Ancient genomic signatures in the Italian population

S. Aneli(1, 2, 3)[†] , F. Montinaro(3), A. Raveane(4, 3), A. Angius(5), G. Athanasiadis(6), A. Baali(7), S. Barlera(8), G. Boncoraglio(9), F. Brisighelli(10), N. Brucato(11), I. Cardinali(12), M. Cherkaoui(7), C. Bycroft(3), F. Cucca(5), AM. Di Blasio(13), C. Di Gaetano(1, 2), JM. Dugoujon(11), G. Birolo(1, 2), V. Grugni(4), G. Hellenthal(14), T. Kivisild(15), H. Lancioni(12), Melhaoui(16), M. Metspalu(17), A. Mulas(5), S. Myers(3), L. Pagani(17), S. Parolo(4), Pascali(10), P. Paschou(18), L. Pereira(19), Peyret-Guzzon(3), A. Piazza(1, 2), F. Ricaut(11), G. Stamatoyannopoulos G(20), M. Zoledziewska(5), A. Torroni(4), O. Semino(4), G. Matullo*(1, 2), A. Achilli*(4), A. Olivieri*(4), C. Capelli*(3)

(1) Italian Institute for Genomic Medicine (IIGM)
(2) University of Turin, IT
(3) University of Oxford, UK
(4) University of Pavia, IT
(5) University of Cagliari, IT
(6) University of Aarhus, DK
(7) University of Marrakesh, MA
(8) IRCCS Mario Negri, Milan, IT
(9) IRCCS Istituto Neurologico Carlo, Besta, Milan, IT
(10) University Sacred Heart Rome, IT
(11) University of Toulouse, FR
(12) University of Perugia, IT
(13) Istituto Auxologico Italiano, Milan, IT
(14) University College London, UK
(15) University of Cambridge, UK
(16) University of Oujda, MA
(17) University of Tartu, EST
(18) University of Alexandroupolis, GR
(19) University of Porto, Portugal
(20) University of Washington, US

෬ක

[†] Email: serena.aneli@gmail.com

Motivation

The genetic make-up of modern-day European populations has been under scrutiny in geneticist's, anthropologist's and archaeologist's communities. Early studies attempted to reconstruct the great migrations and the events of the past that shaped modern-day patterns of genetic variation, by considering only the genetic differences among present-day people. However, in the last few years, the possibility to extract and sequence DNA from fossils have opened an unprecedented window into the major demographic events contributing to human genetic variation. It has emerged that, in Europe, beside the contributions of the early hunter-gathering populations, major genetic components can be traced back to the Neolithic and

Bronze Age. Within Europe, Italy, due to its position in the center of the Mediterranean basin, is an ideal population to recover the genetic signatures of past continent-wide demographic events, offering the opportunity to complement and enrich the scenario depicted by ancient DNA studies.

Methods

In order to characterise the genetic variability of the Italian population and the impact of episodes of migration and admixture which occurred within and around Europe, we assembled a comprehensive genome-wide SNP dataset which included 1,589 individuals representing all of the 20 Italian administrative regions, data from 140 world-wide reference populations, for a final dataset of 4,852 samples. We added 63 representative Eurasian ancient samples selected to maximise SNPs content and past diversity.

We performed principal component (PCA; Patterson et al., 2006) and ADMIXTURE (Alexander et al. 2009) analyses, projecting the ancient individuals onto the components inferred on modern individuals. To disentangle the main components of genetic variations in modern Europeans as the combination of ancient samples, we applied the CHROMOPAINTER (CP; Li and Stephens, 2003; Lawson et al., 2012) pipeline. This software, using Hidden Markov Models, "paints" each haplotype of a sampled "recipient" as a combination of the best matching DNA segments from a set of sampled "donors". Basically, we generated a "painting profile" of ancient and modern samples using only the latter as donors. Subsequently, we applied the mixture fit NNLS-based (Non-Linear Least Squares) analysis, which deconstructs the "painting profile" inferred by CP as the combination of defined sources. In our analysis, we reconstructed the profiles of modern samples as the combination of the ancient ones. We analysed two sets of ancient samples separately to characterise extant variation in relation to ancient (ultimate) and recent (proximate) sources. We considered a combination of Hunter-Gatherer and Neolithic samples as the ultimate sources and three geographically and temporally closer ancient samples as proximate ones.

Results

PCA placed Italians in between WHG, Anatolian Neolithic and Steppes Bronze Age populations, in agreement with previous observations of other European samples. The Italian samples, except for Sardinians, also appeared close to post-Neolithic samples. A similar pattern was also shown by the ADMIXTURE analysis on both ancient and modern samples. A finer dissection of the demographic processes shaping European and Italian genetic diversity has been obtained from the CP/NNLS pipeline. Most of the ultimate and proximate sources resulted associated with longitude and/or latitude across Europe, with differences in ancestry composition highlighted between North and South Italy. The power of this combination of techniques allowed us to disentangle the main ancient components of the Italian genetic structure. In fact, the unique genetic make-up of present-day Italians has been shaped by the complex history of migrations and admixture experienced by the Italian population. As a consequence, the genetic structure within the Italian population deserves further studies, in order to be properly considered also in disease and pharmacogenetic association studies.