Italian isolates today: geographic and linguistic factors shaping human biodiversity*

Giovanni Destro Bisol1-2, Paolo Anagnostou1-4, Chiara Batini1, Cinzia Battaggia1, Stefania Bertoncini3, Alessio Boattini4, Laura Caciagli3, Carla M. Calò5, Cristian Capelli6, Marco Capocasa1, Loredana Castri4, Graziella Ciani4, Valentina Coia7, Laura Corrias5, Federica Crivellaro6, M. Elena Ghiani5, Donata Luiselli7, Cristina Mela3, Alessandra Melis5, Valeria Montano3, Giorgio Paoli3, Emanuele Sanna5, Fabrizio Rufo1, Marco Sazzini4, Luca Taglioli3, Sergio Tofanelli3, Antonella Useli4, Giuseppe Vona5 & Davide Pettener4

1) Dipartimento di Biologia Animale e dell’Uomo, Università di Roma “La Sapienza”
e-mail: destrobisol@uniroma1.it;
2) Istituto Italiano di Antropologia, Roma, Italy
3) Dipartimento di Biologia, Università di Pisa, Pisa, Italy
4) Dipartimento di Biologia Evoluzionistica e Sperimentale, Università di Bologna, Bologna, Italy
e-mail: davide.pettener@unibo.it
5) Dipartimento di Biologia Sperimentale, Sezione di Scienze Antropologiche, Università di Cagliari, Monserrato, Italy
6) Department of Zoology, University of Oxford, Oxford, United Kingdom
7) Dipartimento di Filosofia, Storia e Beni culturali, Università di Trento, Trento, Italy
8) University of Cambridge, Leverhulme Centre for Human Evolutionary Studies, Cambridge, United Kingdom

Summary – We briefly review the current status of anthropological and genetic studies of isolated populations and of their micro-evolutionary and biomedical applications, with particular emphasis on European populations. Thereafter, we describe the ongoing collaborative research project “Isolating the Isolates: geographic and cultural factors of human genetic variation” regarding Italian extant geographical and/or linguistic isolates, aimed at overcoming the limitations of previous studies regarding geographical coverage of isolates, number and type of genetic polymorphisms under study and suitability of the experimental design to investigate gene-culture coevolutionary processes. An interdisciplinary sampling approach will make it possible to collect several linguistic isolates and their geographic neighbours from Trentino, Veneto, Friuli, Tuscany, Sardinia and Calabria. This will be coupled with a shared genotyping strategy based on mitochondrial and Y-chromosomal polymorphisms. The results will be analyzed with a focus on the role of geographical and cultural factors in shaping human biodiversity. The aims of the project go beyond the simple reconstruction of the genetic structure and history of the examined groups. In fact, the study will also include an assessment for future bio-medical studies and the development of genetic and bio-demographic databases. Ethical and educational aspects are also foreseen by the project, by using informed consents together with disseminating activities in loco, completed by the creation of a dedicated web site for both scientific and public audiences.

Keywords – Isolates, geography, language, population genetics, Italy.

* a description of the contribution of each author is available at www.isita-org.com/italian_isolates
Isolated populations as a benchmark for understanding human genetic diversity

Studies of processes shaping genetic diversity in human populations are fundamental for a number of disciplines, including anthropology, medicine and forensics. In fact, they can provide information on the origin and history of human groups, as well as help clarify the microevolutionary effects of social and cultural factors or, finally, pinpoint the genetic factors underlying disease susceptibility.

Compared to large and cosmopolitan populations, the human isolates originated from a restricted number of founders due either to a past migration event or reduction in population size (e.g. a bottleneck) often show characteristics which may facilitate our insights into human genetic diversity. These include greater availability of historical and genealogical records, reduced genetic diversity and pronounced environmental uniformity. However, some issues which may complicate genetic surveys (privacy, exploitation of data and use of samples for commercial purposes) are often exacerbated in the case of studies dealing with small populations. This is well exemplified by the difficulties encountered by the proponents of the Human Genome Diversity Projects and the discussion surrounding the mass genetic screening conducted in Iceland by deCODE Genetics, (Greely, 2001; Arnason, 2004; Rufo, 2007).

Studies of human isolates must take into account the increasing awareness of the important effects exerted by socio-economic, historical, and cultural factors on the genetic structure of human populations. For example, the effects of religion, historical migration patterns, and single historical events or crises have been documented (Bittles & Smith, 1994; Crawford et al., 1995; Relethford & Crawford, 1998; Smith & Bittles, 2002), although agreement as to which of these elements leaves the most important signature has yet to be reached (e.g., Martin et al., 2000; North et al., 2000). Recent works based on Y-chromosome, mitochondrial DNA (mtDNA) and the autosomes support the role played by socio-cultural factors in determining patterns of sex-related gene flow (Seielstad et al., 1998; Oota et al., 2001; Destro-Bisol et al., 2004; Marchani et al., 2008).

In recent times, isolated populations have attracted particular interest also for bio-medical applications. In fact, they have proven extremely useful for mapping genes for rare monogenic disorders, and it has been claimed that the same advantages would also hold for common diseases (Boehnke, 2000; see Peltonen et al., 2000 for a review). These views are based on three key features. Firstly, the confounding effects of different demographic histories and environmental factors between cases and controls are limited (Varilo & Peltonen, 2004). Second, the genetic heterogeneity underlying complex diseases is decreased because of the small population size and shared ancestry, the low immigration rate and the high consanguinity, which can significantly increase genotypic relative risk and hence the ability to identify candidate genes (Shifman & Darvasi, 2001). Finally, isolated populations show an increased level of Linkage Disequilibrium extension when compared to large out bred populations (Service et al., 2006). Heutink & Oostra (2002) have proposed to classify founder populations according to the number of generations since their foundation - very old isolates (>100 generations), young isolates (<100 generations) and very young isolates (minor 20 generations) - arguing that the older the populations are, the better they are for localizing pathogenic mutations.

The analysis of classic and molecular genetic markers has to date revealed a few cases of well documented isolates in Europe. These include the Sardinians, the Basques, the Finns and, in particular, the Saami (Cavalli-Sforza & Piazza 1993; Cavalli-Sforza et al., 1994; Bertranpetit et al., 1995; Sajantila et al., 1996; Tambets et al., 2004). All these populations share a similar demographic history: a small number of original founders, subsequent isolation, rapid expansion and major bottlenecks have allowed genetic drift to mould the gene pool (Kittles et al., 1998; Sajantila et al., 1996).
Among the different facets of human culture, language is probably the most characterising, playing a critical role in the processes of cultural learning, transmission and communication. Despite this key role, few studies have carefully and systematically addressed the influence of language on the genetic structure of human populations using simultaneously families of polymorphic loci with different modes of inheritance. A more focused insight into the role of language could be provided by the investigation of linguistic isolates. However, to date, unfortunately, little has been done in this direction. Previous reports on European populations have suggested that the role played by language seems to be negligible at continental level, while genetic variation appears to be shaped mainly (Rosser et al., 2000; Semino et al., 2000) or exclusively by geography (Novembre et al., 2008; but see Poloni et al., 1997 and Barbujani, 1997). Nevertheless, a recent worldwide analysis of 377 autosomal microsatellite showed, beyond the effects of geographic distances, a detectable effect of language differences on DNA diversity at genomic level (Belle & Barbujani, 2007). On a regional scale, it is worth mentioning that genetic analysis of the Balkan area highlighted a general homogeneity across countries whose languages belong to different Indo-European branches (Bosch et al., 2006). The only exceptions were the Aromouns, which represent a series of linguistic Romance isolates scattered across the Balkans. These isolates appeared to be strongly influenced by drift, but no dedicated test was done to specifically address the level of gene flow between the isolates and the neighbouring populations.

**Setting up a study of Italian isolates**

Among European countries, Italy can be an optimal model area for studies of genetic diversity at micro-geographic level, as its populations host a remarkable cultural and environmental heterogeneity, supported by an extensive documentation on their history and archaeology (Capelli, 2007; Capelli et al., 2007). At the same time, Italy counts a large number of populations with a history of cultural isolation, a feature related to the complex and heterogeneous migratory patterns which have characterized this country since prehistoric times, and which have left an evident signature represented by the presence of numerous linguistic minorities (Salvi, 1975; Tosò, 2006). At present, this is reflected in the Italian legislation which states that the Italian Republic safeguards twelve linguistic minorities: Albanian, Catalan, Croatian, French, Franco-provençal, Friulian, German, Greek, Ladin, Occitan, Slovene and Sardinian (art. 6 of the Italian Constitution, laws n. 482/1999 and n. 38/2001; decree of the President of the Italian Republic n. 345/2001).

A number of linguistic Italian minorities has been previously investigated from a biodemographic point of view, including German speaking Mocheni (Pettener et al., 1994), the Provençal of the Italian Alps (Biondi et al., 2005), and the Arbëreshe (often defined as an Albanian-speaking ethnic minority) of Basilicata (Pettener, 1990; 1995) and Calabria (Biondi et al., 1993; Tagarelli et al., 2007; Fiorini et al., 2007). These studies have shown that marriage exchange patterns are affected by geographic and linguistic barriers and result in higher endogamy, inbreeding, and subdivision into subpopulations. Furthermore, these studies could be useful in complementing genetic studies because selecting individuals bearing surnames with a monophyletic origin may help reduce the confounding effect of migration and identify founder lineages. However, only a small subset of Italian isolates has been so far investigated from a genetic point of view, with further limitations arising from the use of small sample sizes and single sources of genetic variation (generally mtDNA) (Stenico et al., 1998; Mogentale-Profizi et al., 2001; Vernesi et al., 2002; Pichler et al., 2006; Thomas et al., 2007; Borglum et al., 2007). A number of studies have looked for mutations linked to some pathologies, among which multifactorial disorders, such as nephrolithiasis in Talana village in East-Central Sardinia (Gianfrancesco et al., 2003), multiple sclerosis in Sardinia (Pugliatti et al., 2003;
The strictly scientific objectives pursued in this project are:
- to evaluate the degree of isolation of each population through the analysis of its genetic structure using both fast-evolving and slow-evolving uni-parentally inherited polymorphisms;
- to investigate the origin of the isolates by assessing their genetic continuity with the ancestral populations as identified on the basis of historical and traditional records;
- to shed light on the micro-evolutionary processes and cultural factors shaping human genetic diversity which may be relevant to gene-culture co-evolution processes;
- to understand what isolates could be more fruitfully studied in subsequent genetic analyses aimed at elucidating the genetic basis of rare and complex diseases and pharmacogenetic studies;
- to develop a genetic and bio-demographic database which will be freely accessible after the publication of the data.

A point which characterizes the projects regards the interdisciplinary approach. The main criticisms concerning studies which link biology and culture come from scholars outside the “genetic” field, who often point to the extreme oversimplification when including evidence external to genetics and the use of heterogeneous criteria for human group identification, mixing geography, linguistics, culture and history (e.g. MacEachern, 2000). In order to overcome these limitations, specialists from other disciplines, such as linguistics, archaeology and history, will be actively involved in the study.

Communication has also an important place in our project. Being concerned with the concept of (genetic) identity, similarity and diversity, heritage and origins, our study may help promote public engagement with science, and strengthen the links between science, education and research. We will focus on the implications and opportunities of the research project in terms of communication and promotion of
Fig. 1 - Map showing the geographical distribution of the Italian linguistic minorities protected by legislation of the Italian Republic (law n° 482, art. 2, year 1999). Populations studied within the projects "Isolating the Isolates: geographic and cultural factors of human genetic variation" and "Biodiversity and history of the populations from Trentino": 1= Fersina Valley (Trento); 2= Fersina Valley, german speakers (Trento); 3= Folgaria highland (Trento); 4= Luserna highland (Trento); 5= Fiemme Valley (Trento); 6= Fassa Valley (Trento); 7= Cadore Valley (Belluno); 8= Sappada (Belluno); 9= Sauris (Udine); 10= Ampezzo (Udine); 11= Garfagnana area (Lucca); 12= Lucca Plain (Lucca); 13= Benetutti (Sassari); 14= Carloforte (Cagliari); 15= North Pollino area (Cosenza, Potenza); 16= North Pollino area, Albanian speakers (Cosenza, Potenza); 17= South Pollino area (Cosenza, Potenza); 18= South Pollino area, Albanian speakers (Cosenza, Potenza).

The colour version is available online at the JASs website.
Science, through an extensive evaluation of the bio-ethical implications related to the proposed scientific research and through the development of web-based tools (including GPS interactive maps) aimed at promoting public interest in science. More specifically, we have identified three noteworthy educational outputs of our project:

- it addresses key evolutionary concepts in human evolution, offering a comprehensive description of how culture, space and time shape genetic legacy;
- it offers a chance to introduce the “making of” scientific research to the public;
- it helps understand why and how basic science can help develop applications of other disciplines (e.g., medicine and forensics).

The ethical implications related to the project will be specifically investigated, with particular attention being paid to informing the involved communities of the aims and modalities of the research. We will discuss benefits, risks and selection criteria of the study (e.g. Universal Declaration on Bioethics and Human Rights, UNESCO, 2005) within the research participants and the communities involved. The respect for privacy and cultural diversity will be addressed through the use of an appropriate consent form which guarantees anonymity of the data and protection of personal information. In this way, we intend to promote an active participation of the communities and individuals to such a research, and to favour the interaction between the working group and the communities, with the ultimate objective of applying a governance-based model. The educational web page will include a section dedicated to the ethical aspects of the research. Finally, the project will be an important occasion to promote the formation of ad-hoc national and local Committees for the approval of the ethical issues pursued in population-based projects.

An international symposium will be organized at the end of the project, aimed at giving a general presentation of the results achieved in the course of the project, to discuss their implications for other disciplines and to identify future lines of research. Sessions specifically developed for and dedicated to the media and the general public will be an integral part of this event.

In conclusion, with this project we intend to build the first genetic map of Italian linguistic isolates in Italy and help identify case studies suitable for subsequent bio-medical studies, while paying attention to interdisciplinary aspects, educational outputs and ethical issues.

Acknowledgements

The project “Isolating the Isolates: geographic and cultural factors of human genetic variation” is financed by the Ministero dell’Università e della Ricerca (PRIN - Progetti di Ricerca di interesse nazionale, 2007-2009, prot. n. 2007TYXE3X). The authors acknowledge grants from the Istituto Italiano di Antropologia (funds to G.D.B), Provincia Autonoma di Trento (project “Biodiversity and history of the populations from Trentino”, V.C.), the University of Bologna (D.P), the University of Cagliari (G.V.), the University of Rome “La Sapienza” (G.D.B.) and the University of Pisa (G.P and S.T.).

Info on the web

http://mek.oszk.hu/minerva/html/dok/italy.doc

A document by the Italian working group of the MINERVAplus survey on multilingualism.


http://www.eurac.edu/index
Portal of the European Academy Bozen/Bolzano (EURAC), with links to the Institute for Specialised Communication and Multilingualism and Institute for Minority Rights.

http://www.fobiotech.org/ geneticisolates2009/
The web site of the 4th international Meeting on Genetics of complex diseases and isolated populations (20-23 June 2009, Trieste, Italy)

http://www.isita-org.com/italian_isolates
A web site providing more information and data on the project “Isolating the Isolates: geographic and cultural factors of human genetic variation”.

http://web.uniud.it/cip/home_i.html
The web site of the International Centre for the Study of Plurilingualism of the University of Udine, with pages on International and Italian legislation and on-line resources on various aspects of linguistics, including linguistic diversity.

References


Destro Bisol et al.


Exploring biodiversity of Italian isolates


Editor, Giovanni Destro-Bisol