

# **PRIN2020 - Crossing the Sea: ancient and modern human genomes to study the evolutionary dynamics of Sicily, Sardinia, and Corsica**

Supervisor: **Prof. Marco Sazzini**

## **Research Project**

In the last years the technological improvements which allow to characterize the whole genome sequence (WGS) of human subjects, along with the development of appropriate inferential and statistical methods for data analysis, have greatly increased our ability to precisely infer past population structure, demographic processes, and selection signatures that characterize the present-day human genomic variability. Furthermore, the more recent possibility to sequence and analyze paleogenomic data also from ancient human remains has additionally provided a powerful means to further dig into the demographic history and evolutionary dynamics experienced by human populations.

To date, several studies have demonstrated that inferring the complex demographic and evolutionary events experienced by the ancestors of present-day Italians corroborates the understanding of the main population processes having contributed to the formation of the European gene pool. However, the contribution of past population dynamics having involved human groups from the three main islands of the western Mediterranean Sea – i.e., Sicily, Sardinia, and Corsica – remained relatively unknown, especially due to the paucity of paleogenomic data from these geographical areas. Despite that, these islands are characterized by a variety of archaeological records, material culture and traditions, and represented an historical meeting place for several populations, having witnessed a variety of migrations since the first human colonization of the European continent. Therefore, the analysis of modern and ancient whole genome data from individuals/remains from these Mediterranean islands may provide fundamental evidence to disentangle past population processes, identify ancient causes for present-day genetic structure, and highlight genetic relationships with surrounding populations, which would allow researchers to significantly improve the knowledge about the human past.

In this context, the MIUR-PRIN2020 Project “*Crossing the Sea*” is aimed at shedding new light on the colonization processes and evolutionary dynamics experienced by human groups from the three largest islands of the western Mediterranean Sea, through the analysis of ancient and modern genomic variability. Within the framework of this project, the research activities will be specifically focused on the analysis of the high-coverage whole genome sequencing data that will be generated by the Project for a representative panel of modern individuals from Sicily, Sardinia, and Corsica. The

modern genetic variation will be then compared with ancient genomes obtained from human remains covering a wide temporal scale and a variety of cultural groups, with the aim to reconstruct the genomic history of the populations from these islands and to explore the genetic determinants of local biological adaptations evolved by their ancestors in response to peculiar climate conditions, diseases, dietary habits and/ or cultural practices.

## **Activity Plan**

All the activities will be conducted according to those scheduled within the PRIN2020 project “*Crossing the Sea*” and will be organized according to the following specific tasks:

### **Task 1: Bioinformatic analyses of raw whole genome sequence data**

Quality of sequence reads generated by the project will be evaluated through fastQC to exclude low quality records. Filtered reads will then be aligned to the human reference genome (hg19) using the BWA mem algorithm. The GATK best practice workflow will be applied to generate high quality BAM alignments and to call single nucleotide variants (SNVs) and short indels.

### **Task 2: Population structure and admixture analyses**

Ancient and modern genetic variation of the examined Sicilian, Sardinian and Corsican populations will be described and set into a broader geographical context by creating a merged dataset composed by the project data and other genomic data available for Mediterranean and Italian ancient and modern populations. Fine-scale apportionment of genetic variation within and between populations will be evaluated through Principal Component Analysis and clustering approaches, such as ADMIXTURE and Chromopainter/fineSTRUCTURE algorithms. Similarities among diachronic samples within the same island and with respect to other Italian and Mediterranean populations will be assessed by means of the outgroup  $f_3$  statistic, as well as by calculating  $D$ , admixture  $f_3$  and  $f_4$  statistics, which will allow us to formally test for specific admixture events.

### **Task 3: Evolutionary genomics analyses**

A pipeline of analyses based on the combination of multiple selection statistics, algorithms aimed at reconstructing networks of interacting genes and machine learning approaches will be developed to detect subtle genomic signatures associated to soft selective sweeps and polygenic adaptation events evolved by the ancestors of present-day Sicilian, Sardinian and Corsican populations in response to selective pressures different with respect to those of peninsular Italians due to the peculiar Mediterranean environments they have long inhabited. This will enable the project to exhaustively investigate for the first time the adaptive history of the populations under study by testing adaptive models that are more realistic for the human species than those traditionally investigated.