

A PATCHWORK OF GENES AND CULTURES: INSIGHT INTO THE POPULATION HISTORY OF SICILY AND SOUTHERN ITALY



Sarno S¹, Carta M¹, Boattini A¹, Ferri G², Alù M², Tofanelli S³, Sineo L⁴, Luiselli D¹, Pettener D¹.

¹ Lab. di Antropologia Molecolare, Dipartimento di Scienze Biologiche, Geologiche e Ambientali (BiGeA), Università di Bologna, Bologna, Italy
² Dipartimento di Medicina Diagnostica, Clinica e di Sanità Pubblica, Università degli Studi di Modena e Reggio Emilia, Modena, Italy
³ Dipartimento di Biologia, Unità di Antropologia, Università di Pisa, Pisa, Italy
⁴ Dipartimento di Biologia Animale 'G. Reverberi', Università di Palermo, Palermo, Italy

This study was supported by LANGELIN Project (FP7-Ideas-ERC2011-AdG295733), GENOGRAPHIC Project (funded by National Geographic & IBM) and MIUR PRIN 2007 and 2009 Grants to DP.

The Mediterranean Sea has been one of the most important natural hubs for the expansion of human genes and cultures since the first dispersal of modern humans out of Africa. Consequently, this area has received increasing attention by population genetic studies, but the genetic jigsaw of Mediterranean populations is far from being exhaustively resolved. Therefore, to better understand origin, migration and admixture processes shaping current human biodiversity, further studies focused on specific geographic areas are needed. According to this view, the present research aims at exploring the history of **one of the Mediterranean most important midpoints**, namely Sicily and Southern Italy, through a **two-level approach of analysis**.

MACRO-GEOGRAPHIC PERSPECTIVE

Historical and archaeological documents describe **Sicily and Southern Italy** as a melting pot where different peoples and cultures came together over time. In order to explore more in detail specific issues related to the peopling of this area, as well as to clarify its role in the Mediterranean genetic context, a total of **326** individuals from 8 different provinces (**Figure 1**) of Sicily (Agrigento - AG, Catania - CT, Ragusa - RG, Enna - EN, Trapani - TP) and Southern Italy (Matera - MT, Lecce - LE, Cosenza - CS), have been collected according to the standard grandparents criterion and deeply typed for 42 Y-SNPs and 17 Y-STRs. Results were compared with 29 additional populations from Southern Europe, North-Africa and the Levant.

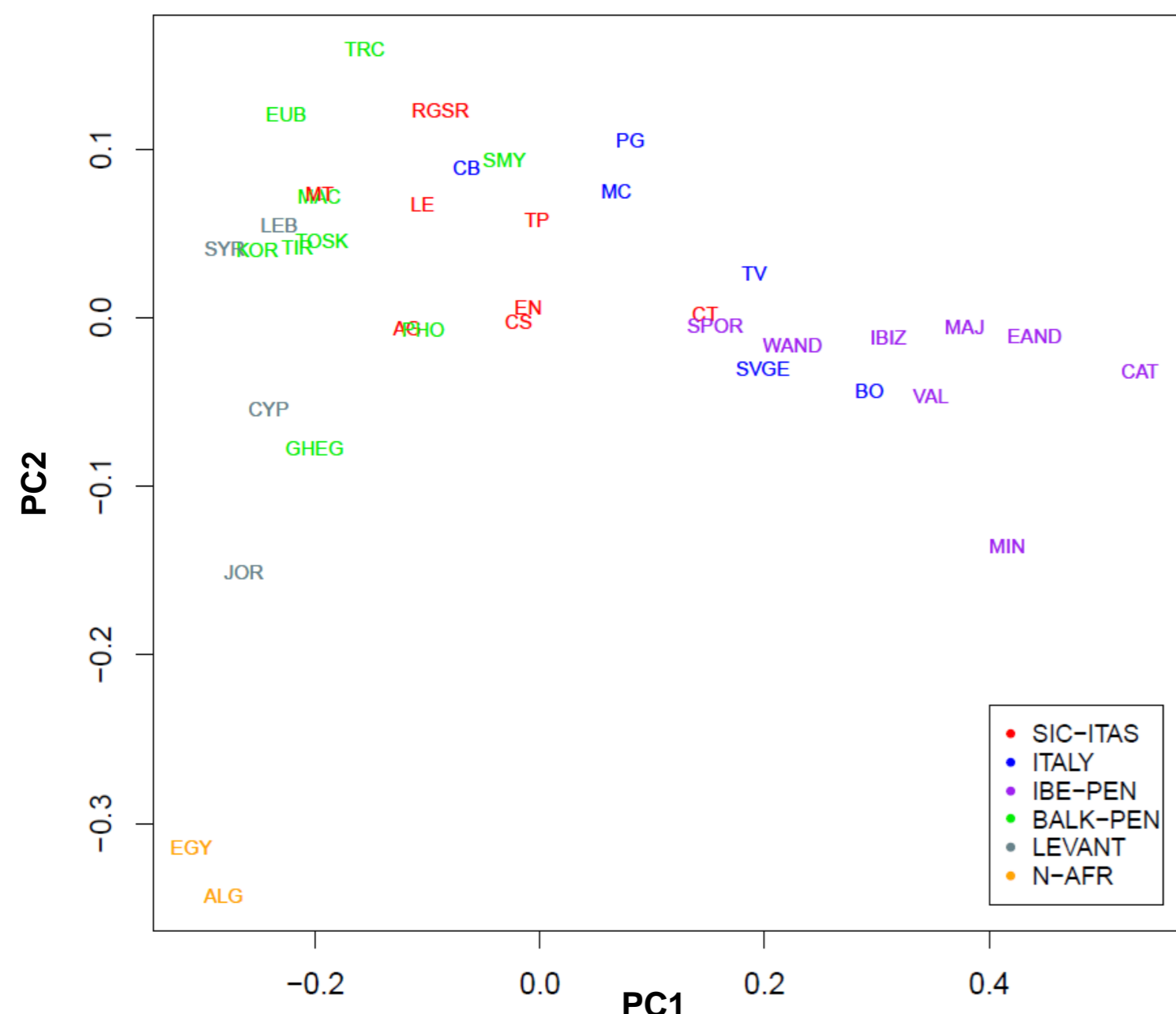


Figure 2. Plot of PC1 (66.8%) against PC2 (12.4%) based on HGs frequencies of Southern Italy dataset (red) and comparison populations from North-Central Italy (blue), Iberian Peninsula (purple), Balkan Peninsula (green), Levant (grey) and North-Africa (orange).

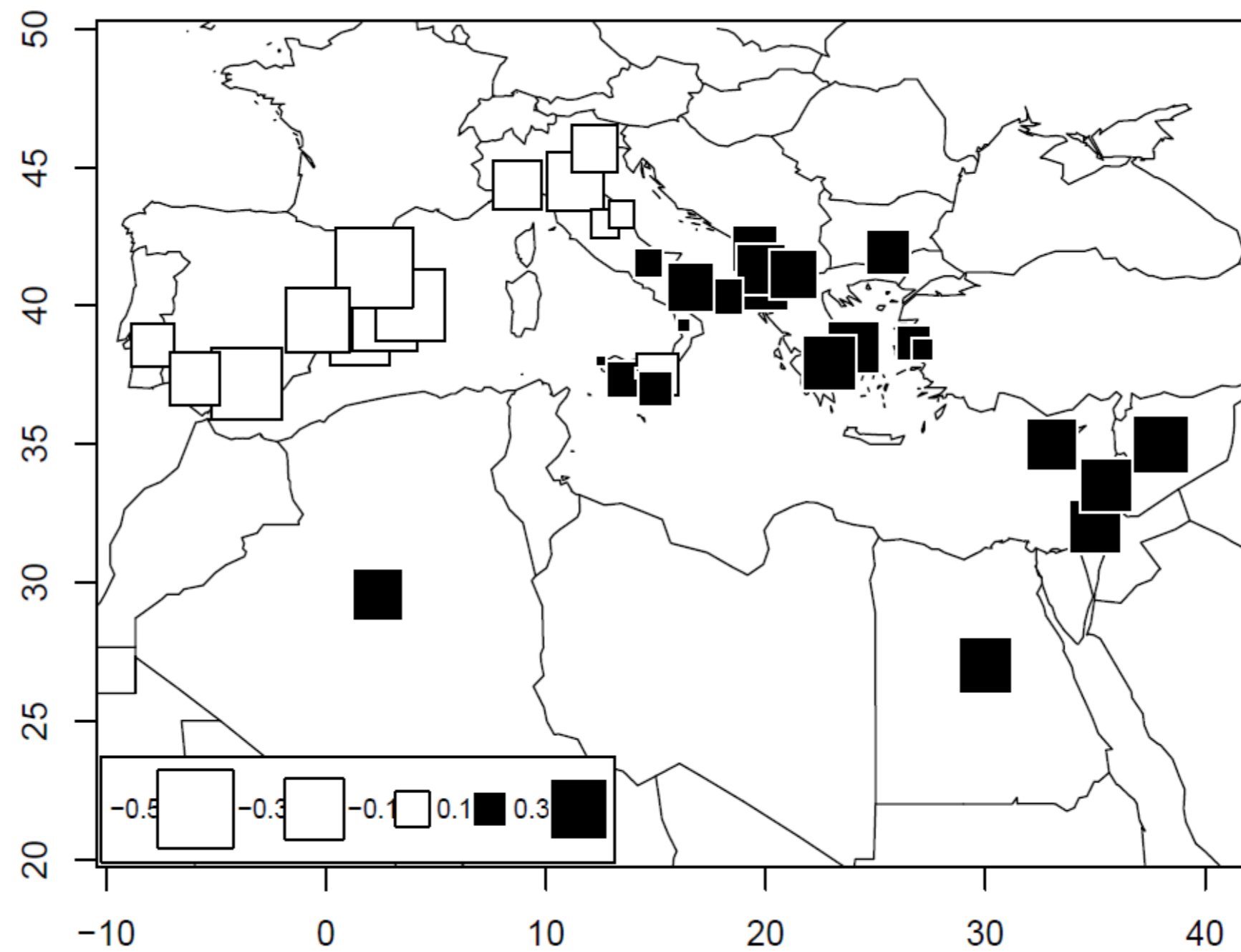


Figure 3. Spatial Principal Component Analysis (sPCA) based on HGs frequencies within the Mediterranean Basin. The first global components (sPC1) is depicted. Positive values are represented by black squares; negative values by white squares; the size of squares is proportional to the absolute value of sPC scores.



Figure 1. Geographical location of sampling points in Sicily and Southern Italy. Sampling area of Arbëreshes of Calabria (Figure 5a) and Arbëreshes of Sicily (Figure 5b) are here indicated by boxes

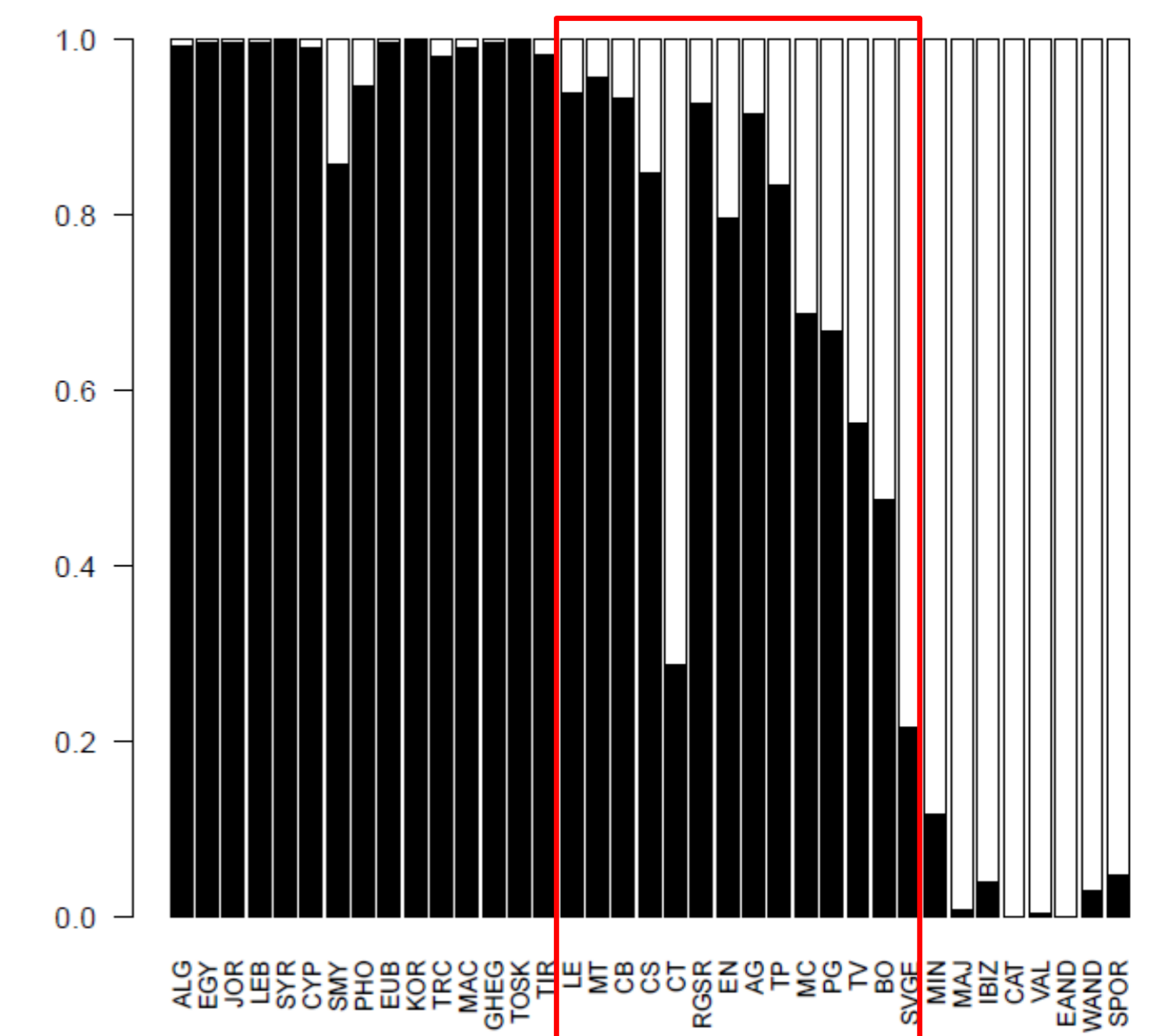
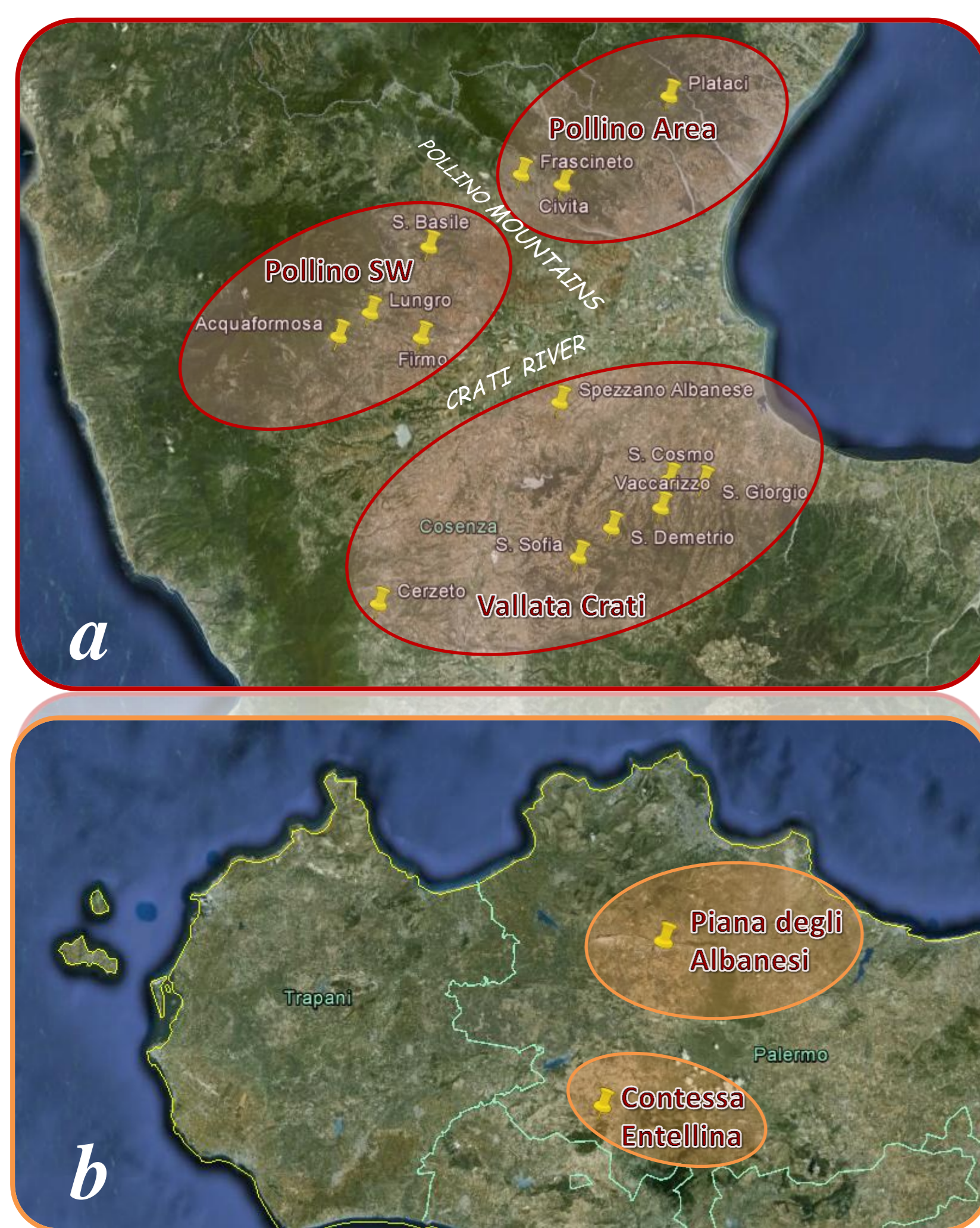


Figure 4. Discriminant Analysis of Principal Components (DAPC). The different colours represent membership probabilities to one of the two sPCA identified clusters (black= SE Mediterranean, white= NW Mediterranean).

Genetic variability within the Mediterranean Basin. Both classic PCA (**Figure 2**) and sPCA (**Figure 3**) based on haplogroup frequencies show a clear-cut geography-related differentiation within the Mediterranean Basin, between a South-Eastern group (Southern Italy, Balkan Peninsula, Levant and North-Africa) and a North-Western one (North-West Italy, Iberian Peninsula). In this context, Sicily and Southern Italy appear to be well set within the genetic variability of the South-Eastern Mediterranean group (SE), the only exception being Catania, which instead reveals a stronger affinity to the NW Mediterranean cluster.

Admixture in Southern Italy. The structure-like plot of membership probabilities (**Figure 4**) highlights the position of Italian samples (red box), which seem to represent a sort of "admixture-zone" between the two sPCA-identified Mediterranean clusters (black, white). In particular Sicily and Southern Italy seem to be more affected by different migration processes occurred from the South-Eastern Mediterranean side.

Figure 5. Geographical location of Arbëreshë groups sampled in Calabria, Southern Italy (a) and Sicily (b). The Calabrian Arbëreshë communities were grouped according to their geographic position and taking also into account previous results on the bio-demographic sub-structure of the area (Tagarelli et al 2007)



MICRO-GEOGRAPHIC INSIGHTS

Looking more deeply within specific migratory events linking Sicily and Southern Italy with the Balkan Peninsula, human isolates can be used as an ideal model to clarify the recent genetic relationships between these two areas of South-Eastern Mediterranean side.

At this purpose, Sicily and Southern Italy host one of the largest ethno-linguistic groups of Italy, the **Albanian-speaking Arbëreshes**, whose origins are related with different migration waves, either originated directly from Albania (Toskeria) or arrived in Italy after intermediate steps in Greece (Peloponnese), occurred between the 15th and 16th centuries. Being the result of well-known population movements and having preserved a certain degree of isolation to the subsequent confounding events of admixture, these linguistic isolates offer a powerful tool to more deeply explore the historical and demographic processes shaping human current biodiversity, as well as to clarify the micro-evolutionary effects of socio-cultural vs. geographic factors.

A micro-geographic approach aimed at comparing different linguistic isolates with their geographic neighbors (Sicily and Southern Italy), as well as with putative source groups (Albania and Greece), has been thus used. A total of **373** individuals from 3 Arbëreshë groups of Southern Italy (**Figure 5a**) and 2 of Sicily (**Figure 5b**), and from the Albanian populations of Tosks and Gheghs, has been typed for 42 Y-SNPs and 17 Y-STRs and compared with Italian and Greek genetic landscapes.

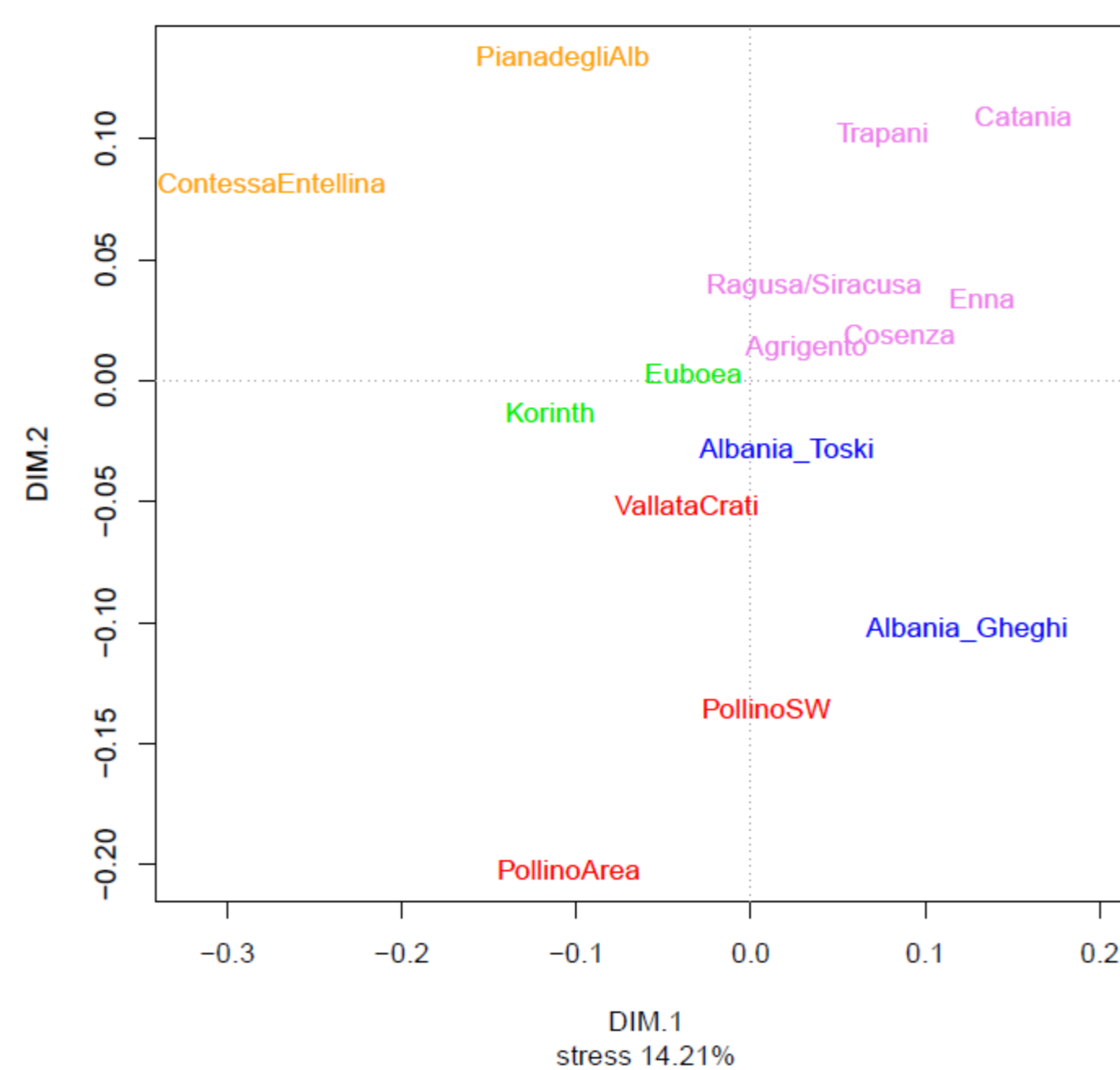


Figure 6. Multi Dimensional Scaling based on STRs distances. Population colours code: Arbëreshë of Calabria (red), Arbëreshë of Sicily (orange), comparison populations from Calabria (Southern Italy) and Sicily (violet), Albania (blue) and Greece (green).

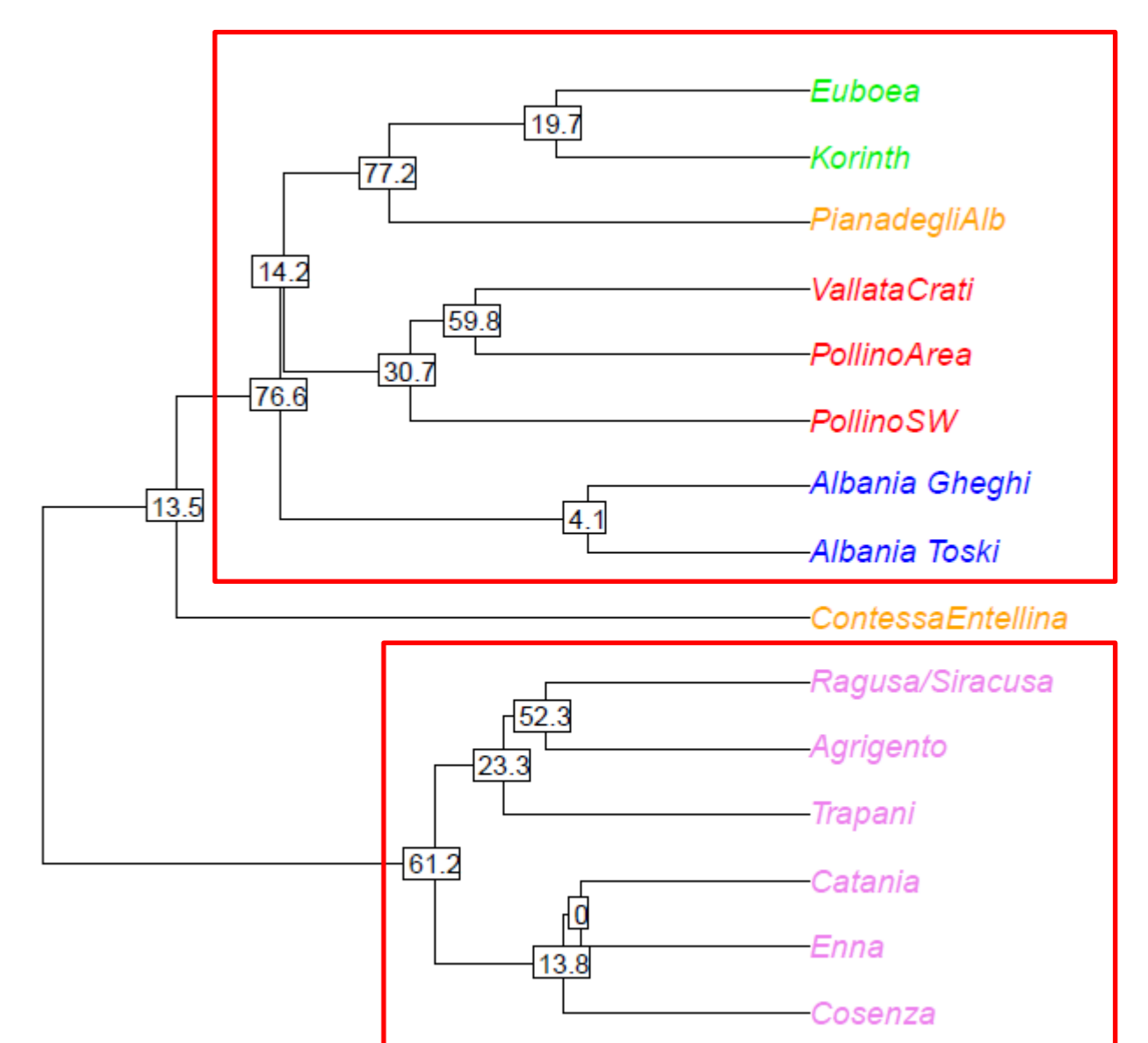


Figure 7. Hierarchical Clustering (1,000 bootstraps) computed on STRs distances. Population colours code: Arbëreshë of Calabria (red), Arbëreshë of Sicily (orange), comparison populations from Calabria (South-Italy) and Sicily (violet), Albania (blue) and Greece (green). Bootstrap values are reported in percentage for each nodes of the tree.

Relationships between genetic and linguistic isolation. From a micro-geographic point of view, Italian and Balkan samples appear to be well distinguished by the two dimensions of the MDS plot (**Figure 6**), also clustering in clearly separated groups (respectively an Italian - bootstrap 61.2% - and a Balkan one - bootstrap 76.6%, **Figure 7**). Interestingly, Arbëreshë communities fall mostly within the Greek-Albanese cluster, thus confirming their shared genetic ancestry with modern Balkan populations. However, not completely overlapping histories can be found respectively for Arbëreshë of Calabria (South Italy) and Arbëreshë of Sicily. In fact, the two Arbëreshë groups are well detached along the second dimension of the MDS plot, also showing differential similarities patterns within the hierarchical clustering analysis. The genetic history of each Arbëreshë group thus seems to be the complex result of distinct founding events, coupled with different degrees of admixture with surrounding populations. As a result, these populations have conserved at different levels their ancestral genetic diversity along with their cultural (linguistic) features.

CONCLUDING REMARKS

Our findings on the genetic structure within the Mediterranean Basin show a significant geographic differentiation along the longitudinal axis between a SE and a NW group, the Italian Peninsula occupying an intermediate position among them. In particular, Sicily and Southern Italy appear to be well set in the genetic context of the first group (SE), suggesting a shared genetic background with the Balkan Peninsula, probably resulting from common or partially overlapping patterns of diffusion within the Mediterranean from North-Africa and the Levant. However, when a more regional scale is considered, Italy and the Balkans show higher complexity and distinct patterns of genetic variation. The Balkan genetic ancestry appears clearly conserved in almost all the Arbëreshë communities, even though different genetic histories, related to different origins and migration waves, as well as to different interplays between socio-cultural and geographical factors, appear depending on the Arbëreshë group considered. Further investigations on Sicilian and Southern Italian historical strata and genetic ancestries, by evaluating also the maternal contributes (mtDNA), will be the next step of our research, in order to obtain an integrated and complementary view of the genetic history of such a key area of the Mediterranean Basin.