greenspan E., Staats S., Krahn T., Tyler-Smith C., Xue Y., Tofanelli S., Francalacci P., Cucca F., Pagani L., Jin L., Li H., Schurr T.G., Greenspan B., Spencer Wells R. & Genographic Consortium. The GenoChip: a new tool for genetic anthropology. *Genome Biol. Evol.*, 5: 1021-31.
- Excoffier L. & Lischer H.E.L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.*, 10: 564-567.
- Fadhlaoui-Zid K., Plaza S., Calafell F., Ben Amor M., Comas D. & Bennamar Elgaaiad A. 2004. Mitochondrial DNA heterogeneity in Tunisian Berbers. *Ann. Hum. Genet.*, 68: 222-233.
- Falchi A., Giovannoni L., Calò C.M., Piras S.I., Moral P., Paoli G., Vona G. & Varesi L. 2006. Genetic history of some western Mediterranean human isolates through mtDNA HVRI polymorphisms. *J. Hum. Genet.*, 51: 9-14.
- Ferri G., Alù M., Corradini B., Radheshi E. & Beduschi G. 2009. Slow and fast evolving markers typing in Modena males (North Italy). *Forensic Sci. Int. Genet.*, 3: e31-e33.
- Fiorini S., Tagarelli G., Boattini A., Luiselli D., Piro A., Tagarelli A. & Pettener D. 2007. Ethnicity and evolution of the biodemographic structure of Arbereshe and Italian populations of the Pollino area, Southern Italy (1820-1984). *Amer. Anthropol.*, 109: 735-746.
- Fraley C. & Raftery A.E. 2002. Model-based Clustering, Discriminant Analysis and Density Estimation. *J. Amer. Statist. Assoc.*, 97: 611-631.
- Fraley C. & Raftery A.E. 2006. MCLUST Version 3 for R: Normal Mixture Modeling and Model-based Clustering. Technical Report No. 504, Department of Statistics, University of Washington (revised 2009).
- Francalacci P., Bertranpetit J., Calafell F. & Underhill P.A. 1996. Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *Am. J. Phys. Anthropol.*, 100: 443-460.
- Franceschi, M.G. & Paoli, G. 1994. Isolation factors and kinship by isonymy in a group of parishes in northern Tuscany (Italy): influence of within-parish similarity level on between-parish similarity pattern. *Hum. Biol.*, 66: 905-916.
- Fraumene C., Belle E.M., Castri L., Sanna S., Mancosu G., Cosso M., Marras F., Barbuiani G., Pirastu M. & Angius A. 2006. High resolution analysis and phylogenetic network construction using complete mtDNA sequences in sardinian genetic isolates. *Mol. Biol. Evol.*, 23: 2101-2111.
- Fraumene C., Petretto E., Angius A. & Pirastu M. 2003. Striking differentiation of sub-populations within a genetically homogeneous isolate (Ogliastra) in Sardinia as revealed by mtDNA analysis. *Hum. Genet.*, 114: 1-10.
- Frigi S., Pereira P., Pereira L., Yacoubi B., Gusmao L., Alves C., Khodjet el Khil H.L., Cherni A., Amorim A. & El Gaaied A. 2006. Data for Y-chromosome haplotypes defined by 17 STRs

- (AmpFLSTR1 Yfiler™) in two Tunisian Berber communities. *Forensic Sci. Int.*, 160: 80–83.
- Fu Y.X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147: 915–925.
- Gagliardi L. 2006. *Mobilità e integrazione delle persone nei centri cittadini, Aspetti giuridici. I. La classificazione degli incolae*. Giuffrè Editore, Milano.
- Ghiani M.E., Piras I.S., Mitchell R.J. & Vona G. 2004. Y-chromosome 10 locus short tandem repeat haplotypes in a population sample from Sicily Italy. *Leg. Med. (Tokyo)*, 6: 89–96.
- Giacomelli G. 1975. Aree lessicali toscane. *La ricerca dialettale*, 1: 115–152.
- Giunta F. 2003. Albanesi in Sicilia. In: M. Mandalà (ed.): *Albanesi in Sicilia*, p. 25. Mirror, Palermo.
- Gonzalez A.M., Larruga J.M., Abu-Amero K.K., Shi Y., Pestano J. & Cabrera V.M. 2007. Mitochondrial lineage M1 traces an early human backflow to Africa. *BMC Genomics*, 8: 223.
- Goudie A.S. 2013. *The human impact on the natural environment: past, present, and future*. Wiley-Blackwell, Oxford.
- Grassi F., De Mattia F., Zecca G., Sala F. & Labra M. 2008. Historical isolation and Quaternary range expansion of divergent lineages in wild grapevine. *Biol. J. Linn. Soc.*, 95: 611–619.
- Guazzelli F. 2001. *Suddivisione dialettale della Garfagnana*. Verlag, Köln.
- Harpending R.C. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.*, 66: 591–600.
- Heutink P. & Oostra B.A. 2002. Gene finding in genetically isolated populations. *Hum. Mol. Genet.*, 11: 2507–2515.
- Hey J. & Nielsen R. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *P. Natl Acad. Sci. USA*, 104: 2785–2790.
- Hin S. 2013. *The Demography of Roman Italy: Population Dynamics in an Ancient Conquest Society 201 BCE–14 CE*. Cambridge University Press, Cambridge.
- Jeran N., Havaš Auguštin D., Grahovac B., Kapović M., Metspalu E., Villems R. & Rudan P. 2009. Mitochondrial DNA heritage of Cres islanders—Example of Croatian genetic outliers. *Coll. Antropol.*, 33: 1323–1328.
- Johns D. 2009. The International Year of Biodiversity—From Talk to Action. *Conserv. Biol.*, 24: 338–340.
- Jombart T., Devillard S. & Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.*, 11: 94.
- Jörg J. 2002. *Storia dei longobardi*. Einaudi, Milano.
- Krings M., Salem A.E., Bauer K., Geisert H., Malek A.K., Chaix L., Simon C., Welsby D., Di Rienzo A., Utermann G., Sajantila A., Paabo S. & Stoneking M. 1999. mtDNA analysis of Nile River Valley populations: a genetic corridor or a barrier to migration?. *Am. J. Hum. Genet.*, 64: 1166–1176.
- Livius Titus. *Storia Romana XXXVIII*.
- Marcuccetti L. 2008. *La lingua dimenticata*. Luna Editore, La Spezia.
- Marcuccetti L. 2012. *La fondazione di Luca e Luna*. Centro Stampa Photos, La Spezia.
- Marks J. 1995. *Human biodiversity. Genes, race, and history*. Transaction Publishers, New Brunswick.
- Marton-Lefèvre J. 2010. Biodiversity is our life. *Science*, 313: 1179.
- Metcalfe A. 2009. *The Muslims of medieval Italy*. Edinburgh University Press, Edinburgh.
- Montinaro F., Boschi I., Trombetta F., Merigoli S., Anagnostou P., Battaglia C., Capocasa M., Crivellaro F., Destro-Bisol G. & Coia V. 2012. Using forensic microsatellites to decipher the genetic structure of linguistic and geographic isolates: a survey in the eastern Italian Alps. *Forensic Sci. Int. Genet.*, 6: 827–833.
- Morelli L., Grosso M.G., Vona G., Varesi L., Torroni A. & Francalacci P. 2000. Frequency distribution of mitochondrial DNA haplogroups in Corsica and Sardinia. *Hum. Biol.*, 72: 585–595.

- Morelli L., Paoli G. & Francalacci P. 2002. Surname analysis of the Corsican population reveals an agreement with geographical and linguistic structure. *J. Biosoc. Sci.*, 34: 289–301.
- Moseley C. 2010. *Atlas of the World's Languages in Danger*. UNESCO, Paris.
- Murru Corriga G. 1995. Il cognome della madre. Tendenze matrilineari nella parentela in Sardegna (XVI-XVIII sec.). *Antropologia Contemporanea*, 18: 47–66.
- Navarra E. 2002. Comportamenti demografici e organizzazione socio-economica in due comunità germanofone delle Alpi orientali: Sappada e Sauris (sec. XVIII e XIX). In A. Fornasin & A. Zannini (eds.): *Uomini e comunità delle montagne*, pp. 113–132. Forum Editrice Universitaria Udinese, Udine.
- Neel J. 1992. Minority populations as genetic isolates: the interpretation of inbreeding results. In A.H. Bittles & D.F. Roberts (eds.): *Minority Populations: Genetics, Demography and Health*, pp. 1–13. The MacMillan Press, London.
- Nei M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Onofri V., Alessandrini F., Turchi C., Fraternali B., Buscemi L., Pesaresi M. & Tagliabracchi A. 2007. Y-chromosome genetic structure in sub-Apennine populations of Central Italy by SNP and STR analysis. *Int. J. Legal Med.*, 121: 234–237.
- Padovani L.M., Carrabba P., Di Giovanni B. & Mauro F. 2009. *Biodiversità. Risorse per lo sviluppo*. ENEA, Roma.
- Patterson J. 2009. *Samnites, Ligurians and Romans*. Edizioni Comune di Circello, Benevento.
- Peratoner A. 2002. *Sappada/Plodn. Storia, etnografia e ambiente naturale*. Tiziano Editore, Pieve di Cadore.
- Petit R.J., Aguinagalde I., de Beaulieu J.L., Bittkau C., Brewer S., Cheddadi R., Ennos R., Fineschi S., Grivet D., Lascoux M., Mohanty A., Müller-Starck G., Demesure-Musch B., Palmé A., Martin J.P., Rendell S. & Vendramin G.G. 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, 300: 1563–1565.
- Petris B. 1980. *Tischlbong Tamau Timau*. Del Bianco, Udine.
- Piazza A., Cappello N., Olivetti E. & Rendine S. 1988. A genetic history of Italy. *Ann. Hum. Genet.*, 52: 203–13.
- Plaza S., Calafell F., Helal A., Bouzerna N., Lefranc G., Bertranpetit J. & Comas D. 2003. Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. *Ann. Hum. Genet.*, 67: 312–328.
- Presciuttini S., Caglià A., Alù M., Asmundo A., Buscemi L., Caenazzo L., Carnevali E., Carrà E., De Battisti Z., De Stefano F., Dominici R., Piccinini A., Resta N., Ricci U. & Pascali V.L. 2001. Y-chromosome haplotypes in Italy: the GEFI collaborative database. *Forensic Sci. Int.*, 122: 184–188.
- Reynolds J., Weir B.S. & Cockerham C.C. 1983. Estimation for the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, 105: 67–779.
- Rizzi E. 1993. *Storia dei Walser*. Fondazione E. Monti, Anzola d'Ossola.
- Robledo R., Corrias L., Bachis V., Puddu N., Mameli A., Vona G. & Calò C.M. 2012. Analysis of a Genetic Isolate: The Case of Carloforte (Italy). *Hum. Biol.*, 86: 735–754.
- Robledo R., Piras I., Beggs W. & Calò C. 2009. Analysis of 31 STR loci in the genetic isolate of Carloforte (Sardinia, Italy). *Genet. Mol. Biol.*, 32: 462–465.
- Rosser Z.H., Zerjal T., Hurler M.E., Adojaan M., Alavantic D., Amorim A., Amos W., Armenteros M., Arroyo E., Barbujani G., Beckman G., Beckman L., Bertranpetit J., Bosch E., Bradley D.G., Brede G., Cooper G., Côrte-Real H.B., de Knijff P., Decorte R., Dubrova Y.E., Evgrafov O., Gilissen A., Glisic S., Gölge M., Hill E.W., Jeziorowska A., Kalaydjieva L., Kayser M., Kivisild T., Kravchenko S.A., Krumina A., Kucinskas V., Lavinha J., Livshits L.A., Malaspina P., Maria S., McElreavey K., Meitinger T.A., Mikelsaar A.V., Mitchell R.J., Nafa K., Nicholson J., Nørby S., Pandya A., Parik J., Patsalis P.C., Pereira L., Peterlin B., Pielberg G., Prata M.J., Previderé C., Roewer L., Rootsi S., Rubinsztein D.C., Saillard J., Santos F.R., Stefanescu G., Sykes B.C., Tolun

- A., Villems R., Tyler-Smith C. & Jobling M.A. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J. Hum. Genet.*, 67: 1526-1543.
- Scheidel M. 2003. The Greek Demographic Expansion: Models and Comparisons. *J. Hellenic Stud.*, 123: 120-140.
- Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 139: 457-462.
- Steinicke E., Walder J., Löffler R. & Beismann M. 2011. Autochthonous linguistic minorities in the Italian Alps: new legislation – new identifications – new demographic processes. *Journal of Alpine Research*, 99: 2.
- Stenico M., Nigro L., Bertorelle G., Calafell F., Capitanio M., Corrain C. & Barbujani G. 1996. High mitochondrial sequence diversity in linguistic isolates of the Alps. *Am. J. Hum. Genet.*, 59: 1363-1375.
- Taberlet P., Fumagalli L., Wust-Saucy A.G. & Cosson J.F. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.*, 7: 453-464.
- Tagarelli G., Fiorini S., Piro A., Luiselli D., Tagarelli A. & Pettener D. 2007. Ethnicity and biodemographic structure in the Arbereshe of the province of Cosenza, southern Italy, in the XIX century. *Coll. Antropol.*, 31: 331-338.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123: 585-595.
- Tofanelli S., Taglioli L., Merlitti D. & Paoli G. 2011. Tools which simulate the evolution of uni-parentally transmitted elements of the human genome. *J. Anthropol. Sci.*, 89: 201-219.
- Toso F. 2008. Le minoranze linguistiche in Italia. Il Mulino, Bologna.
- Turrina S., Atzei R. & De Leo D. 2006. Y-chromosomal STR haplotypes in a Northeast Italian population sample using 17plex loci PCR assay. *Int. J. Legal Med.*, 120: 56-59.
- UNESCO 2001. *Universal Declaration on Cultural Diversity*. UNESCO, Paris.
- Vallebona G. 1974. *Storia di una Colonizzazione*. Edizioni Della Torre, Cagliari.
- van Oven M., Hämmerle J.M., van Schoor M., Kushnick G., Pennekamp P., Zega I., Lao O., Brown L., Kennerknecht I. & Kayser M. 2011. Unexpected island effects at an extreme: reduced Y chromosome and mitochondrial DNA diversity in Nias. *Mol. Biol. Evol.*, 28: 1349-1361.
- 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). *Am. J. Hum. Biol.*, 12: 339-351.
- Veeramah K.R., Tönjes A., Kovacs P., Gross A., Wegmann D., Geary P., Gasperikova D., Klimes I., Scholz M., Novembre J. & Stumvoll M. 2011. Genetic variation in the Sorbs of eastern Germany in the context of broader European genetic diversity. *Eur. J. Hum. Genet.*, 19: 995-1001.
- Vona G., Ghiani M.E., Calò C.M., Memmi M. & Varesi L. 2001. Mitochondrial DNA sequence analysis in Sicily. *Am. J. Hum. Biol.*, 13: 576-589.
- Walsh B. 2001. Estimating the time to the most recent common ancestor for the Y chromosome or mitochondrial DNA for a pair of individuals. *Genetics*, 158: 897-912.
- Zangari D. 1941. *Le colonie italo-albanesi di Calabria*. Editore Casella, Napoli.
- Zavattari P., Deidda E., Whalen M., Lampis R., Mulargia A., Loddo M., Eaves I., Mastio G., Todd J.A. & Cucca F. 2000. Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Hum. Mol. Genet.*, 9: 2947-2957.

**Appendix 1 - Other unpublished and literature data used in this study. Abbreviations: Alt, altitude; G, geographic isolate; GL, geo-linguistic isolate; L, linguistic isolate; O, open population.**

POPULATION	ABB.	STATUS	CENSUS*	ALT	mtDNA		Y CHROMOSOME	
					N	REFERENCE	N	REFERENCE
Adige Valley	TVA	O	166394	199	56	Coia <i>et al.</i> , 2012	56	Coia <i>et al.</i> , 2013
Ascoli Piceno	PIC	O	49892	154	53	Brisighelli <i>et al.</i> , 2012	38	Brisighelli <i>et al.</i> , 2012
Aviano	AVI	O	9270	159	29	Boattini <i>et al.</i> , 2013	-	-
Belvedere	BEL	O	9388	150	-	-	27	Brisighelli <i>et al.</i> , 2012
Benevento	BNV	O	61700	207	36	Boattini <i>et al.</i> , 2013	33	Boattini <i>et al.</i> , 2013
Bologna	BOL	O	373020	54	100	Bini <i>et al.</i> , 2003	-	-
Bolzano	BOZ	O	105774	262	59	Thomas <i>et al.</i> , 2008	-	-
Brescia	BRE	O	188602	149	40	Boattini <i>et al.</i> , 2013	35	Boattini <i>et al.</i> , 2013
Campobasso	CMB	O	48479	701	37	Boattini <i>et al.</i> , 2013	29	Boattini <i>et al.</i> , 2013
Casentino	CAS	O	46039	465	122	Achilli <i>et al.</i> , 2007	-	-
Como	COM	O	83132	201	39	Boattini <i>et al.</i> , 2013	41	Boattini <i>et al.</i> , 2013
Croatian Molise	CRM	L	1884	493	41	Babalini <i>et al.</i> , 2005	15	This study
Cuneo	CUN	O	55627	534	40	Boattini <i>et al.</i> , 2013	30	Boattini <i>et al.</i> , 2013
Fersina Valley	FER	G	51478	1066	25	Coia <i>et al.</i> , 2012	26	Coia <i>et al.</i> , 2013
Foligno	FOL	O	58363	234	-	-	29	Boattini <i>et al.</i> , 2013
Giudicarie Valley	GIU	G	36282	680	52	Coia <i>et al.</i> , 2012	51	Coia <i>et al.</i> , 2013
Grecanici Salento	GRC	L	10000	80	47	Brisighelli <i>et al.</i> , 2012	46	Brisighelli <i>et al.</i> , 2012
Isarco Valley	VIS	G	44500	801	34	Pichler <i>et al.</i> , 2006	-	-
Jenne	JEN	G	407	834	103	Messina <i>et al.</i> , 2010	-	-
La Spezia	SPZ	O	95378	3	50	Brisighelli <i>et al.</i> , 2012	45	Brisighelli <i>et al.</i> , 2012
Ladins (Badia Valley)	LVB	GL	10644	1341	56	Thomas <i>et al.</i> , 2008	44	Coia <i>et al.</i> , 2013
Ladins (Gardena Valley)	LVG	GL	10198	1320	46	Thomas <i>et al.</i> , 2008	51	Coia <i>et al.</i> , 2013
L'Aquila	AQU	O	110268	714	25	Boattini <i>et al.</i> , 2013	27	Boattini <i>et al.</i> , 2013
Lazio	LAZ	O	5536292	-	52	Babalini <i>et al.</i> , 2005	-	-
Lecce	LEC	O	89839	49	39	Boattini <i>et al.</i> , 2013	35	Boattini <i>et al.</i> , 2013

**Appendix 1 - Continued.**

POPULATION	ABB.	STATUS	CENSUS*	ALT	mtDNA		Y CHROMOSOME	
					N	REFERENCE	N	REFERENCE
Macerata	MAC	O	43019	315	39	Boattini <i>et al.</i> , 2013	34	Boattini <i>et al.</i> , 2013
Matera	MAT	O	59973	401	36	Boattini <i>et al.</i> , 2013	24	Boattini <i>et al.</i> , 2013
Non Valley	NON	G	37832	816	48	Coia <i>et al.</i> , 2012	48	Coia <i>et al.</i> , 2013
Ogliastra	SAO	G	57959	392	175	Fraumene <i>et al.</i> , 2003	-	-
Oristano	ORI	O	32156	5	39	Boattini <i>et al.</i> , 2013	40	Boattini <i>et al.</i> , 2013
Primiero Valley	PRI	G	9959	763	40	Coia <i>et al.</i> , 2012	41	Coia <i>et al.</i> , 2013
Pusteria Valley	VPU	G	73000	997	37	Pichler <i>et al.</i> , 2006	-	-
Sole Valley	SOL	G	15235	884	63	Coia <i>et al.</i> , 2012	65	Coia <i>et al.</i> , 2013
Terni	TER	O	109482	130	31	Boattini <i>et al.</i> , 2013	-	-
Tocco da Casauria	TOC	G	2782	356	50	Verginelli <i>et al.</i> , 2003	-	-
Treviso	TRE	O	82535	15	39	Boattini <i>et al.</i> , 2013	28	Boattini <i>et al.</i> , 2013
Udine	UDI	O	100514	113	51	Brisighelli <i>et al.</i> , 2012	45	Brisighelli <i>et al.</i> , 2012
Vallepietra	VAP	G	308	825	21	Messina <i>et al.</i> , 2010	-	-
Venosta Valley	VAV	G	34307	1024	112	Pichler <i>et al.</i> , 2006	-	-
Vicenza	VIC	O	113639	39	40	Boattini <i>et al.</i> , 2013	33	Boattini <i>et al.</i> , 2013

\* Source: ISTAT (2011) (<http://demo.istat.it>)

