Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

SCIENZE DELLA TERRA, DELLA VITA E DELL'AMBIENTE

Ciclo 33

Settore Concorsuale: 05/B1 - ZOOLOGIA E ANTROPOLOGIA

Settore Scientifico Disciplinare: BIO/08 - ANTROPOLOGIA

FROM SINGLE LOCI TO GENE NETWORKS: ADAPTATION GENOMICS OF COMPLEX TRAITS IN HUMAN POPULATIONS

Presentata da: Paolo Abondio

Coordinatore Dottorato

Giulio Viola

Supervisore

Donata Luiselli

Esame finale anno 2021

Acknowledgements

Many thanks to Prof. Donata Luiselli for the opportunity to pursue this PhD, for the endless inputs and ideas, and for the unceasing confidence that I would do everything well.

I will always be grateful to Prof. Marco Sazzini, Prof. Alessio Boattini, Prof. Davide Pettener, Stefania Sarno, Sara De Fanti and Cristina Giuliani for their constant support and collaboration during these three years.

A sincere thanks to my fellow PhD colleagues Maria Fungomeli, Andrea De Giovanni, Giorgia Modenini and Adam Andrews for sharing this experience with me: either at the beginning or at the end of your journey, never cease looking for surprises along the way.

Thanks to Prof. Aida Andres, PI of the Evolutionary Genomics Group, and to the brilliant people that I had the opportunity to meet while a visiting student at the UCL Genetics Institute, University College London: Joshua Schmidt, Harvinder Pawar and Simran Rai. I have had a wonderful time!

Thanks to my mom, my family and all of my historical friends from Parma, as well as my new friends in Bologna, for being there during these long years of study and research.

To Marco (and Kali): you had to endure so much craziness and you still supported me and my sometimes absurd ideas, which was a great relief. Thank you!

Abstract

Over the course of the last two decades, the quick advancement of technological innovations in the field of Molecular Genetics has exponentially enhanced the ability to produce immense amounts of genomic data in a very short time and at a fraction of the cost that was necessary to produce a single human genome at the beginning of the current millennium. This incessant progress has also changed the methods with which human genomic data are stored and handled, as well as the power of the analyses that can be carried out, allowing for a shift from the painstaking study of single loci linked to very specific phenotypes, to the possibility of dissecting the genomic background of entire populations by managing thousands of high-quality whole genome sequences.

In the following chapters, three interlinked arguments are presented after an introduction to the discipline of Molecular Anthropology and an overview of the dispersal of *Homo sapiens* from Africa, the peopling of Europe and the migratory processes that involved the Italian peninsula.

In Chapter 2, a study on the evolutionary dynamics of the Italian population, using high-quality whole genome sequencing data, assesses the contribution of distinct modern and ancient migratory waves to the genetic makeup of Northern and Southern Italy, as well as the existing relationship between the demographic structure of modern Italians and contemporary populations of the Euro-Mediterranean basin. Moreover, the Italian genomes are tested for signatures of positive and balancing selection at single nucleotide polymorphisms, which highlight the impact of diverse local environmental pressures on the inhabitants of the peninsula. Finally, a gene network analysis approach based on the results of the previously mentioned scans for selection reveals a climate-related adaptation in Northern Italy, a pathogen-UV driven adaptation in Southern Italy and the existence of selective pressures on pleiotropic genes modulating longevity. These results have been published in: *Sazzini M, Abondio P, Sarno S, Gnecchi-Ruscone GA, Ragno M, Giuliani C, De Fanti S, Ojeda-Granados C, Boattini A, Marquis J, Valsesia A, Carayol J, Raymond F, Pirazzini C, Marasco E, Ferrarini A, Xumerle L, Collino S, Mari D, Arosio B, Monti D, Passarino G, D'Aquila P, Pettener D, Luiselli D, Castellani G, Delledonne M, Descombes P, Franceschi C, Garagnani P.* ***Genomic history of the Italian population recapitulates key evolutionary dynamics of both Continental and Southern Europeans.*** *BMC Biol. 2020 May 22;18(1):51. doi: 10.1186/s12915-020-00778-4. PMID: 32438927; PMCID: PMC7243322.*

Pertaining the genetic contributions to the achievement of older age, Chapter 3 presents an exploration of the current literature around apolipoprotein E (*APOE*) and provides a perspective on the attainment of extreme longevity as a healthy phenotype. In particular, the worldwide genetic variability of three isoforms of *APOE* is analyzed, revealing the relationships between human genetics and local environmental variables, as well as the evolutionary mechanisms that maintain deleterious

variants in modern human populations. Moreover, a focus on the Italian peninsula shows an increasing North-to-South gradient for the protective isoform, *APOE2*, while an analysis of ancient Eurasian samples provides a provisional framework to the dynamics of *APOE* in Europe. These results have been published as a review in: *Abondio P, Sazzini M, Garagnani P, Boattini A, Monti D, Franceschi C, Luiselli D, Giuliani C.* ***The Genetic Variability of APOE in Different Human Populations and Its Implications for Longevity.*** *Genes (Basel). 2019 Mar 15;10(3):222. doi: 10.3390/genes10030222. PMID: 30884759; PMCID: PMC6471373.*

Chapter 4 introduces a study, currently undergoing revision, that builds on the concepts and methods presented in Chapter 2, by focusing on the genetics of Greco-speaking communities of the Aspromonte region of Calabria, in Southern Italy. These linguistic and geographic isolates are characterized by private ancestral components, closer affinity with modern populations form the Eastern Mediterranean and higher genetic contributions from ancient Neolithic clusters of the Caucasus and the Near East. Analyzing the most differentiated variants that characterize the Aspromonte populations in a comparison with other Southern Italian groups, peculiar biological functions are highlighted relating to neurological pathways, in accordance with recent studies on Alzheimer's and Parkinson's disease in the Calabrian area that show a higher incidence of rare familiar forms of these diseases in close-knit and isolated communities. These results have been published in: *Sarno S, Petrilli R, Abondio P, De Giovanni A, Boattini A, Sazzini M, De Fanti S, Cilli E, Ciani G, Gentilini D, Pettener D, Romeo G, Giuliani C, Luiselli D.* ***Genome-wide variability of Calabrian Greek-speaking groups reveals ancient events and long-term isolation in the Aspromonte mountain area of Southern Italy.*** *Sci Rep. 2021 Feb 4;11(1):3045. doi: 10.1038/s41598-021-82591-9. PMID: 33542324; PMCID: PMC7862261.*

As highlighted in the concluding remarks, this thesis showcases an array of up-to-date investigative methods, that allow for the analysis of both whole genome sequences and SNP-array data at different levels of detail (from single nucleotide polymorphisms to single genes, to gene networks, to biological pathways). Moreover, significant original contributions to knowledge in the fields of Molecular Anthropology and Human Population Genomics are highlighted. In particular, this thesis provides novel and exciting insights into the demography and adaptation events of the Italian population, in relation to the historical and pre-historic migratory waves that contributed to the peopling of Europe and the Mediterranean. Moreover, it focuses on the single APOE gene, exploring its worldwide variability, evolution and its contribution to a very complex phenotype, such as human longevity. Finally, it explores the demo-evolutionary context of isolated populations in Southern Italy, revealing unexpected neurological pathways as drivers of local differentiation in an innovative analytical perspective.

TABLE OF CONTENTS

Chapter 1: General background

1.1. The genetic bases of human biodiversity

Molecular Anthropology has emerged as a relatively novel transdisciplinary field that implements powerful methodologies developed in multiple interconnected study areas, such as Molecular Biology, Genetics and Genomics, to disentangle the patterns of variation that exist within and across modern and ancient human groups. The populations under scrutiny are described through an anthropological perspective, supported by the most recent technological, statistical and bioinformatic advances in the determination and interpretation of molecular variation. Accordingly, these studies perform a detailed analysis of genetic and genomic data, scrutinizing them not only in the context of human biodiversity, but also considering population histories and their socio-cultural implications. Such analytical approach allows the investigation of present-day patterns of human genetic variation and the associated demo-evolutionary events that led a population from its origins to its current bio-socio-cultural diversity. The general aim of Molecular Anthropology is the understanding and description of the evolution of our species, from the origins of *Homo sapiens* up to the determination of the ancient and modern migration paths that supported the movement of modern humans across all continents.

The first investigations carried out with the aim to explore the genetic basis of human biodiversity relied on the analysis of the distribution of allele frequencies across human groups through "classical markers", or proteins, such as the ABO and Duffy antigens that define blood group systems (Cavalli-Sforza *et al.* 1994). It was, however, the development of the Polymerase Chain Reaction (PCR) method that completely transformed the field of Molecular Biology and introduced modern Molecular Anthropology, by shifting the nature of the object under study from a few polypeptides to an ever-increasing number of DNA sequences. A "genetic marker" is, then, defined as a genomic region, denoted by discrete variability patterns, that presents allelic variants with different frequencies in distinct human populations (Jobling *et al.* 2014).

Genetic markers can be classified according to their location and pattern of inheritance (**Figure 1.1.1**). In fact, the cell nucleus contains most of the genetic material, since there are 22 pairs of homolog chromosomes, the "autosomes", and a pair of sexual chromosomes, X and Y; however, the mitochondrial organelle has its own small genome (mtDNA) and its own evolutionary trajectory, partially overlapping that of the nuclear genome. "Autosomal markers" are located on the homolog chromosomes and are involved in recombination during meiosis, giving rise to novel combinations during sexual reproduction so that both parents contribute equally to the genetic makeup of the next generation; the Y-chromosome and mtDNA are called "uniparental markers", because they are

transmitted without recombination from a single parent to the offspring (specifically, the Y-chromosome from fathers to sons, while mtDNA passes from mothers to the entire offspring) and their sole source of change is mutation (**Figure 1.1.1**).

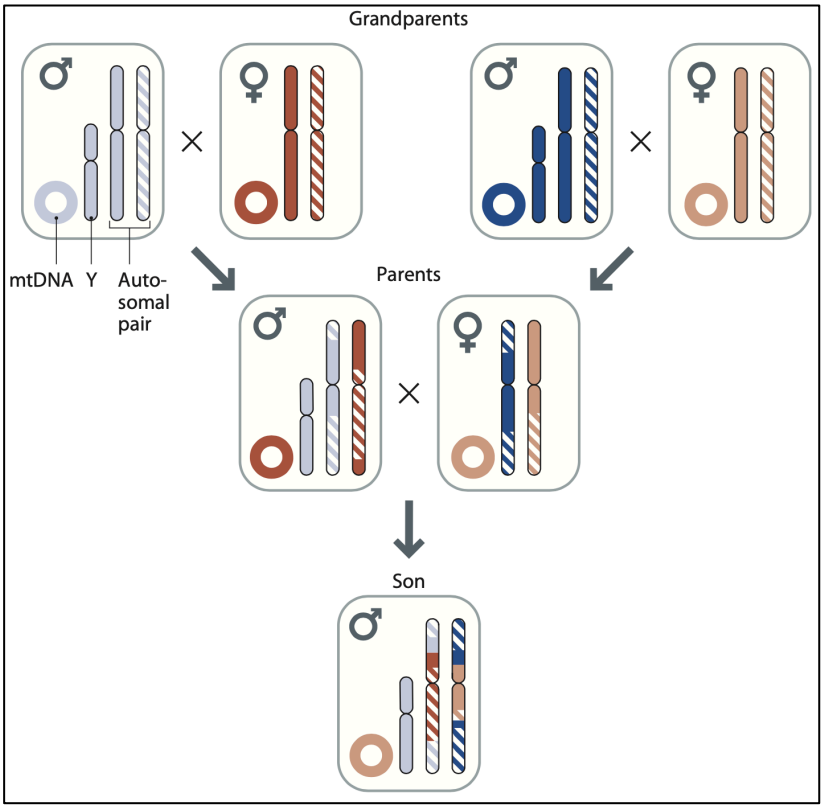


Figure 1.1.1. Genetic elements and their modes of inheritance. Mitochondrial DNA (mtDNA) passes without recombination from mothers to the offspring (both male and female). Y-chromosome goes unchanged from fathers to sons. Autosomal chromosomes are subjects of recombination at each successive generation. From Jobling *et al.* (2014).

1.1.1. Uniparentally-inherited markers

The features of mtDNA and Y-chromosome (uniparental inheritance and absence of recombination) allow their use for the reconstruction of maternal and paternal phylogenies through the application of the coalescence approach. This model, theorized in the early Eighties, but only recently structured (Rosenberg and Nordborg, 2002), is based on the concept that all variation detectable in any given population is inherited from a common ancestor (called the Most Recent Common Ancestor, MRCA). The patterns of variation in contemporary human groups can be used to detect the histories of different genetic lineages by going back in time for generations, until the MRCA. The model is fully functional only with genomic elements that do not recombine, since the mixing of chromosomal segments over generations hinders the ability to distinguish the evolutionary contributions of different genetic lineages (Underhill and Kivisild, 2007).

Uniparental markers are also assumed to be neutral to selection, a useful characteristic that allows to infer the impact of founder effects, gene flow, migration and cultural practices without their

signals being hidden by adaptive events. In particular, variation at these markers is particularly sensitive to random genetic drift, which facilitates geographic clustering and divergence of genetic lineages to assess the origins of present-day human population diversity (Perićić *et al*. 2005). However, the notion that uniparental markers have never been subjected to natural selection and that they do not provide direct advantage to the survival and reproductive success of the individual represents a mere simplification, which has been repeatedly challenged both for the Y chromosome (Wang et al. 2014; Wilson Sayres *et al.* 2014; Jobling and Tyler-Smith, 2017; Fan and Silber, 2019; Lau, 2020) and mtDNA (Otten and Smeets, 2015; Schtolz and Mishmar, 2019; Lawless *et al.* 2020; Stewart and Chinnery, 2021).

1.1.2. Neutral and adaptive autosomal markers

Another breakthrough in Molecular Anthropology was provided by the completion of the Human Genome Project (HGP), an international endeavor to produce a faithful first sequence of the human genome and a detailed genetic and physical map of human chromosomes (Collins *et al*. 2003). The publication of a high-quality whole genome sequence, coupled with the development of high-throughput sequencing technologies, enabled research groups to considerably speed up their scientific work. Although genome-wide data for autosomal loci are unsuited for genealogy reconstruction due to extensive recombination, their high number improves statistical power for testing demographic hypotheses about the population dynamics of human groups (Wilkins, 2006).

The shift from a locus-based perspective to a genome-wide approach greatly increased the level of resolution of Molecular Anthropology research: the simultaneous analysis of millions of single nucleotide polymorphisms (SNPs) provided more clarity on the origins, adaptive evolution and demographic history of human populations, as well as their genetic variation (Stoneking and Krause, 2011).

Autosomal genetic variants also qualify as structures on which natural selection can act and through which human population diversity is shaped, introducing the study of adaptive and disease-causing phenotypic traits: non-neutral markers provide the opportunity to test whether distinct environments and socio-cultural contexts have exerted selective pressures on the human genome. In fact, as modern humans spread out of Africa, prolonged exposure to new environments induced adaptive events useful to cope with novel natural conditions. For instance, climate-related stressors, such as UV radiation and temperature changes, coupled with exposure to new pathogens, may represent the main forces of adaptation in human populations (Hancock *et al*. 2008; Jablonski and Chaplin, 2010; Fumagalli *et al*. 2011; Sazzini *et al*. 2014; Quagliariello *et al*. 2017). As several dietary shifts have occurred after the movement of *H. sapiens* out of Africa, accessibility to novel nutritional

resources is also thought to have imposed notable selective constraints, that were paramount in framing human variability patterns through physiological and metabolic adaptations (Luca *et al*. 2010). Moreover, the dietary and cultural changes that have been supported by significant technological advancements and globalization events in the last centuries, may have played a relevant role in rendering detrimental many of the traits favorably selected in the past, strongly impacting the health and social structures of contemporary human populations (Sazzini *et al*. 2016).

1.1.3. The massive parallel sequencing revolution

The latest technological advancements in sequencing techniques introduced further improvements after the aforementioned genome-wide revolution, with a steep reduction of costs and a high sequencing coverage (i.e., the number of reads covering a specific genomic segment) allowing the development of demanding experimental designs, such as the "1000 Genomes Project" research effort (**Figure 1.1.3.1**), which generated low- and high-coverage sequencing data for 26 human groups worldwide (1000 Genomes Project Consortium, 2015).

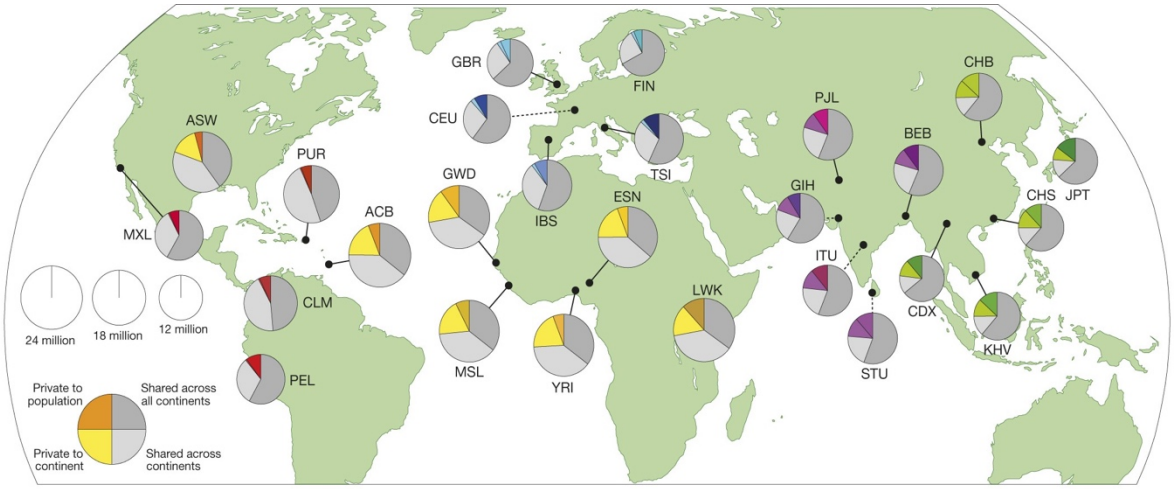


Figure 1.1.3.1. Populations of the 1000 Genomes Project. Human groups are color-coded according to their macroarea: Africa (yellow), America (red), Europe (blue), East Asia (green), South Asia (purple). From 1000 Genomes Project Consortium (2015).

High-resolution genome data also required a new theoretical framework to be developed for demographic, admixture and structure population analysis, that pushed Molecular Anthropology research to reach even finer levels of description and detail (Lawson *et al*. 2012; Hellenthal *et al*. 2014; Schiffels and Durbin, 2014; Sudmant *et al*. 2015). Thanks to the development of new massive parallel sequencing platforms with even lower sequencing costs and faster run times (**Figure 1.1.3.2**), new complete human sequences provide the opportunity to test whether populations inhabiting the same geographical area constitute discrete genetic entities, or intermediate points in an uninterrupted

genetic continuity that extends across the continents, disentangling worldwide patterns of variation (Mallick *et al*. 2016; Pagani *et al*. 2016; Malaspinas *et al*. 2016; Ebert *et al.* 2021).

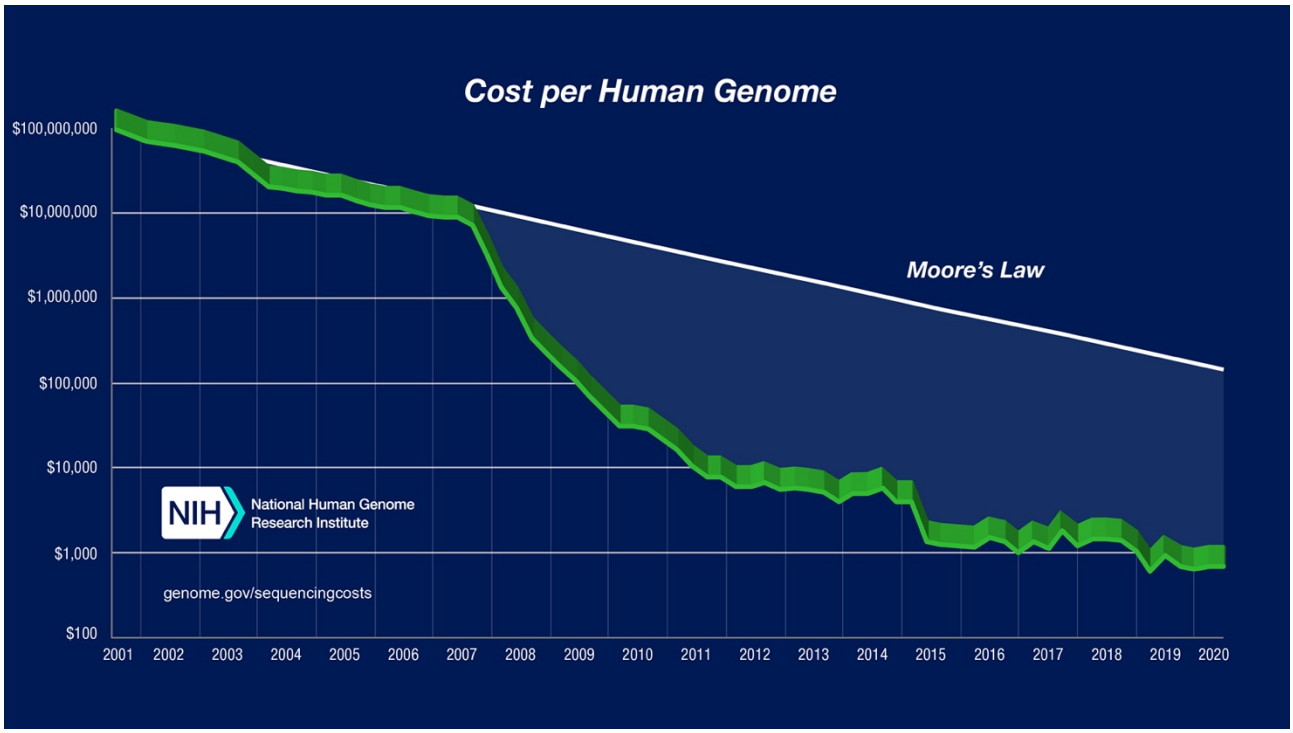


Figure 1.1.3.2. Decreasing cost per human genome over the last 20 years. The green profile highlights the average cost of sequencing a human genome since the years of the Human Genome Project. A sharp reduction can be noted when massive parallel sequencing technologies have been introduced, escaping Moore's Law applied to the theoretical reduction of costs in technology (white line). According to the empirical projection of trends (white line), in fact, the cost of technology is expected to halve every 18-24 months (or, in terms of the logarithmic scale on the vertical axis of this graph, it would reduce by 10 times roughly every 5 years). Figure taken from the National Institutes of Health website at genome.gov/sequencingtools (accessed December 10, 2020).

1.2 Origins of modern humans

1.2.1. The African dawn of the hominid lineage

Archaeological and genomic research overwhelmingly support the theory of an East African origin for the hominid clade, with genus *Homo* being the most recent branch. In fact, specimens excavated in Kenya and Ethiopia (Haile-Selassie *et al*. 2004; Suwa *et al*. 2009) date the earliest hominids to 6.0-4.4 million years ago (Mya); the most ancient remains attributed to *Australopithecus* were discovered in South Africa and dated 4-2 Mya (Falk, 2009), while more recent fossils were found from Chad to Kenya, finally, the first evidence of the genus *Homo* (*H*. *abilis*) to be documented were dated 2.8-1.5 Mya and located in the Kenya-Tanzania territory. As no other remains of the earliest members of our lineage have been uncovered anywhere else, its geographical area of origin can be reliably located in Africa (Brauer, 1978; Wood, 1987). The first species to potentially showcase a range of advantageous characteristics, such as chiefly bipedal locomotion, a larger brain

than its predecessors and the capacity to produce and use lithic artifacts, was *H. erectus*. Moreover, it was the first known human species to move throughout and outside Africa, colonizing both Asia and Europe around 1.8-1.5 Mya. The arguably small population size of the first moving human groups, coupled with limited gene flow due to their physical distance, made them susceptible to strong genetic drift, possibly instigating local adaptations to novel environments and nutritional sources (Richards and Trinkhaus, 2009). Over an extended period of time, this may have led to their differentiation into other species, such as *H. heidelbergensis* (Mounier *et al*. 2009), which fossils have been excavated across the African and Eurasian continents, *H. neanderthalensis* and *H. denisova*, which were arguably limited to Europe and Asia, and possibly *H. floresiensis* and *H. luzonensis* in restricted Asian regions (Hublin, 2009; Detroit et al. 2019; Teixeira et al. 2021). A more recent offshoot of the hominid branch, *H. sapiens* appears to have had a single evolutionary episode in Central Africa, dated to around 200 thousand years ago (Kya), as evidenced by uniparental markers on the non-recombining region of the Y-chromosome (NRY) and the mtDNA. The studies, further supported by more recent analyses including ancient DNA (aDNA) sequences and whole-genome data, traced the origin of all modern humans back to a woman who lived 140-200Kya in Central Africa, while all contemporary male individuals seem to descend from a man who lived 160-120 Kya in Africa (Fu *et al*. 2013; Poznick *et al*. 2013). It has to be noted, however, that these observations have been challenged by a Moroccan specimen from Jebel Irhoud, which was previously dated about 160Kya and has been recently re-dated to 315 Kya (Hublin *et al.* 2017; Richter *et al.* 2017), making it the oldest *H.sapiens* specimen currently in existence.

1.2.2. The Out of Africa dispersal of modern humans

Worldwide collaborative studies in Human Population Genomics, strongly supported by archaeological, cultural and linguistic evidence, theorized the possibility of a second "Out of Africa" event (the first being the one involving *H. erectus*) dated to 130-100 Kya that interested the *H. sapiens* species in the human lineage (Petraglia *et al*. 2010; Lopez *et al.* 2016). This conjecture supports the possibility of modern humans moving away from the African continent in fragmented groups, along at least two dispersion routes (**Figure 1.2.2.1**). A more widely accepted path possibly proceeded along the East African coast to Egypt and the Sinai Peninsula, crossing the Levant and extending into the Eurasian continent, up to the Americas through the North-Western passage during the last 20 Ky. At different times, consecutive and continuous waves of migration radiated from this path and supplied new genetic contributions to the populations of the European continent. The alternative "Southern route" bridged the Red Sea, traversed the Arabian Peninsula, followed the shoreline of South Asia

and finished in Australia around 45-50 Kya, where human remains even predated the most ancient fossils of *H*. *sapiens* in Europe (Petraglia *et al*. 2010; Lopez *et al.* 2016).

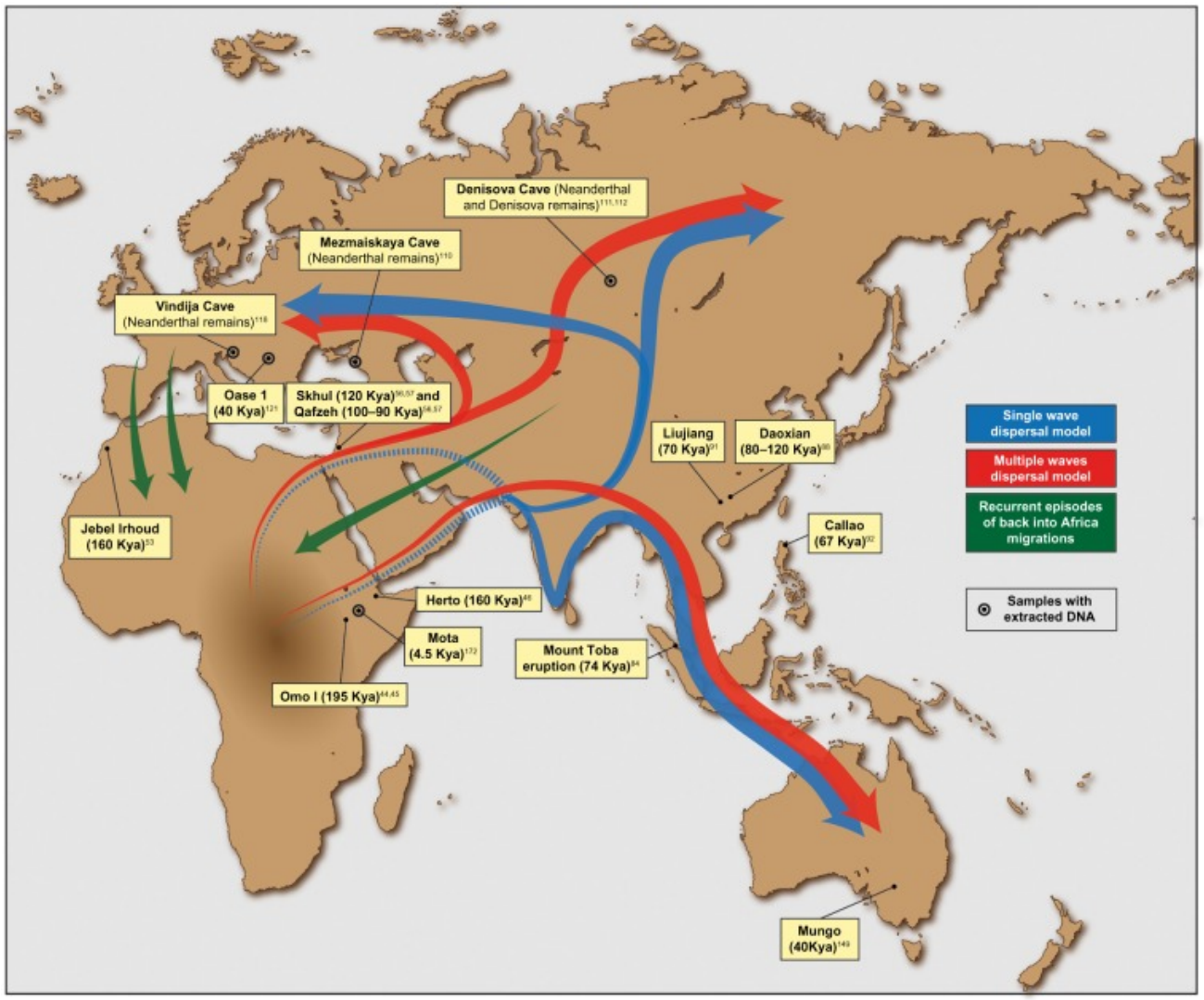


Figure 1.2.2.1. Simplified view of the Out-of-Africa theory. Two models (single-wave dispersal, in blue; multiple-wave dispersal, in red) are presented for the peopling of Eurasia and Oceania, given the most ancient human remains found across these continents. Back-to-Africa movements (in green) are also taken into consideration to highlight the continuous dynamic process of migration. From Lopez *et al.* (2016). It has to be noted that the Jebel Irhoud specimen, previously dated about 160Kya, has been re-dated to 315 Kya (Hublin *et al.* 2017; Richter *et al.* 2017).

1.2.3. Genetic insights into the peopling of Europe

In a first endeavor to disentangle the migration patterns that produced the highly complex demography of the European continent, Luigi Luca Cavalli-Sforza performed early genetic studies on "classical markers". By observing polymorphisms in the antigens that define blood groups and in the structure of immunoglobulins, he could identify differences in allele frequency that defined clines of variation across the continent, an indication of a genetic continuum through the modern populations of Europe (Cavalli-Sforza *et al.* 1994). By computing the genetic distance among populations, based on the same classical markers, he revealed that the genetic distance from the African samples increases going from Europe, through Asia, to Australia and that the non-African populations appear

more closely related to each other than to any African group. Although being inherently flawed by the exclusion of North African groups, this study validated the single origin of *H. sapiens* in Africa and the original worldwide migration of modern humans from West to East, with multiple contributions from Eurasian populations originating the ancestral European gene pool. Given the relevant African genetic component found in populations from the Middle East, the model also alluded to the possibility of more recent migrations, such as the spread of agriculture with the Neolithic Revolution from historical Mesopotamia to Europe 10-8 Kya, enhancing the genetic similarity between Africans and Europeans (Cavalli-Sforza *et al.* 1994).

In the last decade, the genetic structure of the European meta-population was better described by several studies focusing on the distribution of NRY haplogroups, which helped in defining possible migratory paths within the continent, as several relevant haplogroups (E1b1b, I, R1a, R1b) can be identified as markers useful for retracing the colonization of the European continent (Myres *et al*. 2011; Underhill *et al*. 2015). *Haplogroup E1b1b* apparently arose in East Africa between 22-40 Kya and then spread to the Levant and the Mediterranean Coast of the European continent; this hypothesis is supported by the analysis of aDNA sequences from Levantine Natufian fossils dating back 11-15 Kya, which showcase this specific haplogroup (Lazaridis *et al.* 2016). The discovery may support the previously hypothesized migrations from North Africa to the Near East and the diffusion of agriculture in Europe with the Neolithic Revolution (Lazaridis *et al.* 2016). Given that derivatives of the haplogroup can be identified along the Mediterranean coast of Europe and North Africa, but never in mainland Europe, a second more recent migratory route has been also suggested from North Africa to the Iberian Peninsula (Trombetta *et al*. 2011). The origins of *Haplogroup I* trace back to the Caucasus and Western Asia around 35 Kya, although Paleolithic remains carrying this haplogroup are found only in continental Europe and are dated 31-34 Kya (Fu *et al*. 2016). Apparently associated with the Last Glacial Maximum (LGM), it reached the Balkans sometime around 25 Kya and spread to Scandinavia, with the diffusion of the Germanic languages, only around 5 Kya (Fu *et al*. 2016). *Haplogroups R1a and R1b* are the two most widely distributed lineages in contemporary European populations. With origins extending to Western Asia, their parent haplogroup R1 has been dated 18.5 Kya (even though recent studies suggest a date for R1a of 22-25 Kya, which would predate the parent haplogroup). R1a frequency peaks in Eastern European and Scandinavian populations, then it decreases in Central Europe and is almost absent in Southern Europe. R1b is peculiarly found at the highest frequency the Basque population, a genetic and linguistic isolate located between northern Spain and southwestern France (Balaresque *et al*, 2010). The most ancient sample exhibiting the R1b haplogroup currently is the Villabruna1 specimen found in Veneto (Italy) and dated to 14 Kya (Myres *et al*. 2011; Underhill *et al*. 2015; Fu *et al.* 2016). A sub-lineage called R1b1a1a2 (also known as R-

M269), shows a peculiar distribution, since its frequency is extremely high (60- 90%) in populations of Western Europe (Spain, France, Great Britain and Wales) and it decreases towards Central Europe, Italy and Portugal, thus decreasing in frequency from West to East (Myres *et al*. 2011). Several recent aDNA-based analyses around these two haplotypes seemed to corroborate the so-called "Kurgan hypothesis" or "Steppe theory", which was introduced by archaeologist and linguist Marija Gimbutas (see Cavalli Sforza and Piazza, 1993): the studies suggest the origin of these haplogroups in the Pontic-Caspian steppes, being carried into Eastern Europe together with the Indo-European languages somewhere between 4.5 and 10 Kya (Allentoft *et al.* 2015; Haak *et al.* 2015; Mathieson *et al*. 2015; Olalde *et al.* 2018). However, other studies argue for a more careful interpretation of the results, especially related to the reconstructed time of origin for R-M269 (Busby *et al.* 2012) and the contextualization of its sublineages in Europe using genealogy-informative short-tandem repeats (Y-STR) together with single nucleotide polymorphisms (Larmuseau *et al*. 2014, Solé-Morata *et al.* 2014). In fact, analysis of Y-STR and Y-SNPs in Belgium and the Netherlands suggests a low degree of lineage diversity, consistent with a recent radiation in Western Europe and the origin of local subhaplogroups in a short period of time (Larmuseau *et al.* 2014). Similarly, a recent study indicates a Bronze-Age migration wave that introduced sublineages of S-M269 in the Basque region and is concordant with other observations of replacement, local lineage differentiation and founder effects in France and the Iberian Peninsula (Myres *et al.* 2011; Valverde *et al*. 2016; Solé-Morata *et al*. 2017; Luis *et al.* 2021). These observations suggest that R-M269 is a Mesolithic lineage in Europe, while the expansion of several of its sublineages over time would account for the local diversity of Y haplotypes in European populations (Valverde *et al.* 2016; Solé-Morata *et al.* 2017).

Interestingly, mtDNA haplogroups do not present such an extreme geographic distribution, showing a pattern with a high degree of similarity among European groups and low genetic distances. The overwhelming majority of the European mtDNAs belong to ten haplogroups (H, I, J, K, M, T, U, V, W and X) and six of them are unique to European populations: this is an indication of them possibly having risen after the separation of early European settlers from their West Asian ancestors (Kivisild, 2015; Amorim *et al.* 2019). The dissection of ancient samples from Hungary shows a higher frequency of East Asian mitochondrial haplotypes, highlighting contributions to the Eastern Europe genetic makeup, through the maternal lineage, of Siberian populations around 13-6 Kya. Finally, an African input may be dated around 11 Kya, even though it is more probable that the Latinization process (27 BC-1,453 AD) and early Muslim expansion between 1.4-0.7 Kya (Underhill *et al*. 2007), which heavily involved the Mediterranean coast of Africa and the Levant, have been responsible for the introduction of these haplotypes in Southern Europe.

Recently, numerous Eurasian aDNA samples dating 45-5 Kya have been analyzed and, quite notably, it seems that from around 37 Kya, all Europeans, modern or ancient, share the same ancestry (Fu *et al.* 2016, Lazaridis *et al.* 2016). The two studies recognized in these data five main Ice Age populations, each correlated to different geographical areas and material cultures: the Vestonice cluster in Czech Republic, dated around 30 Kya; the Mal'ta cluster from Siberia, which dates to 22 Kya; the El Miron cluster of Spain (dated to 16 Kya); the Villabruna cluster in Northern Italy and the Satsurblia cluster from Caucasus, both dated to 12 Kya (Fu *et al.* 2016, Lazaridis *et al.* 2016).

It is hypothesized, given this information, that the extension of the ice sheets in Northern and Central Europe forced a depopulation of the Northern part of the continent during the LGM (about 27-19 Kya), while the Mediterranean areas (the Iberian Peninsula, Italy and the Balkans) and the coast of the Black Sea retained a suitable climate, becoming refuges for the European populations while remaining largely isolated for thousands of years (**Figure 1.2.3.1**). These groups with a low number of individuals did not interbreed significantly and the bottleneck effect severely reduced the genetic diversity of the isolated groups, while some haplogroups could have had time to diversify due to the strong genetic drift usually associated with small populations. At the end of the LGM, the European re-expansion of the isolated groups from the Southern part of the continent was concomitant with novel migratory waves from Anatolia and the Caucasus, of Mesolithic hunter-gatherers and Neolithic farmers (8-10 Kya). Finally, low proportions of a Yamnaya-related ancestral component, present in all modern European populations and dated 5 Kya, suggest that small groups from the Pontic steppe also provided a genetic component to contemporary Europeans (Sazzini *et al.* 2014; Sazzini *et al.* 2016).

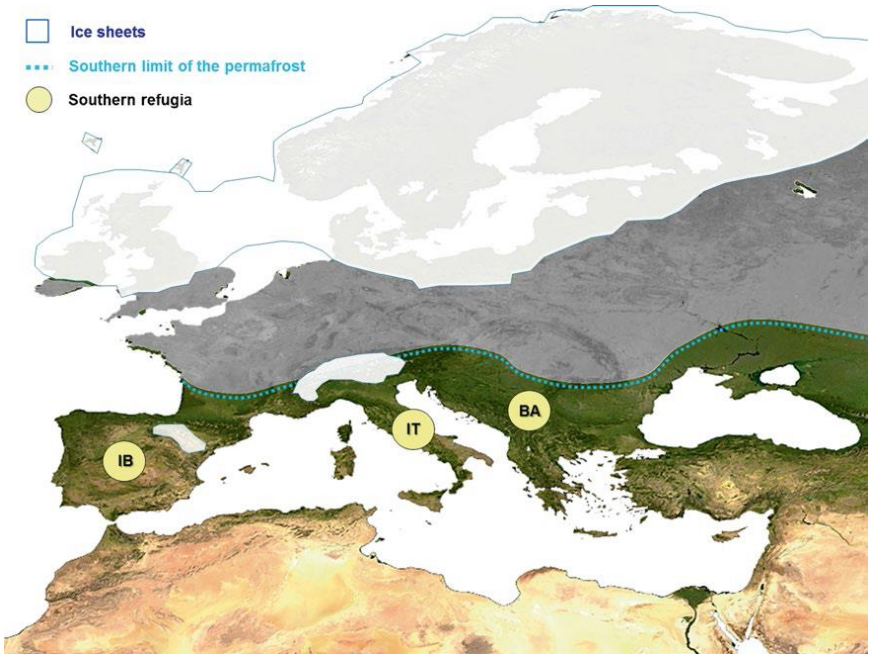


Figure 1.2.3.1 Last Glacial Maximum and European refugia. The Figure highlights the extrapolated extent of the ice sheets and permafrost during the LGM peak, and the three Southern European peninsulas that acted as refugia for the European meta-population (IB, Iberian; IT, Italian; BA, Balkan). From Sazzini *et al.* (2014).

1.2.4. Ancient peopling and recent migrations along the Italian Peninsula

Morphometric analysis of a molar found in Southern Italy, together with chronometric data associated to the layer in which it was found, suggested that the sample, previously attributed to *H. neanderthalensis*, actually belongs to an anatomically modern human dating 45-43 Kya (Benazzi *et al.* 2011). This suggests that *H. sapiens* possibly reached the Italian Peninsula along a coastal path from the Levant at least 45 Kya.

As introduced in the previous paragraph, during the LGM several European populations were forced to move to Southern Europe by the expansion of the glaciers over the Northern half of the continent and, while the importance of the Iberian and Balkan territories has been comprehensively described in literature, the role of the Italian Peninsula as a refuge and reservoir of European genomic diversity has been highlighted only recently (Sazzini *et al.* 2014; Sazzini *et al.* 2016; Raveane et al. 2019). By investigating mtDNA haplogroup patterns in modern Italian samples, the discovery of highly differentiated lineages for haplogroup HV, which can be dated back to the LGM, corroborates this hypothesis (De Fanti *et al*. 2015a).

It also appears that the Italian Peninsula may have been involved in a process of early Neolithization, having been the subject of two expansion routes (both around 8-6 Kya) that introduced agricultural novelties and farming from Southeastern to Western Europe: a coastline path along the Adriatic Sea may have brought the Cardial culture to Southern Italy; later, a migration wave associated to the *Linearbandkeramik* culture of Central Europe extended to the North-West regions of the peninsula (Zeder, 2008; Boattini *et al.* 2013; De Fanti *et al.* 2015a; Natali and Forgia, 2016).

The appearance of proto-historic populations of uncertain origin, together with the settlement of European groups in Northern Italy, such as the Celts, then contributed to a complex socio-cultural stratification that may have had repercussions on the genetic makeup of the inhabitants of the Peninsula during the Late Neolithic and the Metal Age (Boattini *et al.* 2013; Sazzini *et al.* 2014).

Around 2.6 Kya, the historical thalassocracies of Phoenicians and Greeks extended their political power and commercial paths to the coasts of North Africa, Sicily, Southern Italy, France and Spain, strongly impacting the demography and cultures of the Mediterranean populations (Lazaridis, 2018; Mathieson *et al.* 2018). More recently, the growing political and economic influence of the Roman Empire led to the establishment of new colonies in the Mediterranean basin, as it expanded across Europe and North Africa over centuries. Finally, the Muslim conquests and dominion of the Levant, Anatolia, Sicily, Iberia and Northern Africa during the Arab-Byzantine wars (1.3-0.9 Kya) changed again the geo-political and social structure of the Mediterranean region, possibly leaving traces in the genomic background of its inhabitants (Lazaridis, 2018; Mathieson *et al.* 2018; Antonio *et al.* 2019; Arauna et al. 2019). For thousands of years, each commercial route involved not only the

movement of goods, but the opportunity for people to travel through the entire continent, confounding the already complex genetic makeup of the Mediterranean populations, so that the independent influence of each historical event on the genetic background of contemporary Italians is almost impossible to disentangle (Lazaridis, 2018; Mathieson *et al.* 2018; Antonio *et al.* 2019).

1.2.5. Insights into the genetic makeup of the Italian population

Recent studies tried to tackle the complex genomic background of the Italian population by focusing on a description of the geographic distributions and ancestry proportions along the peninsula of both uniparental markers and autosomal variants (Boattini *et al*. 2013; Sarno *et al*. 2014; Sazzini *et al*. 2016; Anagnostou *et al*. 2017; Sarno *et al*. 2017). From the analysis of NRY haplogroups, three clusters could be identified, rather than a continuous North-South cline: a Northwestern group (NWI), a Southeastern one (SEI) and an isolated Sardinian cluster, with NWI individuals exhibiting a higher internal differentiation. This observation is corroborated by the notion that the island has been colonized sparingly over the course of tens of thousands of years, until the advent of the seafaring civilizations, with a concomitant delay in the arrival of the European cultures and the Neolithic revolution; instead, more continuous waves of migration blended with the already present genomic background of the peninsula (Sazzini *et al*. 2016).

Genome-wide research into the genetic makeup of modern Italians supposedly revealed inputs from five ancestral populations (**Figure 1.2.5.1**). The first component (purple in **Figure 1.2.5.1**) was widespread in Southern European populations, homogeneous along the Italian Peninsula and showed the highest proportions in Sardinians: taking into account the extensive isolation that characterized the island after the Neolithic, this genetic contribution may potentially derive from early Neolithic European settlers. Two other components (green and blue respectively in **Figure 1.2.5.1**), predominant in Southern Italian people and in Sicily, are respectively prevailing in populations from the Caucasus/Middle East and Northern Africa/Middle East: their distribution in Southern European populations facing the Mediterranean basin suggests, over an extended timeframe, subsequent coastal migratory instances to Southern Italy, overlapped with the geopolitical dynamics of the Byzantine Empire and the more recent Muslim conquests. A fourth component (red in **Figure 1.2.5.1**), prevalent in the North and reducing towards Southern Italy, is uniformly distributed across Central and Eastern European populations and supposedly related to the changing demography of Europe concurrent with migratory waves from the Steppes, 4-5 Kya. It is also paramount to remember that the impact and extent of the commercial routes during the Roman Empire and the systematic colonization of foreign countries from the Middle Ages may have contributed to waves of migration that modified the pre-existing genetic pattern. A last, minor factor showed low frequencies in the Italian population and

was detected in Northern and Eastern European individuals: it may be a relic of the pre-Neolithic component that reached the peninsula, a refuge for the ancient European diversity, during the LGM (Sazzini *et al.* 2016).

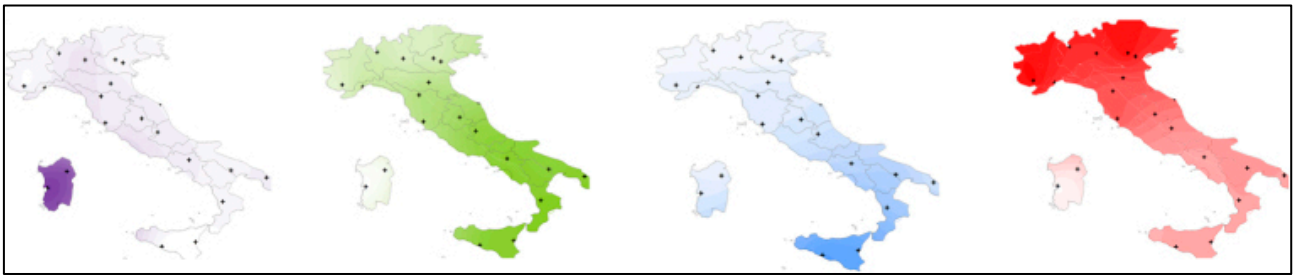


Figure 1.2.5.1. Four ancestral components projected on the profile of the Italian peninsula. The colours and clines correspond to what presented in the main text: a prevalent Neolithic component (purple) in Sardinia; a Caucasus/Levantine contribution (green) on a decreasing South-to-North cline; an African/Near East component (blue) especially in Sicily; a Central/Eastern European contribution (red) on a decreasing North-to-South axis. From Sazzini *et al.* (2016).

Chapter 2: Evolutionary dynamics of the Italian population

2.1. Introduction

Numerous recent studies tried to disentangle the genetic background of contemporary European groups, showing that populations of distinct ancestries contributed to its complexity both historically and prehistorically (Lazaridis *et al.* 2014; Haak *et al.* 2015; Mathieson *et al.* 2015; Fu *et al.* 2016; Hofmanova *et al.* 2016; Lazaridis *et al.* 2017; Mathieson *et al.* 2018; Olalde *et al.* 2018). Some of these lines of enquiry highlighted a subtle change in genetic variation across the continent, with divergence especially between Central and Southern Europe (Lao *et al.*2008; Novembre *et al.* 2008). The same pattern of variation is uniquely mirrored in the Italian Peninsula, indicating that the examination of demo-evolutionary occurrences in which it has been involved may facilitate the description of the dynamic processes and interactions having produced the genomic background observable in Europe today (Capelli *et al.* 2007; Brisighelli *et al.* 2012; Boattini *et al.* 2013; Sarno *et al.* 2014; Fiorito *et al.* 2016; Sazzini *et al.* 2016; Sarno *et al.* 2017).

Both Y-chromosome and mtDNA markers (Capelli *et al.* 2007; Brisighelli *et al.* 2012; Boattini *et al.* 2013; Sarno *et al.* 2014), as well as autosomal variants from genome-wide sequencing data (Fiorito *et al.* 2016; Sazzini *et al.* 2016; Sarno *et al.* 2017), provided a detailed description of how the Italian population is structured. For the same reason, several linguistically and geographically isolated groups inhabiting the peninsula have also been studied, with very intriguing theories being introduced about the demography and ancestry of present-day Italians (Destro Bisol *et al.* 2008; Capocasa *et al.* 2014; Sarno *et al.* 2016; Anagnostou *et al.* 2017). For instance, both Y-chromosome and autosomes seem to have conserved traces of pre-Neolithic contributors (Boattini *et al.* 2013; Sazzini *et al.* 2016; Sarno *et al.* 2017), while maternal lineages in Southern Italy implied an association, predating the Neolithic, with groups from the Caucasus/Near East, which may corroborate the key function of the Peninsula as a refuge for Central European populations during the Last Glacial Maximum (LGM) (Sarno *et al.* 2014; De Fanti *et al.* 2015a). The genetic uniqueness of the Sardinian population was inferred as an Early Neolithic genomic remnant that escaped post-LGM population movements due to the geographic isolation of the island (Piras *et al.* 2012; Francalacci *et al.* 2013; Sikora *et al.* 2014; Sazzini *et al.* 2016; Olivieri *et al.* 2017; Chiang *et al.* 2018); on the contrary, intense migratory waves strongly impacted the standing genomic background along the Italian peninsula. The most striking trend in the cline of Y-chromosome variation is from north-west to south-east, and possibly ascribable to two migration waves, one bringing Neolithic innovations and ancestries to Southern Italy along the Adriatic coast, and a later one involving only the northern territories (Boattini *et al.* 2013; Sazzini *et al.* 2014; De Fanti *et al.* 2015b). Late Neolithic/Bronze Age

demographic changes left strong signals on the paternal lineage and autosomal genome (Boattini *et al.* 2013; Sarno *et al.* 2017), while recent influences could be linked to colonial conquests and movements of people along commercial routes across Europe, throughout the Roman Empire and the Middle Ages (Fiorito *et al.* 2016; Sazzini *et al.* 2016). A stronger Eastern Mediterranean gene contribution has been found in Southern Italy, as the population presented genetic affinity with people from the Eastern Mediterranean (Crete, Cyprus and Anatolia), while an ancestral contribution ascribable to the Arab conquest of Sicily could be identified for the population of the island (Sarno *et al.* 2014; Sazzini *et al.* 2016; Sarno *et al.* 2017). These findings highlight and corroborate an extremely complex history of population movements and gene flow that involved the predecessors of contemporary Italians; they also justify the high degree of genetic and cultural heterogeneity that distinguishes the Italian population from any other European group (Capocasa *et al.* 2014).

The geographical span of the Peninsula, its physical connection to Central Europe and the notable ecological diversity that characterizes a relatively small territory, as well as its position in the middle of the Mediterranean basin, likely forced the first settlers to deal with an extremely diverse and variable environment that could exert diverse selective pressures at a local level. A mixed genomic context, such as the one previously described for the Italian population, may have been a promising background for natural selection to act locally, triggering the evolution of novel adaptations (Sazzini *et al.* 2016). Accordingly, despite being largely understudied, investigation of the adaptive history of the Italian people promises to pinpoint a valuable compendium of gene-environment interactions having played a relevant role in the evolution of European populations.

Nevertheless, studies based on single genetic systems (i.e., mitochondrial DNA and Y chromosome) or on autosomal variants at high frequency prevented a thorough analysis of the variation range in the Italian population. In particular, low frequency variants and small-size effect single nucleotide polymorphisms (SNPs) are inadequately scrutinized by SNP arrays, although they are highly effective in revealing population structure and are crucial factors in complex and adaptive traits, especially considering a model of polygenic adaptation, that is more realistic than a simple adaptive sweep (Pritchard *et al.* 2010a; Pritchard *et al.* 2010b; Gravel *et al.* 2011; Hernandez *et al.* 2011; Keinan *et al.* 2012; Mathieson *et al.* 2012; Schrider *et al.* 2017; Gnecchi-Ruscone *et al.* 2018). In particular, SNP arrays make it difficult to disentangle the appearance of rare variants due to a recent population expansion from their contributing effect due to polygenic adaptation, so that high-quality and high-coverage genomes need to be retrieved for analysis and putative signals of adaptation have to be contextualized in the demographic history of the populations under scrutiny (Schrider *et al.* 2017; Gnecchi-Ruscone *et al.* 2018).

High-coverage (90x) whole-genome sequences have been produced from 38 Italian individuals distributed along the peninsula to adequately scrutinize the demography and adaptive events that led to the modern Italian genomic background. By selecting individuals belonging to two uniform sub-populations (indicated as N_ITA and S_ITA) at the extremities of the variation pattern observable in the peninsula (Sazzini *et al.* 2016), the relationship of these two groups with modern and ancient Euro-Mediterranean clusters could be properly assessed in terms of population dynamics and demographic structure; furthermore, novel instances of adaptive evolution could be identified that, having acted differentially on the two groups under scrutiny, favored their adaptation to local environmental and cultural selective pressures.

2.2. Materials and methods

2.2.1. Sample sequencing

38 unrelated and healthy individuals were selected from the Italian regions of Piedmont, Lombardy, Veneto, Emilia-Romagna, Apulia, Calabria and Sicily; in order to be proper representatives of the populations inhabiting the geographical areas from where they were collected (Sazzini *et al.* 2016), the grandparents' criterion was applied, according to which all parents and grandparents for each individual must come from the same place. To obtain high-coverage WGS data (90x), sequencing libraries were produced with the TruSeq DNA PCR-Free Library Preparation Kit (Illumina San Diego, CA, USA) by applying a 350-bp insert size and following the producer's protocol. Sequencing was performed with a HiSeq X Ten Reagent Kit v2.5 for 2 × 150 cycles and a HiSeq X Ten sequencing platform (Illumina San Diego, CA, USA). The Isaac aligner (version 01.14.02.18) was used to align the attained reads against reference sequence hg19 (GRCh37) of the human genome, by selecting a minimum PHRED quality score threshold of 20 from the 3′-end; variant calling was performed with the default setting of the Isaac Variant Caller (version 1.0.7). This tool automatically executes noise filtration using sequencing and alignment metrics, realigns the reads, filters base calls through pattern recognition mismatch, corrects same-strand base call quality to reveal error dependencies between calls, and computes Bayesian genotype probabilities (Raczy *et al.* 2013). The developers proved that this analytical approach is as accurate and sensitive as the mainstream methodology based on the GATK but is up to five time faster in terms of computational time (Raczy *et al.* 2013). Polymorphisms positioned in tandem repeats and homopolymer regions were removed, together with those having a call rate lower than 98% (Anderson *et al.* 2010), so that the set of high-quality variants for the dataset included 20,075,710 SNVs. The accuracy of the method

was supported by a computed transition/transversion (Ts/Tv) ratio of 2.071, which sits within the expected 2.0–2.1 range (DePristo *et al.* 2011).

2.2.2. Quality control and dataset production

The genotypes produced for the 38 Italian sequences were stored in PLINK (Purcell *et al.* 2007) format files (*.bim, .fam, .bed, .ped, .map*) and specific quality control procedures were applied using functions implemented in the PLINK package (Purcell *et al.* 2007). The above-mentioned PLINK format files supply a detailed description of the variant dataset: the *.bim* file generally contains genetic information organized in six columns, which correspondingly describe, for each SNV, the chromosome it belongs to, a potential SNP identifier (i.e. the rs code), the mapping genetic distance in centiMorgans (cM), the nucleotide position on the chromosome and the alleles that characterize the variant, either as letters or as numbers; the *.fam* file is a pedigree/phenotype file that again contains six columns, defining the family ID to which an individual belongs, an individual identifier, paternal and maternal identifiers, the sex of the individual and a code that describes a potential individual phenotype; the *.bed* file is a non- human readable binary file that contains the same genotype information as its non-binary counterpart (i.e. the *.ped* file); finally, the *.map* file is a shortened version of the .bim file that contains only information belonging to its first four columns.

In particular, SNVs were filtered out if they exhibited more than 5% of missing data and/or significant departures from the Hardy-Weinberg equilibrium, after Bonferroni correction for multiple testing ($p < 5.2 \times 10^{-10}$). Only autosomal variants were considered, by removing SNVs located on sex chromosomes and mitochondrial DNA, and possible ambiguous SNVs (i.e., characterized by A/T or C/G substitutions that would lend themselves to random chromosome copy allocation during haplotype reconstruction). After quality control, a "high-quality Italian dataset" was obtained comprising genotypes for 17,495,290 SNVs.

The WGS dataset was merged with a previously studied Italian dataset (Sazzini *et al.* 2016) of genome-wide genotypes for 737 samples, uniformly distributed along the peninsula and with known origin at the level of local constituencies. This union, performed by only retaining variants that were present in both datasets, produced a "low-density Italian dataset" including 251,648 SNVs that allowed for the allocation of the WGS population clusters along the cline of variation defined by the previously analyzed dataset (Sazzini *et al.* 2016). A Procrustes analysis (Wang *et al.* 2010) was implemented for this purpose: first, a Principal Component Analysis (PCA) was performed using the *smartpca* tool in the EIGENSOFT package v6.0.1 (Patterson *et al.* 2006); then, the coordinates of each individual for the most informative PCs (the ones that overall account for more than 90% of the dataset's variance) were averaged within provinces and projected onto their geographic coordinates

by using the R *vegan* package. For the PCA to be accurately performed, from the merged "low-density Italian dataset" was paramount to remove variants in high linkage disequilibrium (LD), by applying a threshold of $r^2 > 0.2$ within windows of 50 SNVs and advancing by five steps. Identity by descent (IBD) was also computed for each pair of subjects as the proportion of shared alleles, to detect unexpected genetic relatedness among subjects: a kinship coefficient of 0.125 (equal to the expected genome-wide identity of third-degree relatives) or lower was deemed acceptable.

The "high-quality Italian dataset" was also separately merged with publicly available data, produced with the same sequencing technology, for 69 individuals belonging to 35 Euro-Mediterranean populations (Mallick *et al.* 2016). This "high-density Euro-Mediterranean dataset" comprised 6,993,871 SNVs and was used for a dual purpose: first, to perform haplotype-based population structure analysis, for which the dataset was phased (i.e. haplotypes were estimated from genotypic data) with SHAPEIT2 v2.r790 (Delaneau *et al.* 2013) by using standard parameters, a reference panel from the 1000 Genomes Project (1000 Genomes Project Consortium, 2015) and HapMap phase 3 recombination maps; second, to carry out aDNA-informed population dynamics inferences, for which the dataset was further merged with a genome-wide panel of 559 ancient samples assembled from literature (Lazaridis *et al.* 2014; Mathieson *et al.* 2015; Fu *et al.* 2016; Lazaridis *et al.* 2017; Lipson *et al.* 2017; Mathieson *et al.* 2018; Olalde *et al.* 2018), to create a "modern + aDNA dataset" including 47,806 SNVs.

The original "high-quality Italian dataset" was also phased as described for the "high-density Euro-Mediterranean dataset"; in this case, though, a reconstructed human genome was used to discern between ancestral and derived alleles. The state of each allele the reference human sequence was determined by aligning it with the Ensembl Compara 6 primates EPO genome sequences (link accessed 9 May 2019; Herrero *et al.* 2016), with the criterion that only alleles present in all the compared genomes were conservatively deemed ancestral. Selection scans were finally performed on the resulting "phased high-quality Italian dataset" with known ancestral/derived states and including 13,381,038 SNVs.

2.2.3. Haplotype sharing clustering analyses

The phased "high-density Euro-Mediterranean dataset" was submitted to haplotype-based analysis using the CHROMOPAINTER/fineSTRUCTURE pipeline (Lawson *et al.* 2012), with the aim of testing whether the Italian high-quality sequences were part of genetically uniform and distinct Italian clusters. CHROMOPAINTERv2 reconstructs haplotype sharing for each individual, by using the other samples in the dataset as "donors" but preventing self-copy. Mutation and recombination rates for the method were first estimated using 10 steps of the expectation-

maximization (E-M) algorithm on a representative subset of chromosomes (4,10,15,22); mean values were then computed throughout all autosomes for all individuals and weighted by the number of variants; finally, mutation and recombination rates were plugged in to perform the complete CHROMOPAINTER analysis, using a number of expected haplotype chunks per genomic region k=100. CHROMOPAINTER outputs a matrix of counts of shared haplotype chunks that can be given as input to fineSTRUCTURE version fs2.1 (Lawson *et al.* 2012). The parameters for the algorithm were as follows: 1,000,000 "burn-in" iterations of MCMC; 1,000,000 iterations with sampling of the inferred clustering patterns every 10,000; 100,000 hill-climbing steps to improve posterior probability. The clusters were defined by collapsing branches of the obtained dendrogram up to the second-to-last splitting point, to reduce the number closely related and unsupported clusters; clades splitting with a posterior probability threshold lower than 80% were also collapsed until a branching point with probability higher than the threshold was reached.

2.2.4. Inferring and dating recent admixture events

The phased "high-density Euro-Mediterranean dataset" was also parsed with GLOBETROTTER (Hellenthal *et al.* 2014) to fine map recent events of admixture between the Italian and Euro-Mediterranean clusters detected by fineSTRUCTURE.

In this case, CHROMOPAINTER was run so that the length of haplotype chunks copied from every donor by each individual was averaged across samples for each given cluster and the two Italian groups, N_ITA and S_ITA, were excluded from the donors but kept as recipients. The GLOBETROTTER pipeline was applied on this matrix to first deduce admixture proportions in the Italian clusters with the *nnls* function (Hellenthal et al. 2014). The result was used to provide a date for each admixture event. In particular, the dating procedure was first run applying the "null individual" option, which produces a collage of random painted chunks from different individuals in the dataset, to which each cluster is first compared: this is supposed to provide a standardization for the coancestry curves resulting from the subsequent direct comparison of the clusters. Then, each pair of parental groups, excluding N_ITA and S_ITA, was tested twice: once performing a first run to infer admixture sources, proportions and dates; the second time, the algorithm was run by inputting the previously obtained results and applying bootstrap resampling (100 times) to provide confidence intervals for the dates of admixture.

2.2.5. Exploring relationships between modern and ancient populations

To formally contextualize the genetic relationships of the N_ITA and S_ITA clusters inferred by fineSTRUCTURE with the ancient Eurasian genomic background, the "modern + aDNA dataset"

was used to perform a PCA. The EIGENSOFT package v6.0.1 (Patterson *et al.* 2006) provides a *smartpca* method with the option *lsqproject* to allow the inclusion of data with a high rate of missing genotypes, such as those coming from ancient DNA samples.

Ancient samples were grouped by their archeological/cultural background and according to the results of the previously performed PCA. Then, an outgroup *f3* statistics to evaluate the degree of genetic drift was implemented in the form of *f3* (CHB; X Italian population cluster, Y ancient population cluster) as made possible by the *qp3pop* function present in ADMIXTOOLS (Patterson *et al.* 2012). N_ITA and S_ITA clusters were finally compared by computing the residuals in their outgroup f3 scores for each ancient population group, and differences exceeding ± 2 SDs from the mean of the distribution of residuals were deemed significant. The Chinese population (CHB) from the 1000 Genome Project (1000 Genome Project Consortium, 2015) was used as a proxy for an unadmixed group with no shared ancestry to both the Italian clusters and the modern or ancient Euro-Mediterranean groups.

2.2.6. Estimates of effective population sizes and split times

The demographic history of the two Italian clusters was modeled by comparing it with the CEU population from the 1000 Genomes Project through the sequential Markovian coalescent (SMC++) method (Terhorst *et al.* 2017). This tool requires LD data emerging from a coalescent hidden Markov model and a population frequency spectrum to estimate changes in effective population size and genetic split times. According to simulations reported in Chiang *et al.* (2018), which presented models of growth for populations of European ancestry, 150 generations were set as the earliest time point for population size inference (T1) and 10 interpolation timepoints were provided (Chiang *et al.* 2018). A mutation rate of 1.25×10^{-8} mutations/nucleotide/generation (Chiang *et al.* 2018) and a generation time of 29 years (Langergraber *et al.* 2012) were applied to convert the estimates of N_e and split times provided by SMC++.

2.2.7. Genomic signatures of natural selection

As the fineSTRUCTURE analysis confirmed two distinct Italian population clusters, the "phased high-quality Italian dataset" could be split in two groups, corresponding to the individuals associated with either N_ITA or S_ITA, and used to detect signatures of selection for the separate groups. The complementary statistics nSL and DIND were applied to infer different instances of positive selection.

The first applied test was the nSL statistics (i.e. the number of segregating sites by length) that was designed in order to be able to detect both hard and soft sweeps having affected patterns of

haplotype homozigosity as a consequence of the action of relatively recent positive selection, respectively on newly arisen or already existent genetic variants (Ferrer-Admetlla *et al.* 2014). In fact, a selective sweep represents an unexpected reduction of nucleotide variation surrounding an inheritable nucleotide substitution. In details, a hard sweep (or classic selective sweep) is expected to happen when a rare beneficial de novo mutation appears in the gene pool of a given population and being advantageous, rapidly increases in frequency thus reducing population diversity. A soft sweep instead happens when a pre-existing neutral variant (i.e. standing variation) strives in a new environment, becoming beneficial and thus spreading rapidly in the population, but insisting on a previously existing genomic background.

In particular, the nSL test was proved to be more robust to variation in recombination rates and to the confounding effects due to demography with respect to other haplotype-based tests and was thus highly suitable to be applied to WGS data (Ferrer-Admetlla *et al.* 2014).

Computation of nSL scores for each SNV require a matrix H, with rows denoting the haplotypes and columns containing the segregating sites. Each value in the matrix could either be 0, if the haplotype had an ancestral allele in that position, or 1 there was a derived allele. For each SNV, the haplotypes were divided in two groups: one (D(k)) having the derived allele for the core variant (pk) and the other (A(k)) with the ancestral allele in the same position. Any two haplotypes i and j in each group were picked and the length of the longest continuous section containing identical alleles in both haplotypes was computed as hap_{max}(pk). For each variant, then, all these lengths in each were put together and divided by the squared dimension of the group; the value obtained for the ancestral allele at each SNV is divided by the value obtained for the derived allele and, finally, the natural logarithm of this ratio gives the value of the nSL statistic at each site:

$$nSL\ (k) = ln \frac{\frac{2 \sum_{i<j \in A(k)} hap_{max}(p_k)}{n_A(k)\ (n_A(k) - 1)}}{\frac{2 \sum_{i<j \in D(k)} hap_{max}(p_k)}{n_D(k)\ (n_D(k) - 1)}} \quad (1)$$

While the nSL infer extensive genomic regions that plausibly underwent positive selection, the DIND (derived intra-allelic nucleotide diversity) statistic has the potential to fine map adaptive variants. The same matrix H and a similar grouping concept is applied for DIND computation, but given a window of fixed size, the direct number of differences (hap_{diff}) among haplotypes in the same group is assessed (Fagny *et al.* 2014) and the score not logarithmically scaled:

$$DIND\ (k) = \frac{\frac{\sum_{i<j \in A(k)} hap_{diff}(p_k)}{n_A(k)^2}}{\frac{\sum_{i<j \in D(k)} hap_{diff}(p_k)}{n_D(k)^2}} \quad (2)$$

The combination of these two tests seems advantageous, since nSL focuses on the examination for the longest common fragments among haplotypes, can highlight instances of positive selection represented by both hard and soft sweeps, the DIND test accounts for the overall differences between haplotypes around the derived allele, and was specifically developed for the identification of hard sweeps. In particular, when compared to other haplotype-based tests, the DIND statistics is robust to uneven sequencing coverage and small sample size (Barreiro *et al.* 2009), while the nSL is less influenced by disparity in recombination rates and demographic effects (Ferrer-Admettla *et al.* 2014). For the purpose of this study, variants showing DAF lower than 0.2 were filtered out, as DIND results were demonstrably biased around these frequencies (Fagny *et al.* 2014), and the statistical score was computed for each variant using self-customized Python scripts. nSL comptation was instead carried out through the *selscan* v1.1.0b package (Szpiech *et al.* 2014), by considering a maximal window size of 4500 consecutive loci and an extension of haplotype homozygosity of 200kb.

The BALLET pipeline (DeGiorgio *et al.* 2014) was further applied to test for the occurrence of ancient events of balancing selection in the two Italian clusters by accounting for the spatial distribution of polymorphisms and substitutions along the genome. The genome of an outgroup species (i.e., *P. troglodytes*) is compared to each individual in the tested group: all SNVs that did not show any differences between the test and the outgroup were removed, while the number of polymorphisms (inter-individual changes in the test group) and substitutions (inter-specific changes) for each site was computed. Each accounted variant was then associated to its recombination value, and a date of six million years ago for a coalescent event between *H. sapiens* and *P. troglodytes* was provided as the final input for the BALLET method.

Finally, in order to focus on signatures of selection that were peculiar of the selective pressures endured by the Italian groups, all of the scans for selection were replicated on CEU and IBS sequences provided publicly by the 1000 Genomes Project (1000 Genomes Project Consortium, 2015): shared signals between Italians and other populations of Western European ancestry were removed, as they may have arisen in the common ancestor of these populations and are not specific adaptations to the environment of the Italian Peninsula. The lack of a congruent number of high-coverage genomes from other populations of Southern European origin prevented further filtering.

2.2.8. Gene network analysis

The paramount assumption underlying the polygenic model is that several variants conferring an adaptive advantage through haplotypes at low-to-moderate frequency may be much more indicative of a realistic model of adaptation, than a significant spike in frequency of a single favorable haplotype (Pritchard *et al.* 2010b, Gnecchi-Ruscone *et al.* 2018). In particular, single-gene selection scans have low power in detecting relatively weak selective events, which appear to insist on alleles that are relatively sparse in a population. Instead of studying the results obtained at the single variant or gene level for nSL, DIND and BALLET, the genome-wide distribution of the above-mentioned selection scores was tested in a model closely resembling polygenic adaptation, so that gene networks enriched for weaker signals of selection could be inferred. The approach suggested by the *signet* algorithm (Guoy *et al.* 2017) tests whether natural selection has acted at the level of gene networks in order to drive the adaptation of overall functional pathways (Gnecchi-Ruscone *et al.* 2018). The input to the *signet* algorithm was provided by locating the genes in which the SNVs fall and considering, for each gene, the highest DIND, nSL, and BALLET score among those associated to the variants that it contains. Functional pathways can be provided by referring to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, on which the algorithm proceeds with two computations: first, for each provided network, the highest scoring subnetwork (HSS) is estimated through a simulated annealing approach, by starting with a random subnetwork, computing the normalized mean score of its nodes (or genes), and then iteratively adding or subtracting a node and computing the normalized mean score of the new subnetwork, in search of the global maximum score; then, a "null distribution" of scores is produced by shuffling gene values across networks, then a network is randomly picked with a probability proportional to its size, its HSS is computed as described previously and this procedure is iterated 10.000 times to obtain a distribution of values. Finally, a p-value is computed by comparing the HSS obtained in the first step with the null distribution in the second step and assessing the proportion of random subnetworks of comparable size that have a normalized mean score equal to or larger than the HSS score. Gene subnetworks with significant p-values ($p < 0.05$) were identified with this procedure for each population cluster (N_ITA, S_ITA, CEU, IBS) using as input each of the computed selection statistics, and graphically mapped using Cytoscape v3.6.0 (Shannon *et al.* 2003).

2.3. Results and discussion

A high-quality dataset including 38 Italian samples and 17,495,290 SNVs was assembled and self-reported ancestry for the donor individuals were verified with a Procrustes analysis including genotype data from 737 Italians of certain geographical origin (Sazzini *et al.* 2016). The 38 samples accordingly distributed among the clusters identified as N_ITA and S_ITA (**Figure 2.3.1**), at opposite

extremes across the latitudinal cline of Italian variation (Ogata *et al.* 1999; Franceschini *et al.* 2013; Sazzini *et al.* 2016).

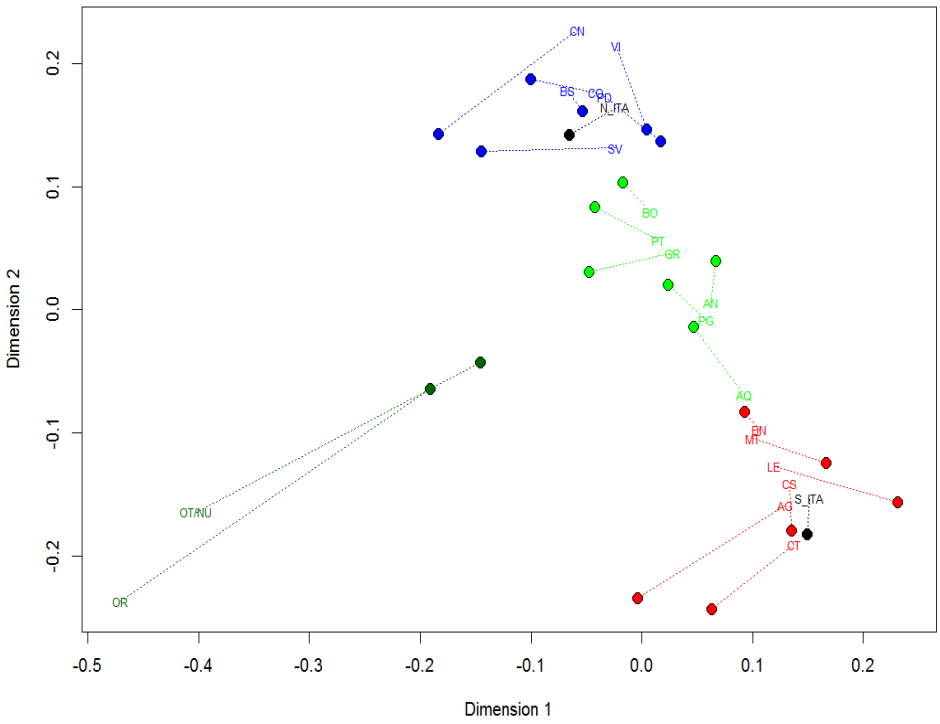


Figure 2.3.1: Procrustes analysis of the Italian data. The genomic coordinates (group labels) of the first and second principal components are averaged across individuals for the same location and projected onto the geographical coordinates (colored dots) of the population samples. Dotted lines highlight the differences in genetics with respect to geography. According to Sazzini *et al.* (2016), Northern Italian provinces are colored in blue; Central Italian provinces in bright green; Southern Italian provinces in red; Sardinian provinces in dark green. The mean of genomic (N_ITA and S_ITA) and geographic coordinates for the high-quality samples sequenced in the present study are displayed in black. From Sazzini *et al.* (2020).

2.3.1. The Euro-Mediterranean genomic landscape

WGS data (Mallick *et al.* 2016) were retrieved from literature and assembled with the 38 high-quality Italian genomes to create a "high-density Euro-Mediterranean dataset" on which to run fineSTRUCTURE analysis, retrieve a co-ancestry matrix and build a dendrogram based on its results (Fig. 1a). Italian samples located on two opposite branches of the dendrogram. In detail, S_ITA subjects clustered apart from N_ITA ones and in close proximity the Cretan cluster, branching out from the node originating also the bulk of individuals from Caucasus, which can be separated in Northern Caucasian (Adygei, Chechens, Lezgins and North Ossetians) and Southern Caucasian (Armenians, Georgians, Turks, and Abkhasians) clusters. The S_ITA/Cretan/Caucasian macro clade further diverged from a node basal to both Near Eastern (i.e., Bedouins, Palestinians, and Jordanians) and North African (i.e., Mozabites and Saharawi) populations. In the other branch of the dendrogram, N_ITA individuals formed a cluster that included with Bulgarian, Albanian and Iberian samples; this clade resulted appreciably different from Sardinians and both diverged from a node that takes to the Basques and the rest of the European populations in the dataset, which can be separated in two main

clusters: a Central-Western European one (i.e., Hungarians, French, British, Czechs, Polish, and Orcadians people) and a North-Eastern European (i.e., Estonians, Russians, Norwegians, Finnish, and Icelandic people) (**Figure 2.3.1.1a**).

fineSTRUCTURE clustering, then, highlights a substantial separation of N_ITA and S_ITA clusters, and is supported by genome-wide estimates of differentiation ($F_{st} = 0.0021$; p value $< 10^{-6}$). Along the Italian Peninsula, these two groups have been shown to include considerable frequencies of the most divergent ancestries in the Italian population, while being genetically connected by a Central Italian cluster deprived of other ancestry proportions (Boattini *et al.* 2013; Fiorito *et al.* 2016; Sazzini *et al.* 2016; Sarno *et al.* 2017). fineSTRUCTURE also hints at a more general trend of divergence at the Euro-Mediterranean level, with S_ITA being genetically closer to groups from Crete and the Caucasus, while N_ITA shows higher affinity with populations from the Balkans (**Figure 2.3.1.1a**). The pattern described for S_ITA may be linked to genetic contributions of people from the Near East and Caucasus during the Neolithic and Bronze Age, as the Caucasus was the likely source of a Bronze Age migration that involved Southern Italy around the same time as the steppe-related migrations that encompassed Central Europe, Northern Italy and the Balkans (Sarno *et al.* 2017).

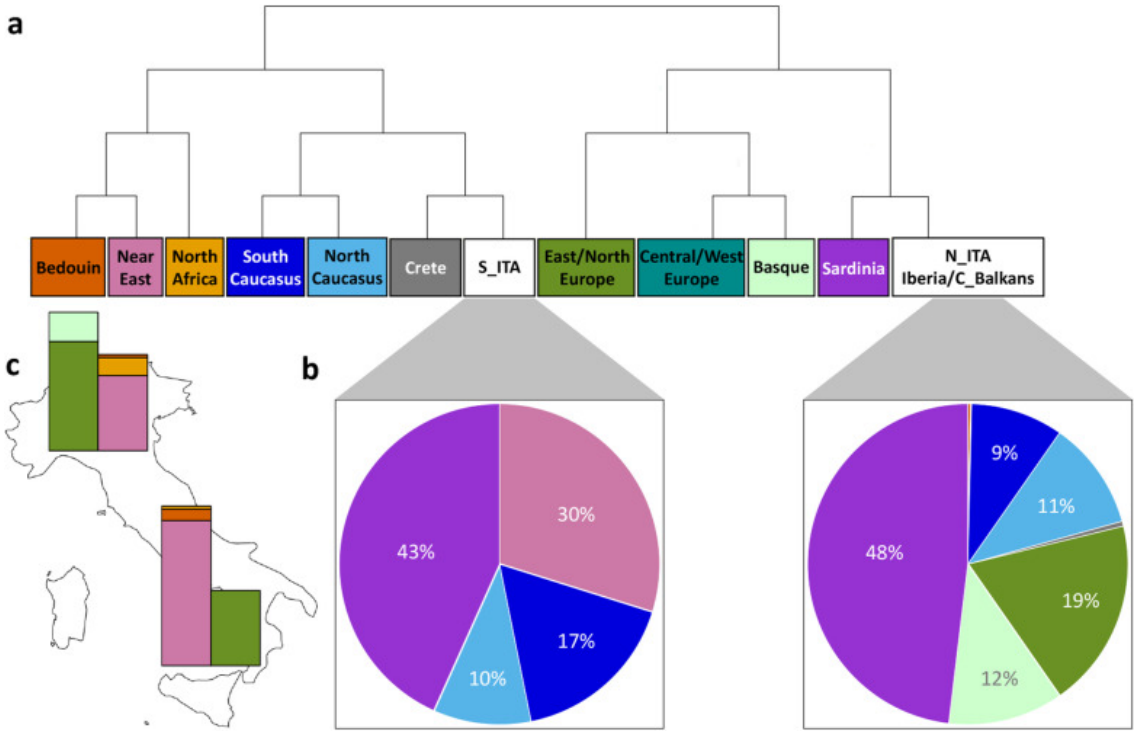


Fig. 2.3.1.1 Clustering analysis, admixture and ancestry proportions. a) Population clusters as defined by fineSTRUCTURE. N_ITA clusters with Iberian and Balkan individuals with a posterior probability higher than the threshold of 80%. b) Proportion of chunks shared between Italian clusters (recipients) and the other populations in the dendrogram (donors), according to CHROMOPAINTER. c) GLOBETROTTER inference of ancestry proportions for the

Italian clusters. The bar on the left represents the major source of admixture. The Figure is color-coded so that each color corresponds to a cluster in the fineSTRUCTURE dendrogram. From Sazzini *et al.* (2020).

2.3.2. Recent admixture events shaped Italian clusters diversity

CHROMOPAINTER was used to investigate proportions of chromosome sharing among individuals for the different identified clusters. Italian groups both shared similar percentages of DNA chunks with Sardinians (N_ITA, 48%; S_ITA, 43%) and Northern Caucasians (~ 10%), which is in line with the theory that Early Neolithic and Bronze Age populations contributed essentially to the European genetic landscape. S_ITA was characterized by a 30% genome sharing with Near Eastern groups, a contribution that is absent in N_ITA, as well as 17% of chunks shared with Southern Caucasian groups, although it has been noted that Southern Caucasian populations share close genetic characteristics with people from the Near East (Mallick *et al.* 2016) and this similarity may influence the sharing pattern. N_ITA, on the other hand, showed elevated proportions of the genome in common with Eastern and Northern European groups (19%) and the Basques (12%), which are not observed in S_ITA (**Figure 2.3.1.1b**).

CHROMOPAINTER profiles were then used to infer plausible proportions related to a recent admixture event involving the Euro-Mediterranean clusters to produce the genetic background of N_ITA and S_ITA have been estimated using GLOBETROTTER on the profiles previously generated by CHROMOPAINTER. Both Italian clusters were characterized by admixture between a Northern European population (plus a Basque contribution in case of N_ITA) and a Near Eastern/North African source (**Figure 2.3.1.1c** and **Table 2.3.2.1**), although with inverse proportions of admixture: 59% European and 41% Near East for N_ITA; 32% European and 68% Near East for S_ITA. Gene flow dates were estimated with a 95% confidence interval to highlight that the admixture events seemed to overlap in time, in a range that goes from 1.2 to 2 kya and with the S_ITA gene flow event being relatively more recent (**Table 2.3.2.2**). Previous studies (Fiorito *et al.* 2016; Sazzini *et al.* 2016) support the notion that the contributions from Europe may be ascribable to the key commercial routes established during the Roman Empire and the Middle Ages, involving predominantly Northern Italy as a passageway to the continent, while Middle Eastern and North African gene flow appear to be concordant with the historical Muslim conquests in Central and Southern Italy. It has also to be taken into consideration that these admixture events may have not introduced novel genetic components to the genetic background of the Italian population as a whole, but rather shaped the different genomic landscapes of N_ITA and S_ITA by emphasizing a differential allocation of ancient genetic characteristics that were already present beforehand.

Table 2.3.2.1. GLOBETROTTER results for inferred admixture proportions.

Cluster	Admixture source	Admixture proportion	Relative contributions		
N_ITA	Major	59%	N_Europe (79%)	Basque (21%)	
	Minor	41%	Near East (78%)	N_Africa (19%)	Beduin (2.4%)
S_ITA	Major	68%	Near East (91%)	Beduin (7%)	N_Africa (2%)
	Minor	32%	N_Europe (100%)		

Table 2.3.2.2. GLOBETROTTER results for inferred admixture dates.

Cluster	Date (N gen.)	95% CI (N gen.)	Date (ya)	95% CI (ya)
N_ITA	62	51.7 - 71.5	1,798	1,499 - 2,074
S_ITA	50	41.9 - 63.1	1,450	1,215 - 1,829

N gen., number of generations; CI, confidence interval; ya, years ago.

2.3.3. Exploring the ancient genetic legacy of Italian population clusters

Genome-wide data for 559 ancient samples were assembled with the modern Euro-Mediterranean dataset in order to push the inference of genetic ancestry for the Italian population clusters back in time, beyond the timescale of the GLOBETROTTER method.

A two-step "nested" PCA was performed, with the modern patterns of genetic variation used as the space in which the ancient data were then projected, so that a differentiation in clustering with modern and ancient samples became noticeable for the two Italian groups (**Fig. 2.3.3.1**). N_ITA individuals, which associated with samples from the Iberian Peninsula (IBS) in the mass of contemporary Southwestern Europeans, showed a peculiar proximity with Neolithic and Copper Age samples from Central Europe, England and the Balkans, as well as a Czech individual associated with the Corded Ware culture. It is also noteworthy that the plotted centroid of the N_ITA cluster is proximal to the Remedello sample from the Copper Age found in Northern Italy. On the other hand, S_ITA individuals showed relatedness with modern Cretans and Greeks, along with Anatolian and Levantine remains from the Neolithic, Copper Age, and Bronze Age, and a Cretan sample belonging to the Minoan civilization (**Figure 2.3.3.1**).

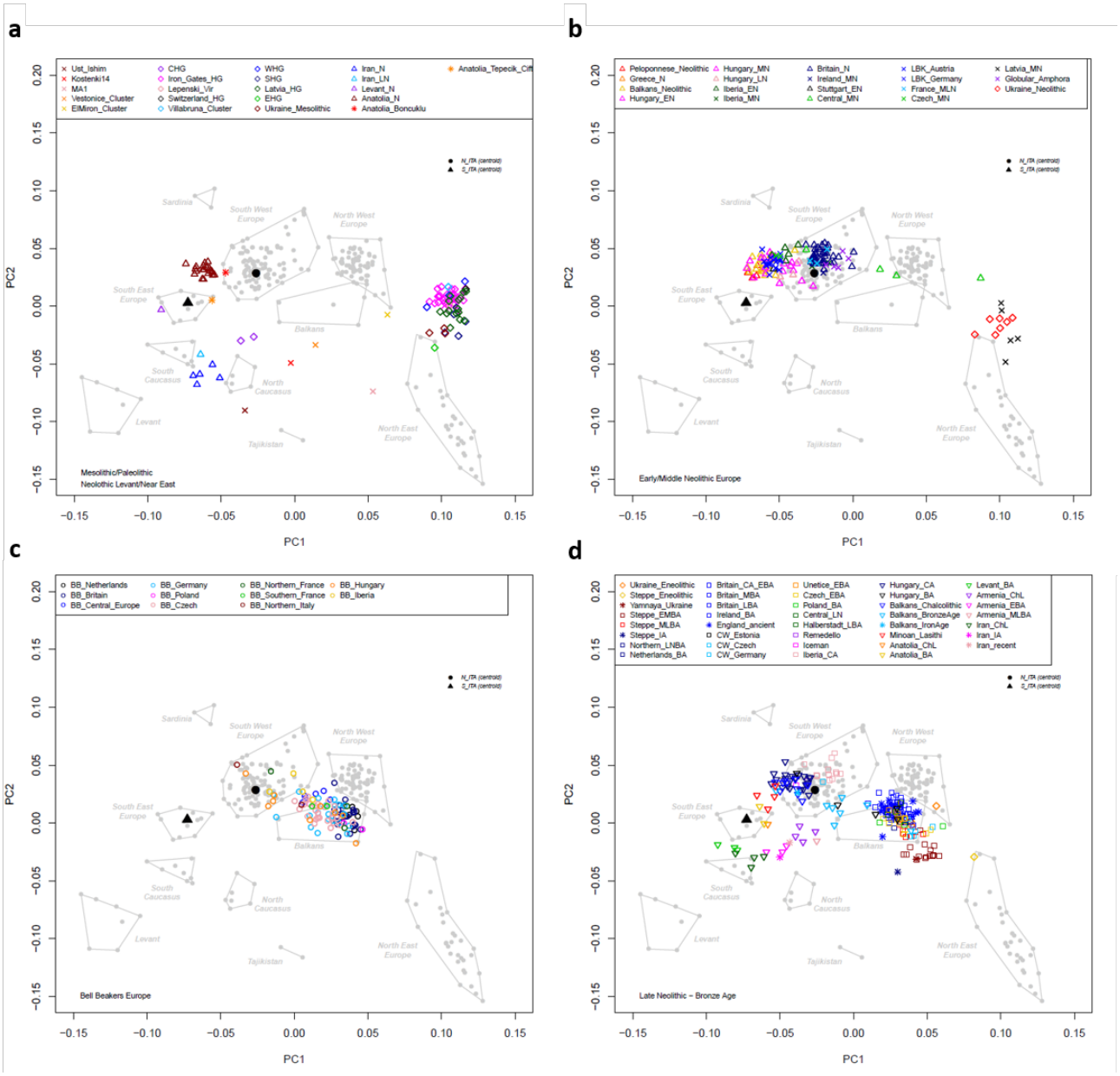


Fig. 2.3.3.1 PCA analysis including aDNA samples. In all panels, centroids for N_ITA and S_ITA clusters are respectively reported in a black dot and black triangle, while modern samples are reported in gray dots. PC1 accounts for 2.24% of the total variance; PC2 accounts for 1.87 of the residual variance. a) Mesolithic/Paleolithic samples, plus Neolithic Levant. b) Early/middle Neolithic European samples. c) European samples from the Bell Beaker culture. d) Late Neolithic/Bronze Age samples. From Sazzini *et al.* (2020).

Ancient samples were then used to quantify shared genetic drift with the present-day Italian groups, by computing f3 statistics. For this purpose, the ancient individuals were clustered in a PCA-informed manner and according to their archeological context. The outgroup *f3* scores for N_ITA and S_ITA were then compared to evaluate differences in ancestral contribution between them. Significant residual of outgroup f3 values were observed for specific ancient clusters (**Figure 2.3.3.2**): negative residuals, indicating closer genetic similarity to the S_ITA group, exceeding one standard deviation (SD) from the mean of the residual distribution highlighted affinity to Caucasus hunter-gatherers as well as Anatolian, Near East, Greek and Balkan Neolithic and Bronze

Age samples, with Levant and Iranian Neolithic remains surpassing two SDs. On the other hand, the N_ITA cluster was characterized by closer affinity (positive residuals over one SD from the mean of the distribution) with the Copper Age Remedello specimen, as well as Bronze Age and hunter-gatherer remains from Central-Eastern Europe; values exceeding two SDs were instead peculiar to the Villabruna cluster and Central European hunter-gatherers (**Figures 2.3.3.1** and **2.3.3.2**).

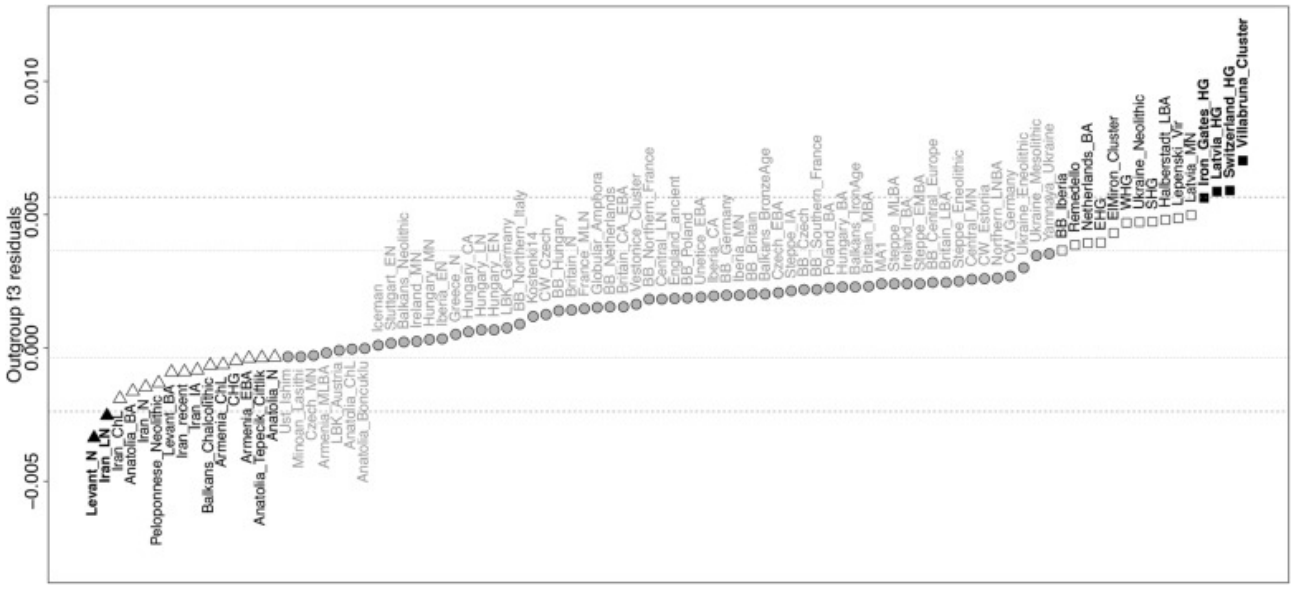


Fig. 2.3.3.2 Distribution of outgroup f3 residuals. The difference in score, computed as f3 (Han Chinese (CHB); X Italian population cluster, X ancient population cluster) to test the N_ITA and S_ITA clusters against each ancient population group, is reported on the x-axis. Negative values suggest closer affinity of each ancient group to S_ITA, while positive values suggest the same for N_ITA. A continuous grey line and a dashed line indicate the limit of one and two standard deviations from the mean of the distribution. Residuals exceeding one SDs are in white; those exceeding two SDs are in black. Squares are used for ancient groups closer to N_ITA; triangle for groups closer to S_ITA. From Sazzini *et al.* (2020).

The results obtained with aDNA samples are consistent with the hypothesis of Southern Italy having been at the limit of a Mediterranean passageway that facilitated the dissemination of farming from Southeastern Europe (Sazzini *et al.* 2014; Hofmanova *et al.* 2016; Omrak *et al.* 2016), with Neolithic migrations from Anatolia and the Caucasus having left stronger traces in the genomic landscape of the S_ITA group. It is also coherent to suppose that a Chalcolithic/Bronze Age migration wave from the same geographic area may have reached the peninsula later, from a Mediterranean route (Mathieson *et al.* 2018). This observation also justifies the patterns of haplotype sharing that are predominant in S_ITA by providing a direct link with present-day populations from the Near East and Southern Caucasus (**Figure 2.3.1.1b**).

When considering the N_ITA cluster, the increased affinity with Copper and Bronze age samples from Central and Eastern Europe justify the higher ancestry proportions shared with modern Northeastern European populations. However, signatures referring to European hunter-gatherers, the Late Glacial "El Miron" cluster and the post-Ice Age "Villabruna cluster" (**Figure 2.3.3.2**) cannot be

simply justified by a direct link between these Upper Paleolithic groups and N_ITA. A plausible theory, recently supported by studies on both uniparental markers and autosomal SNP-chip data (Boattini *et al.* 2013; Sazzini *et al.* 2016; Sarno *et al.* 2017), postulates an almost complete replacement of the local Paleolithic genetic background during the Neolithic migrations to Northern Italy and the reappearance of early European genetic components reintroduced since 4.5 kya by populations retaining consistent pre-Neolithic ancestry proportions, such as the Yamnaya pastoralists from the Caspian steppe during the Early Bronze age (Haak *et al.* 2015; Mathieson *et al.* 2015; Allentoft *et al.* 2015; Olalde, 2018). For example, the "El Miron" Cluster, dated 19-14 kya, retained an ancestry fraction that was found all over Europe before the Last Glacial Maximum (LGM) and was reintroduced in the continent from the Southwestern refugia (Fu *et al.* 2016). Similarly, the closeness to the "Villabruna Cluster" may attest the influence exerted by the spread of the Epigravettian culture on the Northern Italian genomic background since the end of the LGM (Fu *et al.* 2016).

2.3.4. Inferring effective population sizes and divergence time between Italian clusters

Population size dynamics in terms of change in effective population size (N_e) and genetic separation between N_ITA and S_ITA clusters were explicitly modeled Sequential Markov Coalescent + plenty of unlabeled samples (SMC++) method, also including a population of Northern and Western European ancestry (CEU) for comparison purposes. Results highlight that ancestors of all groups experienced a decline in N_e in the period 130-70kya, possibly reflecting the characteristic bottleneck (reduction in population numerosity and genetic diversity) that happened during and after the Out-of-Africa migrations of modern humans. Colonization of Europe was followed by a characteristic demographic expansion around 30 Kya, during which the ancestors of Italians maintained an increasing N_e that is higher than the Central European test cluster (**Figure 2.3.4.1**). This result agrees with previous estimations of past population size for Southern and Continental European populations (Schiffels and Durbin, 2014). Moreover, if a scenario with no post-divergence gene flow is applied, separation between N_ITA and S_ITA was found to clearly occur around 9 kya (**Figure 2.3.4.1**).

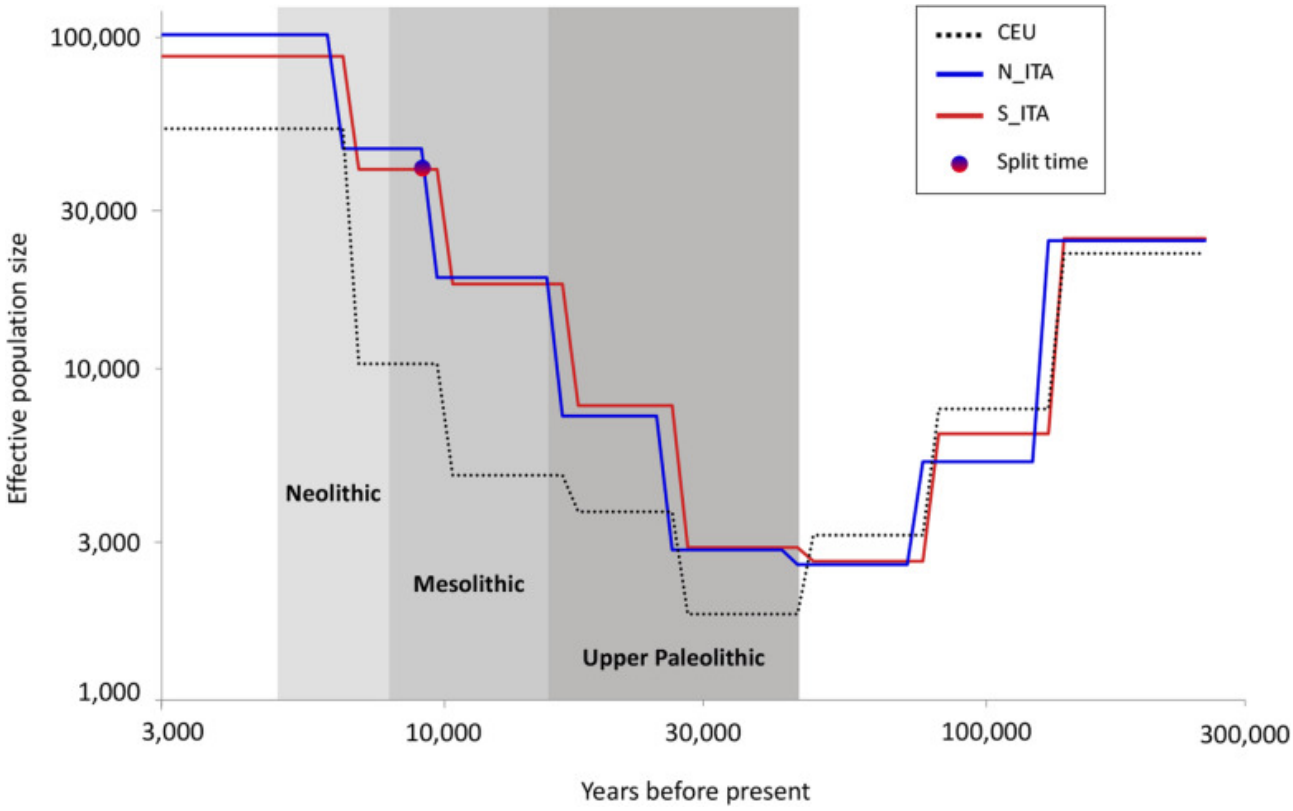


Figure 2.3.4.1. Results from the SMC++ method. Coalescent-based inference of effective population size trajectories of N_ITA and S_ITA groups are indicated in blue and red, respectively. The inferred trajectory for the reference population, CEU, is marked as a dotted line. From Sazzini *et al.* (2020).

Overall, results of the coalescent-based SMC++ method fits well with the differentiation in ancestry composition explained for N_ITA and S_ITA. A higher N_e deduced for S_ITA up to the Late Glacial Maximum might be fitting with the idea of Southern Italy as a refuge during this period (Sarno *et al.* 2014; Sazzini *et al.* 2014; De Fanti *et al.* 2015b), although its statistical significance is impossible to be tested. A more substantial contribution to gene flow in S_ITA may also be causative of the difference in N_e observed with N_ITA, as already estimated for Southern and Continental European groups (Schiffels and Durbin, 2014). Moreover, we have to consider that the estimated population split time is a rough estimate, based on the theoretical assumption that gene flow did not occur after N_ITA and S_ITA diverged. In this case, genetic differentiation between the Italian clusters can be pushed back to the end of the LGM (**Figure 2.3.4.1**), when post-glacial demographic processes started to differentially shape the gene pool of the two groups.

2.3.5. Disentangling the action of positive and balancing selection on the Italian genomes

Instances of positive and balancing selection acting on ancestors of N_ITA and S_ITA populations were identified through three statistical tests: the derived intra-allelic nucleotide diversity (DIND) score, the number of segregating sites by length (nSL) test, and the BALancing selection LikElihood Test (BALLET). The distributions of these statistics were then given as input for gene

network analyses aimed at modeling polygenic adaptation. All the presented analytical steps were repeated on the IBS and CEU populations from the 1000 Genomes Project, and signatures of selection shared between any of them and any Italian group were removed. The filtering approach allowed to limit the selective signatures most probably dependent on environmental pressures peculiar of the Italian Peninsula.

According to the DIND test, both Italian clusters show signals of positive selection for the *RIMS2* and *PCLO* genes involved in *insulin exocytosis* (**Figure 2.3.5.1** and **Table 2.3.5.1**). The nSL also identified a gene network in the *insulin secretion* pathway in the N_ITA cluster but made up of different loci (**Figure 2.3.5.2**). Several of them (*ADCY2*, *ADCY3*, *ADCY9*, and *GNAS)* also play a role in the regulation of lipolysis and thermogenesis; adenylate cyclase (ADCY) genes make the highest number of connections and are relevant to the *longevity regulating* pathway. Variants encoding for the main subunit of long-lasting high voltage activated calcium channels (*CACNA1C* and *CACNA1D*) are also involved in the development of type II diabetes (Reinbothe *et al.* 2013).

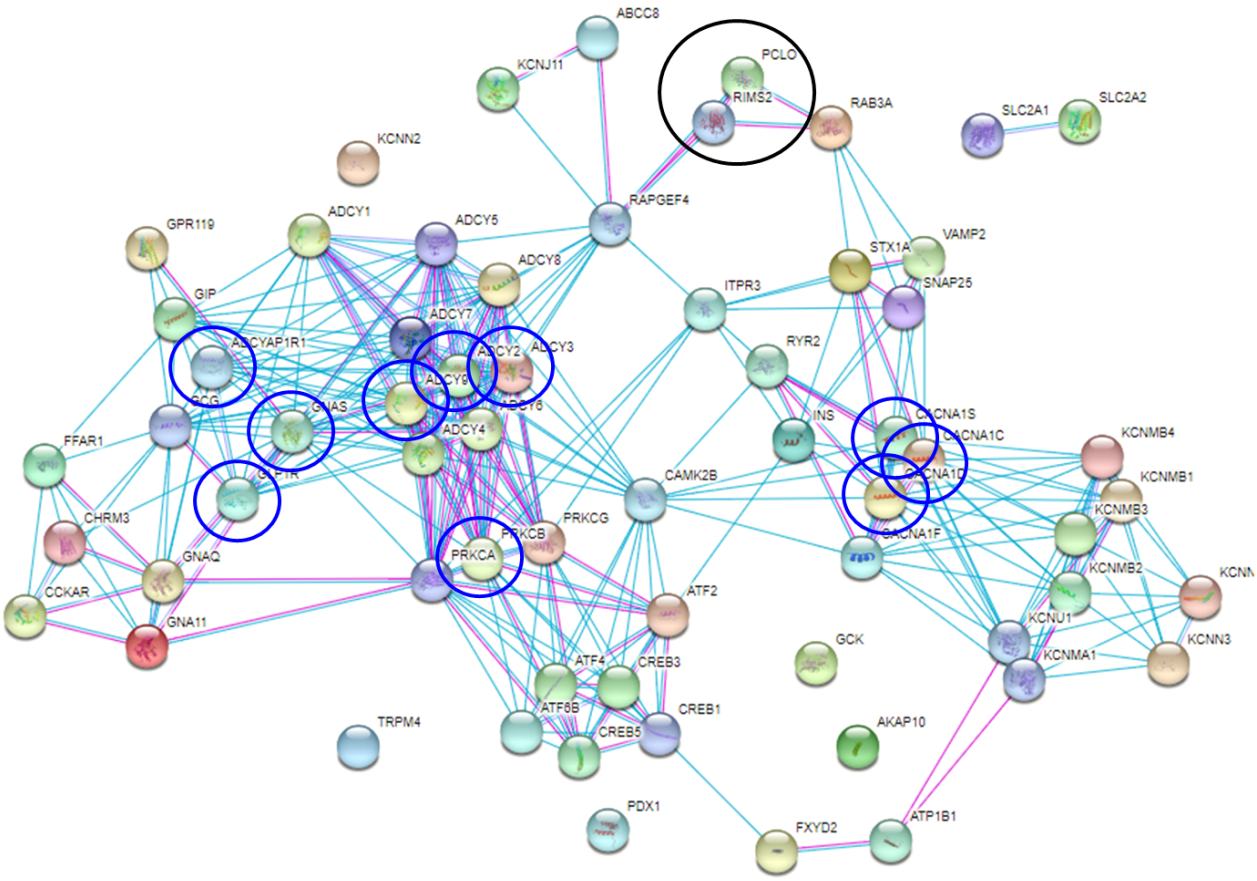


Figure 2.3.5.1. The *Insulin secretion* pathway. Supported interactions (from experimental data in pink; from database information in light blue) among genes in the *Insulin secretion* pathway (ko04911) were inferred on STRING using the reference gene list present in the Kyoto Encyclopedia of Genes and Genomes (KEGG). Blue circles highlight genes subjected to positive selection in N_ITA only; loci with black circles appear under positive selection in both N_ITA and S_ITA clusters. Most of genes under selection concur to the composition of the *Longevity regulating* (ko04211), *Thermogenesis* (ko04714), *Glucagon signaling* (ko04922), *Regulation of lipolysis in adipocytes* (ko04923), *Type II diabetes mellitus* (ko04930) pathways. From Sazzini *et al.* (2020).

Table 2.3.5.1. *signet* analysis on the genome-wide distribution of DIND scores.

Cluster	Pathway	Pathway size	Network size	HSS	p-value	Genes
N_ITA	Insulin secretion	50	2	7.238	0.034	*RIMS2 PCLO*
S_ITA	Insulin secretion	50	2	7.972	0.032	*RIMS2 PCLO*
	Mucin type O-glycan biosynthesis	29	5	8.386	0.023	*GALNT10 C1GALT1 GALNT13 GALNTL6 ENSG00000259075*

Pathway size, number of genes belonging to the pathway; Network size, number of genes composing the identified network; HSS, highest scoring subnetwork.

Evolutionary adaptation of the *Mucin type O-glycan biosynthesis* pathway was supposed to have happened in S_ITA, according to DIND results (**Figure 2.3.5.2** and **Table 2.3.5.2**). The identified network encodes for mucins, glycosylated proteins that are found on mucosal surfaces and cellular membranes and prevent pathogen adhesion (McGuckin *et al.* 2011).

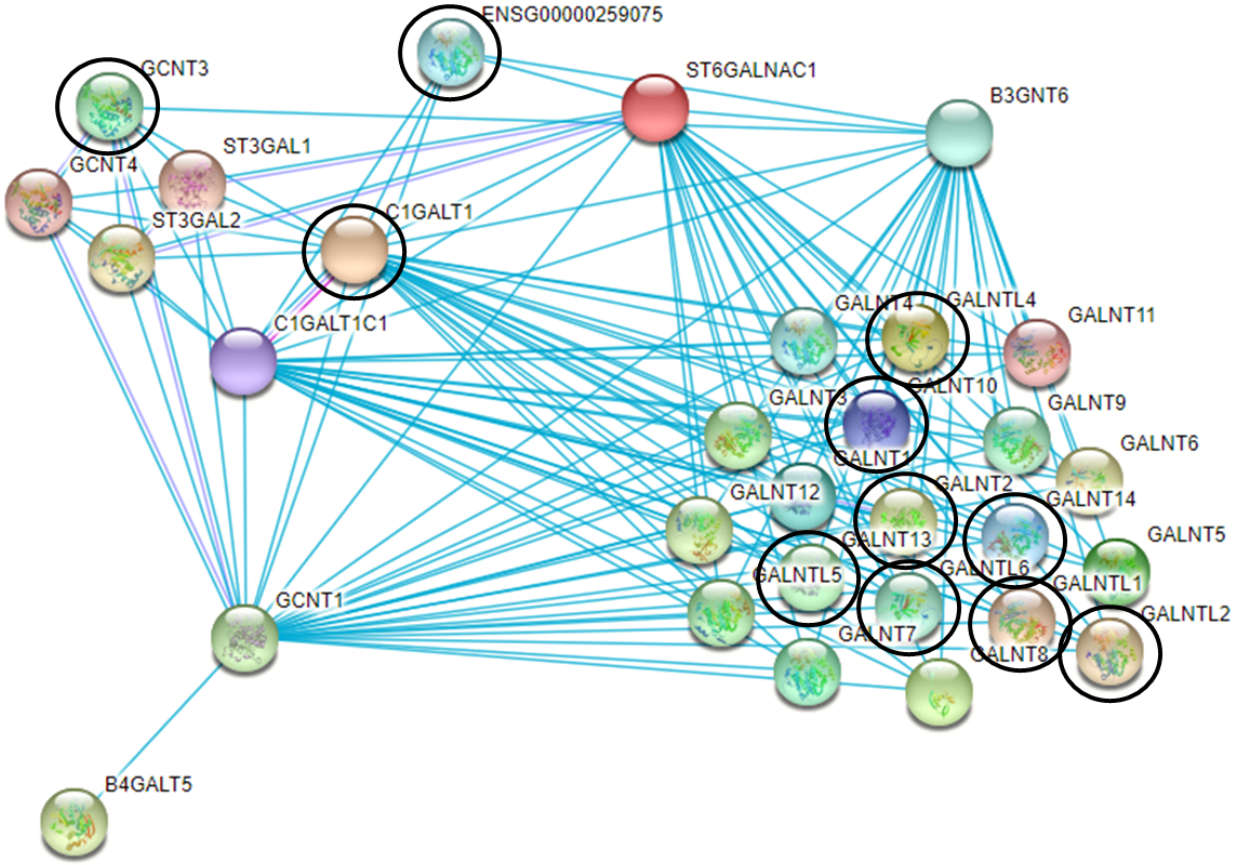


Figure 2.3.5.2. The *Mucin type O-glycan biosynthesis* pathway. Supported interactions (from experimental data in pink; from database information in light blue) among genes in the *Mucin type O-glycan biosynthesis* pathway (ko00512) were inferred on STRING using the reference gene list present in the Kyoto Encyclopedia of Genes and Genomes (KEGG). Loci with black circles appear under positive selection in S_ITA clusters. From Sazzini *et al.* (2020).

Table 2.3.5.2. *signet* analysis on the genome-wide distribution of nSL scores.

Cluster	Pathway	Pathway size	Network size	HSS	p-value	Genes
N_ITA	Insulin secretion	50	10	6.556	0.016	*ADCY2 ADCY3 ADCY9 ADCYAP1R1 CACNA1C CACNA1D CACNA1S GLP1R GNAS PRKACA*
S_ITA	Mucin type O-glycan biosynthesis	29	11	5.573	0.034	*GALNT2 GCNT3 GALNT10 C1GALT1 GALNTL1 GALNTL6 GALNT14 GALNT13 GALNTL2 GALNTL4*
	Basal cell carcinoma	45	13	5.542	0.035	*BMP2 FZD2 GLI2 GLI3 PTCH1 WNT7A WNT7B WNT8A WNT2B FZD1 FZD10 WNT4 WNT5B*

Pathway size, number of genes belonging to the pathway; Network size, number of genes composing the identified network; HSS, highest scoring subnetwork.

nSL results support the findings based on the DIND statistics in S_ITA, highlighting a gene network in the *mucin type O-glycan biosynthesis* pathway (**Figure 2.3.5.2** and **Table 2.3.5.2**). Furthermore, genes belonging to the the *basal cell carcinoma* pathway are found to be positive selection (**Table 2.3.5.2**). Most of these genes encode for G protein-coupled receptors (FZD) and Wnt glycoproteins that are relevant to melanin production.

Events of balancing selection at several different gene networks were inferred for both the Italian population clusters (**Table 2.3.5.3**): a cluster of loci, known to play a role in the *FoxO signaling* and *longevity regulating pathways,* included several ADCY genes, some of which (*ADCY2* and *ADCY9*) were identified by scans for positive selection in N_ITA, as well as mitogen-activated protein kinase (MAPK) genes (**Table 2.3.5.3**); different network was related to glycerolipid and arachidonic acid metabolisms, as it included glycerol phosphate acyltransferase (*GPAT/AGPAT/MBOAT*), diacylglycerol kinase (*DGKA*), lipid phosphatase (*LPIN/PLPP*) and phospholipase (e.g., *PLA2G*/*PLB*). N_ITA was characterized by two gene networks, one pertinent to glycolysis/gluconeogenesis and a second one including protein kinase C (*PRKC*) and phospholipase (*PLCG/PLCB*) loci, which play a relevant role in *glucagon signaling*, and *insulin resistance* pathways (**Table 2.3.5.3**). A network specifically under balancing selection in S_ITA is the *longevity regulating pathway*, characterized by MAP kinase 1 (MAPK1), phospholipases (PLA2G4) and the metabotropic glutamate receptor, GRM1 (**Table 2.3.5.3**).

Table 2.3.5.3. *signet* analysis on the genome-wide distribution of BALLET scores.

Cluster	Pathway	Pathway size	Network size	HSS	p-value	Genes
N_ITA	FoxO signaling	59	16	4.849	0.017	*ADCY2 ADCY5 ADCY8 ADCY9 CACNA1A CACNA1B CNR1 GNG10 KCNJ6 PRKACG MAPK1 MAPK11 MAPK10 MAPK12 GNG13 GNG2*
	Glycerolipid metabolism	87	12	4.648	0.024	*DGKB PLA2G1B PLD1 PLA2G6 PLPP1 DGKI CEPT1 PEMT PISD PLA2G3 LPCAT1 PLB1*
	Glycolysis/ gluconeogenesis	30	10	4.764	0.019	*ALDH2 ALDH3A1 ALDH1A3 ALDH3B2 ALDH3A2 AOC2 GAD2 AOC3 UPB1 CARNS1*
	Glucagon signaling/insulin resistance	88	7	4.367	0.039	*ITPR1 PLCG2 PRKCE PRKCG PRKCH PLCB1 TRPV4*
S_ITA	FoxO signaling	59	20	4.242	0.031	*ADCY1 ADCY2 ADCY5 ADCY6 ADCY8 ADCY9 CACNA1A CACNA1B CNR1 GNG7 GNGT1 GNGT2 KCNJ3 KCNJ5 KCNJ6 MAPK1 MAPK11 MAPK10 MAPK12 GNG2*
	Glycerolipid metabolism	87	27	4.843	0.012	*CDS1 DGKB DGKG PLA2G4A PLD1 PLA2G6 DGKD PLA2G4C PLPP1 PLPP3 CDS2 DGKI LPIN2 CEPT1 PEMT LPIN1 PISD PLA2G3 LPCAT2 AGPAT4 GPAM SELENOI PLD4 PLB1 MBOAT1 PLA2G4F PNPLA7*
	Longevity regulating	55	9	4.224	0.033	*GNA12 GNAZ GRM1 PLA2G4A PPP2R1A PRKCG MAPK1 PLA2G4C PLA2G4F*

Pathway size, number of genes belonging to the pathway; Network size, number of genes composing the identified network; HSS, highest scoring subnetwork.

According to the presented results, insulin-related genes seem to be under complex positive and balancing selection along the Italian peninsula. In particular, it appears as though the first selective events related to insulin modulation have occurred in the common ancestors to both N_ITA and S_ITA (**Figure 2.3.5.1** and **Table 2.3.5.1**), while a distributed event of more recent positive selection characterized ten genes in Northern Italian people only, regulating key processes pertaining to glucose homeostasis (**Figure 2.3.5.2** and **Table 2.3.5.2**). In particular, ADCY genes (especially *ADCY3*) are essential in thermogenesis processes, adiposity control and modulate predisposition to T2D and obesity (Cooper, 2003; Keele *et al.* 2018; Tian *et al.* 2018). Moreover, balancing selection tests revealed specific adaptive events in N_ITA: for ALDH genes involved in glycolysis, as well as PRKC and PLCG/PLCB in insulin resistance and the onset of diabetic complications (**Table 2.3.5.3**).

We can, then, speculate that the signatures of selection shared between N_ITA and S_ITA clusters might exemplify the legacy of retreating human groups towards Southern Italy during the LGM (Sarno *et al.* 2014; Sazzini *et al.* 2014; De Fanti *et al.* 2015b). In a colder environment and with an animal-based high-energy/high-fat diet (Vercellotti *et al.* 2008), ancestral populations could freely admix for 10,000 years, long enough to have optimized their energy metabolism. This proposition finds further support when it is considered that most populations of Western European ancestry showcase adaptive events dating to the LGM and strongly correlating with climate cooling and short-term temperature instability as main drivers of selection (Büntgen *et al.* 2011; Quagliarello *et al.* 2017).

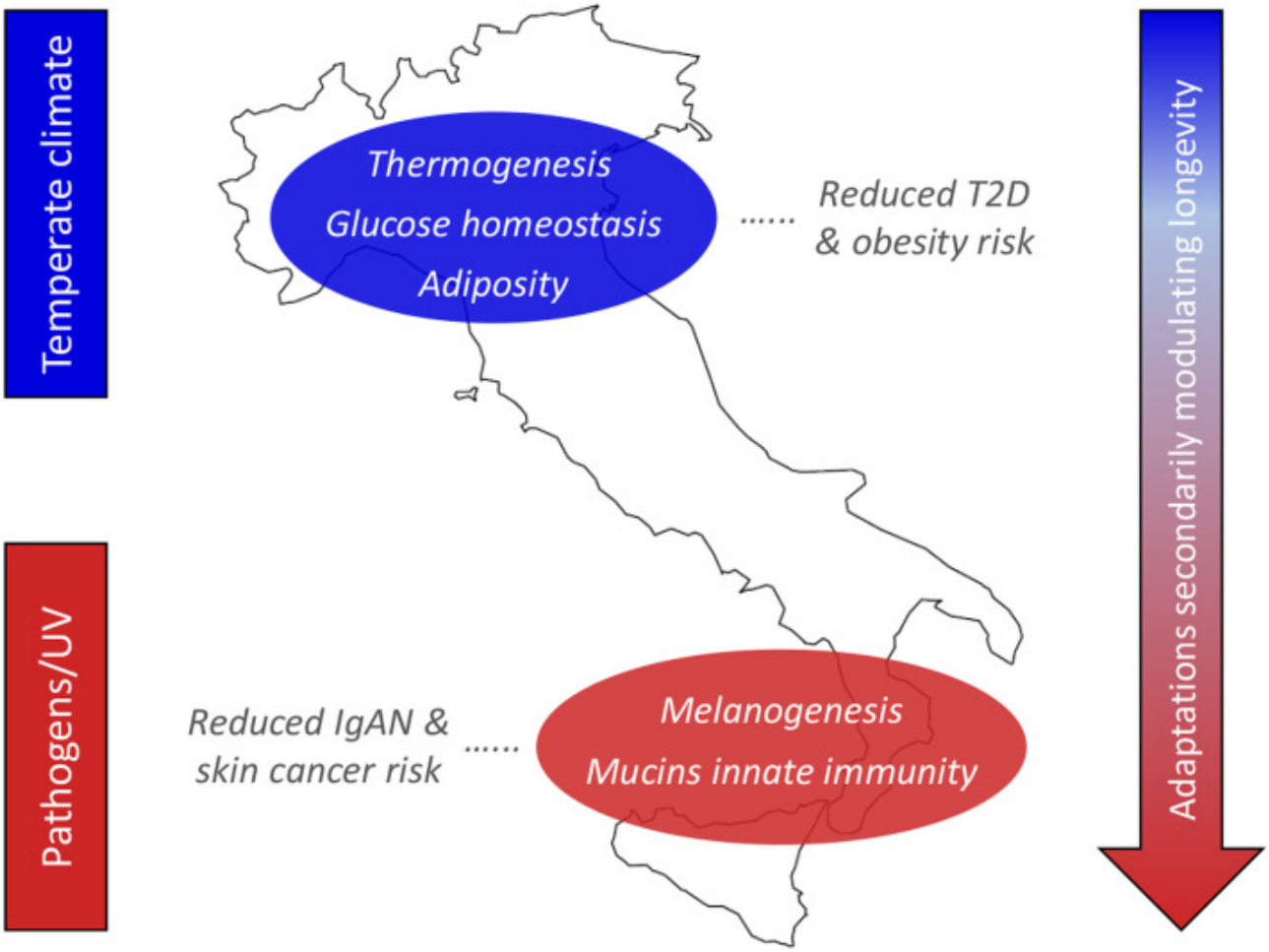


Fig. 2.3.5.4 Environmental pressures, adaptive events and health consequences. Putative selective pressures prompting local adaptations are on the left; biological processes under selection in groups ancestral to modern Italians are reported in the center, on a colored background (blue for N_ITA, red for S_ITA); health repercussions on modern Italians are reported in grey text. The cline of adaptations potentially modulating the longevity phenotype, especially in people from Southern Italy, is symbolized by the arrow. From Sazzini *et al.* (2020).

Populations expanding towards Northern Italy during the Late Glacial would have followed the edge of the retiring ice sheets, continuously experiencing a colder climate with respect to individuals in the Southern Italian region, thus being subjected to a seemingly constant selective pressure for several more millennia, until recent historical times (Kaniewski *et al.* 2013; Tournebize *et al.* 2019). This climatic background could have contributed, along with differential gene flow and ancestral contributions to Northern and Southern Italy in relatively recent times, to the differentiation of selection signatures between the two Italian groups, with prolonged selective pressures especially on the ancestors of the N_ITA cluster that stimulated adaptations in metabolic pathways related to insulin sensitivity, adipose tissue and thermogenesis (**Figure 2.3.5.4**). As most of the loci specifically under selection in N_ITA, when dysfunctional, are related to the insurgence of obesity and type 2 diabetes, these evolutionary events may have biomedical relevance: today's richer nutritional environment, in fact, may be counterbalanced by past adaptive evolutions to a sparce but highly caloric diet, as reflected by the incidence of T2D being almost halved in Northern Italy (https://www.istat.it/it/files/2017/07/REPORT_DIABETE.pdf). This observation also corroborates contemporary research concerning *ADCY3* as a target for the elaboration of anti-obesity drugs (Wu *et al.* 2016; Grarup et al. 2018).

Putative targets of selection specific to the S_ITA cluster included loci in the mucin synthesis pathway and in melanogenesis (**Tables 2.3.5.1** and **2.3.5.2**). *C1GALT1* in the mucin pathway is particularly relevant, as several genome-wide association studies indicated a correlation of some of its variants with the most diffused kidney inflammation, immunoglobulin-A nephropathy (IgAN; Gale *et al.* 2017). Epidemiologic data point out a higher prevalence of this pathology in Northern Italian regions than in Southern Italy (Kiryluk *et al.* 2012). Moreover, several microorganisms, such as *Pseudomonas aeruginosa*, *Entamoeba histolytica* and *Burkholderia cepacia* are known to enzimatically deactivate mucins to overcome mucosal barriers and their geographical distribution positively correlates to temperature, but negatively with IgAN incidence (Bevivino et al. 2002; Blessman et al. 2003; Gale *et al.* 2017). This observation points to a potentially higher frequency of exposure to these and other similar pathogens in the warmer Southern Italian climate after the LGM, triggering ancestral adaptive evolution and justifying reduced susceptibility to IgAN in modern S_ITA individuals (**Figure 2.3.5.4**).

Moreover, the selection signatures at FZD/Wnt genes involved in melanogenesis, that are specific for Southern Italy, may have been triggered by the exposure to higher UV-B radiation, thanks to a mean value of annual solar radiation that is nearly double with respect to Northern Italy (**Figure 2.3.5.4**). The FZD/Wnt genes, under selection in S_ITA, regulate the expression of the microphthalmia-associated transcription factor (*MITF*), which modulates pigmentation genes

(e.g., *TYR*, *TYRP1*, and *TYRP2*) and, by extension, melanogenesis in response to environmental stimuli (D'Mello *et al.* 2016). The participation of the FZD/Wnt genes under selection in the *basal cell carcinoma* pathway is also supported by studies revealing an oncogenic role of the melanin-regulating loci in skin cancers (Levy *et al.*2006), as it is known that extended UV exposure is the main risk factor for the development of skin malignancies. Several studies pointed out that protection to UV damage would also prevent micronutrient photodegradation in the skin and favor sweat-mediated thermoregulation (Jablonski and Chaplin, 2000), thus pointing to reduced skin cancer predisposition as a secondary advantage in S_ITA (**Figure 2.3.5.4**); this is in agreement with the incidence of melanomas in Southern Italy being half of what reported for the northern regions (Palmieri *et al.* 2015).

2.3.6. Pleiotropic genes and the longevity phenotype

Several of the biological processes under selection, especially in the S_ITA cluster, include loci that are also related to the achievement of the longevity phenotype (**Figure 2.3.5.4** and **Table 2.3.5.3**). Positive selection of the FZD/Wnt genes in the *mTOR signaling* pathway are particularly relevant, as variants at these loci have been shown to deter age-related ailments or to directly influence the attainment of longevity in humans (Weichhart, 2018). It is also notable that several genes under selection in S_ITA explicitly belonging to *the longevity regulating* pathway, especially different mutations in phospholipase A2 isozymes that have been invariably linked with the development of neurodegenerative diseases (Yagami *et al.* 2014; Ong *et al.* 2015; Deng and Li 2019), indicating that protection from neurological ailments could enhance the attainment of longevity as a healthy phenotype. These observations also corroborate recent findings showing that Italian centenarians, irrespective of their geographic origin, genetically cluster with people from Southern Italy (Giuliani *et al.* 2018a; Giuliani *et al.* 2018b).

Selection around loci responsible for the synthesis of arachidonic acid also suggests that adaptive events in response to pathogen exposure, when maintained in the modern Italian gene pool, may be secondarily useful to counterbalance the pro-inflammatory effects of contemporary diets, contributing to longevity (Giuliani *et al.* 2018b). Balancing selection around genes in the FoxO signaling pathway is also relevant to exceptional longevity, as this pathway exerts control and monitoring of metabolic stress induced by dietary restriction, insulin imbalance and pathogen exposure (Li *et al.* 2009). This is also a novel observation, as dietary- and pathogen-driven selective pressures have been observed so far only at the single-gene level, for *FOXO3* (Anselmi et al. 2009; Flachsbart *et al.* 2009; Flachsbart *et al.* 2017).

Chapter 3: Worldwide genetic variability of APOE in longevity

3.1. Introduction

Candidate gene studies and genome-wide analyses pertaining the genetics of extreme longevity as a healthy phenotype invariably pinpoint apolipoprotein E (APOE) as one of the most influential genes in achieving old age. In fact, its multiple effects and its pattern of variability across human populations offer intriguing perspectives on the evolutionary relationships between human genetics and environmental variables. In the following paragraphs, the APOE gene and protein are described in terms of structure, together with their physiological functions and pathological implications; worldwide human variability of APOE is introduced and integrated by a novel analysis of ancient and modern human data; finally, APOE trade-off mechanisms that would explain the conservation of deleterious isoforms are explained.

3.2. APOE structure and models

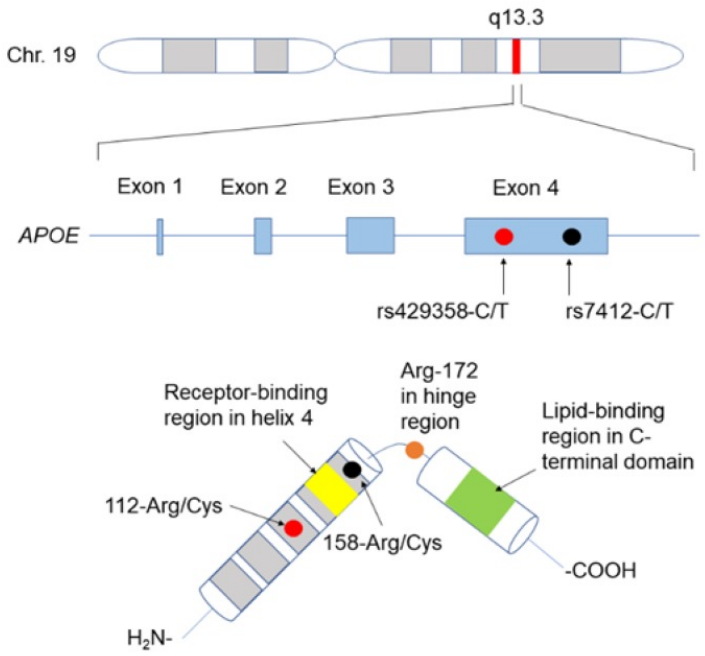
Human APOE is a 299-amino acid long circulating protein mainly produced by hepatocytes, adipocytes and astrocytes, with the chief function of collecting plasma lipoproteins and cholesterol (Tedla *et al.* 2004; Kockx *et al.* 2018). The lipid-free protein (**Figure 3.2.1a**) is characterized by two folded domains: a N-terminal elongated domain covering residues 1–167 and a more relaxed C-terminal domain spanning residues 206–299 (Wetterau *et al.* 1988; Wilson *et al.* 1991). Connecting the domains is an unstructured hinge that allows for extreme flexibility (Mahley *et al.* 1984).

Functionally, the protein links to lipids through a stretch of hydrophobic residues at the end of the C-terminal domain (residues 260–299), while the lipoprotein receptor is contacted by a binding region in the N-terminal domain, with an Arg-172 residue in the hinge putatively involved in the binding as well (Morrow *et al.* 2002). While it is the monomer that binds lipids, the C-terminal domain also allows for oligomerization of APOE in its lipid-free form, so the rate of protein lipidation directly depends on oligomer dissociation (Morrow *et al.* 2002; Chou *et al.* 2005; Huang *et al.* 2011).

The gene itself is located on chromosome 19 (**Figure 3.2.1a**), together with genes APOC1, APOC2 and APOC4, as well as the mitochondrial translocase of the outer membrane gene (TOMM40): these loci are all involved in the development of Alzheimer's disease (Roses *et al.* 2016; Subramanian *et al.* 2017; Kulminski *et al.* 2018).

Three main APOE isoforms, called ε2, ε3, and ε4 (or *APOE2*, *APOE3* and *APOE4)* arise from the combination of the alleles in the two variants rs7412 C/T and rs429358 C/T (Weisgraber, 1990; Chetty *et al.* 2017): isoform ε3 has a cysteine in position 112 and an arginine residue in position 158; isoform ε2 has two cysteine residues; isoform ε4 has two arginine residues (**Figure 3.2.1b**).

A)



B)

Allele	ε2	ε3	ε4
Haplotype	rs429358-T rs7412-T	rs429358-T rs7412-C	rs429358-C rs7412-C
Residue combination	112-Cys 158-Cys	112-Cys 158-Arg	112-Arg 158-Arg

Figure 3.2.1: Variants determining three human isoforms of APOE. A) From top to bottom: chromosome location; gene structure with identity of the polymorphisms in the gene; consequent mutating residues in the protein. Red and black dots highlight the correspondence between the two variant sites in the genomic sequence and in the protein. B) Table reporting the combination of alleles and corresponding residues associated to each APOE isoform. From Abondio *et al.* (2019).

The change in structural features of APOE, given by the different residue combinations, provides insight on the different behavior of its isoforms (Huang and Mahley, 2014; Nguyen *et al.* 2014; Henry *et al.* 2018). For example, the affinity of the protein for the lipoprotein receptor is greatly reduced by the loss of an electrostatic interaction given by the Arg158Cys mutation in isoform ε2 (Weisgraber *et al.* 1982). The mutation Cys112Arg in isoform ε4 does, in turn, modify the preference of the ligand-receptor structure to lower density lipoproteins by establishing trans-domain interactions that modify the protein structure (Dong *et al.* 1996). Mutagenesis experiments proved effective in re-establishing the functionality of the isoforms to those of "normal" APOE3 (Weisgraber, 1990; Dong *et al.* 1996).

Denaturation experiments verified that ε2 N-terminal domain is the most stable, followed by ε3 and ε4, which is the least resistant, but shows a higher number of stable intermediates between folding and unfolding: this could be a pH-dependent phenomenon that maintains ε4 in partially

unfolded states that facilitate conformational change and speed up the rate of ligation (Gursky and Atkinson, 1996; Acharya *et al.* 2002; Morrow *et al.* 2002). Lower plasma concentration of the protein in APOE4 homozygotes may be explained by higher ε4 catabolism (Martinez-Morillo *et al.* 2014), although recent studies have postulated that intermediate-state APOE4 is more sensitive to cleavage and that trans-domain interactions may render it more easily recognized as misfolded, especially in the brain (Williams *et al.* 2015; Love *et al.* 2017).

3.3. APOE function and pathology

The most diffused isoform, ε3, is crucial for maintaining a balanced lipid metabolism, being a structural component of cholesterol-rich lipoproteins and redirecting any excess of cholesterol to the liver for elimination (Mahley *et al.* 2006; Ang *et al.* 2008). However, the *APOE* genotype is involved with the development of several debilitating pathologies: for example, inheriting one or two copies of *APOE4* increases, respectively, four times and twelve times the risk of developing Alzheimer's disease (Hyman *et al.* 1996; Spinney, 2014), which is also in line with accelerated formation of β-pleated amyloid. Furthermore, clinical data from a United States cohort report that 50% of AD is associated with the ε4 allele (Raber *et al.* 2004).

Knock-out experiments have proved that mice not expressing APOE have a shorter lifespan, caused by the onset of an altered lipoprotein profile, neurological disorders, type II diabetes and immune response deficits (Robertson *et al.* 2000; Moghadasian *et al.* 2001). This observation highlights the pleiotropic effect of *APOE,* being simultaneously involved in multiple phenotypes, depending on the site of protein synthesis. Studies also emphasize that its impairment leads to a decline in biological functions and elevated susceptibility to disease, which increases mortality with age (Kregel *et al.* 2007; Vasto *et al.* 2007).

According to GWAS and whole genome sequencing studies, isoform ε4 in the brain is the "functionally altered" form of APOE associated with an impairment of human longevity and the onset of AD (Ang *et al.* 2008; Broer *et al.* 2015). As relevant to its structural integrity in AD pathogenesis, the truncated segment ε4-165 folds into a more compact helical structure that was shown to have deleterious effects on Aβ clearance, favoring instead its deposition (Argyri *et al.* 2014; Dafnis *et al.* 2016). Enhanced lipid remodeling activity of APOE4 has also been linked with a higher degree of lysosome leakage in neurons, in association with Aβ (Ji *et al.* 2006).

Mice experiments have shown how isoform ε4 can also cause behavioral impairments similar to AD in humans, with spatial and memory deficits increasing with age and are observed primarily in females (Hartman *et al.* 2001; Bour *et al.* 2008). Similar studies showed that isoform ε4 inhibits neurite extension and branching, as well as GABAergic input in newly formed neurons (Li *et al.* 2009).

Interestingly, studies have highlighted a decrease in cerebral glucose metabolism that anticipates the cognitive impairment given by this isoform, which is indicative of interactions with the mitochondrial membrane proteins at early stages of the disease (Chou *et al.* 2005; Tanaka *et al.* 2006). Moreover, mitochondria and the endoplasmic reticulum (ER) are associated via mitochondria-associated membranes (MAMs) and their proteins, so that mitochondrial functionality can affect the secretory pathway in the ER (Roses *et al.* 2016, Eysert *et al.* 2020).

Moreover, the cholesterol-binding efficacy of APOE isoforms has been pointed out as a modulator of viral infection, since cholesterol is the main component of the envelope of many human-infecting viruses (Bankwitz et al, 2017). For example, herpes simplex virus HSV-1 is recurrently discovered in the brain of healthy elderly individuals as well as younger AD-affected patients, and the cause of neural decay at a younger age is ascribed to an enhancement of viral colonization and activation processes through ε4-stimulated inflammation (Itzhaki, 2016; Itzhaki, 2017; Itzhaki, 2018). In this case, antiviral therapy may be effective in slowing AD progression (Itzhaki, 2016; Itzhaki, 2017; Itzhaki, 2018). Now that HIV affected individuals can keep the infection at bay thanks to anti-retroviral therapy and can live to older ages with this chronic disease, several studies tried to investigate (with varying success) the correlation between APOE and modulation of HIV infection, revealing plausible correlations of isoform ε4 with premature brain aging, impaired cognition and neurocognitive disorders (Suwalak *et al.* 2015; Geffin and McCarthy, 2018).

However, most studies investigating APOE in disease link its isoforms to the frequency and development of cardiovascular diseases (CVD): one study including European and Chinese cohorts of middle-aged men demonstrated that ε4 frequency variation among populations could justify up to 75% of the difference in coronary heart disease (CHD) deaths (Stengard *et al.* 1998); another study on heart attack survivors highlighted that ε4 carriers have an 80% increased risk of a second heart attack (Gerdes *et al.* 2000); post-mortem studies also confirm that, in a cohort of patients that died of cardiovascular disease, there were significantly more APOE4 carriers (34%) and significantly less APOE2 carriers (12%) than in the rest of the group (29% and 14% respectively; p-value < 0.05) (Kumar *et al.* 2012). It is also noteworthy that APOE variants have demonstrably been associated with the level of circulating saturated and unsaturated fatty acids, so that a link has been established between APOE and nutrition in the occurrence of cardiovascular events (Satizabal *et al.* 2018).

Finally, as introduced in the previous paragraph 3.2, the SNPs related to the three isoforms of APOE that are being analyzed impact the quantity of CpG dinucleotide, which in turn modifies gene methylation of the CpG island they are located in. This feature allowed to differentiate AD-affected subjects from healthy controls in postmortem brain biopsies (Foraker *et al.* 2015) and it has been

suggested that DNA methylation of the circulating cells may be used as an early indicator for the development of dementia at an older age (Liu *et al.* 2018).

3.4. Studies on APOE genotype and human longevity

The involvement of APOE isoforms in several post-reproductive pathologies seems to be the main reason for its contribution to the complex genetics of the longevity phenotype (Pilling *et al.* 2017; Giuliani *et al.* 2018a; Revelas *et al.* 2018). Candidate gene studies, genome-wide association studies (GWAS) and WGS all suggest that the causative alleles are population-specific, but there are modulating mechanisms that are shared among populations (Giuliani *et al.* 2018b) and interesting discrepancies in results.

Concerning candidate gene studies, for example, a comparative analysis of centenarian cohorts from Italy, Spain, and Japan highlighted that ε4 was negatively associated with the attainment of extreme ages in a trans-ethnic fashion, while ε2 showed positive association in two of the three groups (Garatachea *et al.* 2014a). A study concerning a cohort of Ashkenazi Jewish ancestry (minimum age 95 years) also found a significant enrichment in the presence of the ε2 allele (*p-value* = 0.003) and the ε2/ε3 genotype (*p-value* = 0.005), as well as a depletion of the ε3/ε4 genotype (Ryu *et al.* 2016). Accordingly, two recent meta-analyses of polymorphisms associated with human longevity highlighted how the likelihood of reaching extreme longevity is negatively associated with carrying the ε4 allele in any genotype (*p-value* < 0.001), while it is positively associated with the ε2/ε3 genotype ($p = 0.017$) regardless of ethnicity; however, they obtained discordant results for the significance of the ε2 allele (Garatachea *et al.* 2014b; Revelas *et al.* 2018). A third meta-analysis involved data from AD cases ranging from 96 to 119 years and confirmed that carrying the ε2 allele is associated with a significant increase in the probability of reaching extreme longevity, with a declining risk reduction at the most extreme ages, while the opposite trend is observed for ε4 (Sebastiani *et al.* 2019). Finally, a noteworthy paper focused on trans-ethnic and trans-generational effects of variants affecting evolutionary fitness found few common variants significantly related to mortality, all tagging the *APOE* ε4 allele and the *CHRNA3* gene (Mostafavi *et al.* 2017).

Among the GWAS tackling the association between APOE isoforms and longevity, a study on Japanese centenarians successfully highlighted the polygenic nature of attaining extreme ages by finding a novel association between gene *FNDC5* (which is upregulated by physical exercise) and *APOE* variants in the oldest olds (Fuku *et al.* 2017). However, as pointed out in a recent paper, although it is feasible to replicate previous results for known APOE variants, the heterogeneous genetic influences on longevity and the contributions of rare variants introduces difficulties in finding new alleles associated with survival after 90 years of age (Broer *et al.* 2015). To bypass these problems, a new GWAS method informed by previous knowledge of age-dependent and disease-

related traits in longevity was developed, to amplify signals difficult to pick up in reduced centenarian cohorts. Accordingly, the *APOE*/*TOMM40* locus was found to be associated with longevity at GWAS significance ($p < 10^{-8}$) and the result was subsequently replicated in three validation cohorts (Fortney *et al.* 2015). On the contrary, a study on Chinese individuals with at least one parent over the age of 90 found no significant correlation between *APOE* isoforms and familial longevity (Wang *et al.* 2018), while the first study on an older Brazilian cohort (aged 85 or more) could not successfully replicate known associations between APOE and several phenotypes, possibly because of limited power of the 200 gathered samples (Silva-Sena *et al.* 2018).

Finally, it is worth hifhlighting a study on Northern European populations (Gerdes *et al.* 2000), which clarified the role of APOE as a "frailty gene" able to slightly influence mortality, rather than a "longevity gene" that guarantees long life by itself. Indeed, the risk of death associated with the ε4 variant does not disappear with age but persists even when individuals do achieve extreme longevity (Sebastiani *et al.* 2019).

3.5. APOE evolution and variability among human populations

From an evolutionary standpoint, a comparison of human and primate APOE protein sequences reveals that non-human APOE, just like human isoform ε4, has arginine in position 112, suggesting that ε4 is actually the ancestral variant (Huebbe and Rimbach, 2017). This hypothesis is also corroborated by the analysis of an archaic specimen of Denisova hominin found in the Altai Mountains in Siberia (Reich *et al.* 2010), while Neanderthal genomes did not provide reliable information so far. Since the other mutations do not alter the structure of the protein or are located in non-coding portions of the gene, the only difference between human and primate APOE involves a threonine residue in position 61, which became an arginine in *H. sapiens*. The Thr61Arg substitution produces a positively charged residue that projects Arg112 in ε4 outwards and allows for Arg61 to destabilize the protein structure, making it less stable, but more reactive. Several theories have been proposed for the evolutionary advantages that this mutation could have provided in the ancestors of modern humans, including protection from the damages of a high-cholesterol diet, exposure to food-driven pathogens and increasing brain size (Huebbe and Rimbach, 2017).

A very interesting hypothesis suggests that the increase in movement and physical activity throughout the evolution of the genus *Homo* could have counteracted the deleterious effects of APOE4, reducing the risk of CVD and secondarily supporting longevity (Raichlen and Alexander, 2014). In fact, it has been determined that isoforms ε2 and ε3 originated almost simultaneously 200,000 to 300,000 years ago (Fullerton *et al.* 2000), while the change to a more active lifestyle could be dated to the first hunter-gatherer groups of *Homo erectus,* up to 1.5 Mya: the aerobic activity necessary to walk and run long distances to follow prey or flee from danger imposed the conversion

of fat into energy for endurance and could have counteracted the effects of the ε4/ε4 haplotype (Bramble and Lieberman, 2004; Malina and Little, 2008). This observation is supported by paleodemographic analyses suggesting an increase in the proportion of older individuals throughout the evolutionary arc of *H. erectus* and *H. sapiens*, which coincides with an extension of post-reproductive lifespan with the acquisition of the hunter-gatherer lifestyle (Caspari and Lee, 2004; Raichlen and Alexander, 2014; Hawkes, 2016).

However, ε3 is currently the most frequent isoform in all modern populations, peaking at 94% in the Hutterite people of Canada, 90% in Mayas, 88% in the Basque and Sardinian populations, and 86% in Han Chinese people. Isoform ε4 shows the maximum percentage in a range of indigenous populations of Central Africa (40% in Aka Pygmies, 38% in Tutsis), Oceania (49% in the Hui population of New Guinea) and Central America (27% for the Huychol people of Mexico). Moreover, a decreasing North-South gradient (**Figure 3.5.1** and **Appendix A1**) characterizes this isoform in Europe (5% to 10% in the Mediterranean basin; 16% in Central and Western Europe; 23% in the Scandinavian peninsula, with peaks of 31% in Finland) and China (Zekraoui *et al.* 1997; Corbo and Scacchi, 1999; Singh *et al.* 2006; Hu *et al.* 2011; Ojeda-Granados *et al.* 2017). Isoform ε2 has an inconsistent distribution, with peaks in Southeast Asia, Australia, Africa (19%) but is absent Native American groups (Singh *et al.* 2006).

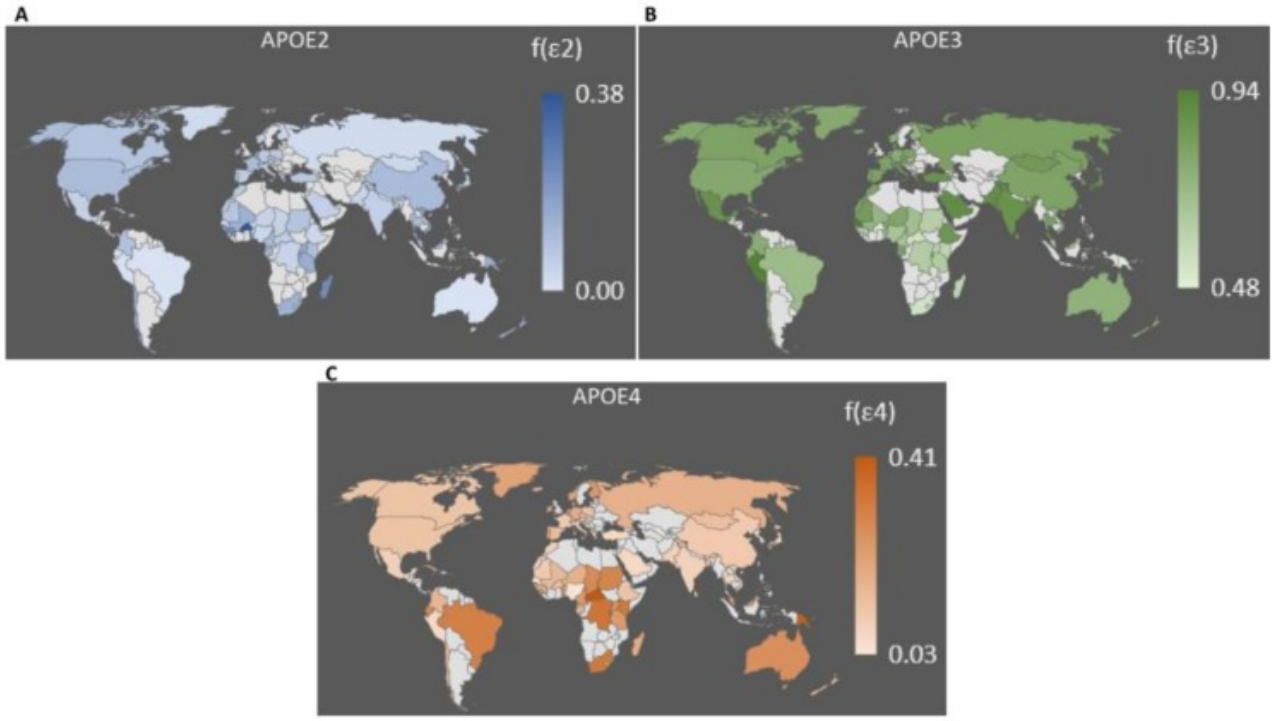


Figure 3.5.1. Frequency distribution of APOE alleles. Data have been recovered for 82 countries by combining data extrapolated from the 1000 Genomes Project with those reported in Singh *et al.* (2006). White areas represent absence of data. (A) Frequency distribution of the ε2 variant. (B) Frequency distribution of the ε3 variant. (C) Frequency distribution of the ε4 variant. From Abondio *et al.* (2019).

Given the differentiation in ancestry and selective pressures observed across the Italian peninsula, as described in Chapter 2, the presence of a latitudinal gradient has been tested by gathering a novel cohort of 134 Italian centenarians and 350 healthy controls, that have been clustered in three Italian macroareas (North, Center, and South Italy) according to their place of origin (Boattini *et al.* 2013; Sazzini *et al.* 2016). Analyzing the distribution of APOE genotypes, a definite North-South decreasing gradient could be observed for the ε4 allele in both centenarians and controls, with centenarians showcasing lower frequencies of the deleterious allele (**Table 3.5.2**). Moreover, a higher proportion of centenarians from Southern Italy carry the supposedly protective ε2 variant, when compared to all other groups. Although sample size is relatively small in centenarians, the frequency distributions follow what has been already observed at regional and continental level for other loci implicated in lipid metabolism (Sazzini *et al.* 2016; Ye *et al.* 2017; Buckley *et al.* 2017).

Table 3.5.2. Variants and genotype frequency for a new Italian dataset of centenarians and healthy controls.

Groups	n	f(e2)	f(e3)	f(e4)	f(e2/e2)	f(e2/e3)	f(e2/e4)	f(e3/e3)	f(e3/e4)	f(e4/e4)
N_cent	48	0,104	0,771	0,125	0,000	0,167	0,042	0,625	0,125	0,042
N_cont	97	0,108	0,768	0,124	0,000	0,155	0,062	0,608	0,165	0,010
C_cent	67	0,037	0,910	0,052	0,000	0,075	0,000	0,821	0,104	0,000
C_cont	215	0,060	0,877	0,063	0,000	0,112	0,009	0,763	0,116	0,000
S_cent	19	0,184	0,789	0,026	0,000	0,368	0,000	0,579	0,053	0,000
S_cont	38	0,079	0,882	0,039	0,026	0,105	0,000	0,789	0,079	0,000
TOT_cent	134	0,082	0,843	0,075	0,000	0,149	0,015	0,716	0,104	0,015
TOT_cont	350	0,076	0,847	0,077	0,003	0,123	0,023	0,723	0,126	0,003

f(e2), frequency of isoform APOE2; f(e3), frequency of isoform APOE3; f(e4), frequency of isoform APOE4; f(e2/e2), frequency of APOE2 homozygous; f(e2/e3), frequency of APOE2/APOE3 heterozygous; f(e2/e4), frequency of APOE2/APOE4 heterozygous; f(e3/e3), frequency of APOE3 homozygous; f(e3/e4), frequency of APOE3/APOE4 heterozygous; f(e4/e4), frequency of APOE4 homozygous.

This suggests that isoform ε3 may be selected against ε4 at lower latitudes, but this does not explain the evolutionary advantage of the single amino acidic mutation Arg112Cys provided in giving rise to the now most frequent *APOE* variant worldwide (Zekraoui *et al.* 1997; Singh *et al.* 2006; Hu *et al.* 2011). Studies trying to understand the strong evolutionary advantage of isoform ε3, which made it the most frequent variant of APOE, over the much more ancient ε4 suggest higher structural stability and an array of novel functions, such as metal binding, micronutrient uptake and enhanced neuronal repair (Huebbe *et al.* 2015; Huebbe *et al.* 2016; Huebbe *et al.* 2017). However, responsiveness to environmental changes and nutritional shifts does not explain why most ailments carried by ε4 nowadays tend to occur in the post-reproductive age, especially in females. Recently introduced cultural and kinship-based theories (the "grandmother effect") postulate that the extension of post-menopausal lifespan increases younger kin survival under grandparental care if cognitive

decline is prevented through a late-life selection of protective variants, such as APOE3 (Hawkes, 2016). It is known, in fact, that cultural transmission and multi-generational sharing of knowledge shape the social structure of modern foraging populations by enhancing the chance of survival of the entire group (Hawkes, 2016; Hawkes, 2020).

3.6. APOE trade-offs

Given the involvement of APOE4 in (especially post-reproductive) cardiovascular and neurodegenerative diseases, several recent studies tried to understand why such a deleterious variant predominantly affects individuals of affluent populations but has been maintained at particularly high frequencies in many indigenous human groups.

A study on a rural population from Ghana, characterized by high levels of mortality from infectious diseases, suggested an association with a higher degree of fertility (ε4 homozygous women have 3.5 more children on average) where limited survival spontaneously delays the harmful effects of APOE4. Mortality in modern societies is much less defined by pathogens, however, and the positive reproductive advantage is not needed, but the extension of lifespan implies a higher probability of manifesting detrimental repercussion in older age (van Exel *et al.* 2017). On the contrary, a significant association has been observed in industrialized countries between lipid clearance dysfunction induced by ε4 alleles in expectant women and recurrent pregnancy loss (Goodman *et al.* 2009; Yenicesu *et al.* 2010). Moreover, as introduced at the end of paragraph 3.5, the practice of trans-generational child caring may have been crucial in exceptionally prolonging post-fertile age and favoring the diffusion of the ε3 variant in early human clusters after the shift to a hunter-gatherer lifestyle. However, if the ε4 variant is present, the early positive effects of survival and growth of the younger individuals may be equally mirrored by cognitive deficiencies in older age (Hawkes *et al.* 2016).

Alternatively, a shift in dietary habits has been proposed to support the advantage introduced by isoform ε3 around 200,000 years ago, when the higher quantity of fat-rich meat provided by organized hunting methods enhanced the quality of the diet and concurred to the extension of lifespan. However, efficient cholesterol clearance and a rapid response to food-borne pathogens would be provided by the ancient isoform ε4 (Luca *et al.* 2010). A possible immunity-based explanation has also been introduced to explain the latitudinal distribution of APOE isoforms. In fact, ecology and biogeography both highlight that harmful bacteria tend to thrive in hot, wet climates (such as the Tropical belt) and in densely populated areas of the world (Guernier *et al.* 2004; Cashdan, 2014). In this context, it has been observed that individuals with the ε4 isoform have a higher activated T-cell count and a more ready immune system (Bonacina *et al.*, 2018). As described in paragraph 3.2, several viruses take advantage of the more common variant ε3 to invade human cells. In the case of

the hepatitis C virus, isoform ε4 may compete with the virus for access to LDL receptors, reducing damage in the affected population: in the Italian peninsula, the gradient of hepatitis C incidence overlaps with a reverse gradient in ε4 distribution (Picardi *et al.* 2007). On the other hand, studies have suggested that APOE4 may be deleterious in the case of Gram-negative infections, as their toxic lipopolysaccharides (LPS) can be collected in lipoproteins and redirected to the liver by APOE. A protein with reduced functionality may lead to hindered endotoxin clearance and a stronger immune response leading to sepsis (van Oosten *et al.* 2001).

Recently, the relationship between APOE and the gut microbiota has been considered of importance, especially for the double function of APOE in lipid assimilation and immunity. Knock-out mice fed on a chow with macronutrient proportions recalling a typically Western diet have shown an enrichment in intestinal Firmicutes, which stimulated a chronic inflammatory state (Kasahara *et al.* 2017). Immunizing mice against their own microbiota resulted in a reduction of circulating inflammatory cytokines (Saita *et al.* 2016). Similarly, restoring the loss of *Akkermansia muciniphila* caused by a Western diet contrasted vessel inflammation, fat dysmetabolism, and atherosclerosis (Zhao *et al.* 2017). These studies show that there is a balance between the immune response to intestinal microbes and the inflammatory response driven by APOE according to the diet, with local and systemic responses shaping the equilibrium of the intestinal biome (Saita *et al.* 2016).

Trade-off mechanisms may also justify difficulties in the replication of the same association signals for different human populations, to which it has to be added that many WGS studies now show how the genetics of complex traits, including longevity, may be explained by polygenic models in which smaller contributions from numerous rare variants contribute to the observable phenotype (Boyle *et al.* 2017; Gnecchi-Ruscone *et al.* 2018).

3.7. Ancient dynamics of APOE in Europe

In accordance with what presented in the previous two paragraphs, a theory was introduced that the ε4 variant is a relic of an ancient hunter-gatherer genetic background, that failed to adapt to (or, indeed, did not experience the selective pressures represented by) a modern, energy-rich, and exercise-poor lifestyle (Alvergne *et al.* 2016). To assess the possibility of a temporal description for the distribution of APOE in Europe, 97 publicly available ancient genomes, with both rs7412 and rs429358 already directly genotyped, were selected. This choice was aimed at avoiding bias that could be introduced in the dataset by an imputation procedure on highly deteriorated DNA, which presents lengthy stretches of missing data. The overall samples (**Figure 3.7.1** and **Appendix A2**) cover the Euro-Mediterranean area on a temporal range of 1,500 to 42,000 years ago. ε3/ε3 was the most frequent genotype (83%), followed by ε4/ε4 (13%), and ε2/ε2 (3%). A single ε3/ε4 heterozygote sample was represented by a 42,000-year-old specimen of early hunter-gatherer from Siberia.

Regarding the temporal distribution of the genotypes, ε2/ε2 individuals are limited to Northern European Bronze Age samples; moreover, most individuals carrying the ancestral genotype are more recent than 5000 years, while the majority of ε3/ε3 samples are much older, especially those coming from the Eastern part of Europe and the Near East. As highlighted by older local specimens (Allentoft *et al.*, 2015; Haak *et al.* 2015; Mathieson *et al.* 2015), both alleles could be found in Europe before and around the Late Glacial phase (12-15 Kya), with APOE3 possibly reintroduced from the Levant during the Neolithic Revolution (8-10 Kya) and the Paleolithic APOE4 carried at a higher frequency by the Yamnaya people of the Pontic Steppe during the Bronze Age (4-5 Kya). This is also in line with the ancestral contribution to the European and Italian genetic background described in Chapter 2. It is, however, acknowledged that the reduced number of samples, their patchy distribution and the possibility of genotyping errors for deteriorated specimens allow only for a limited discussion of the ancient samples.

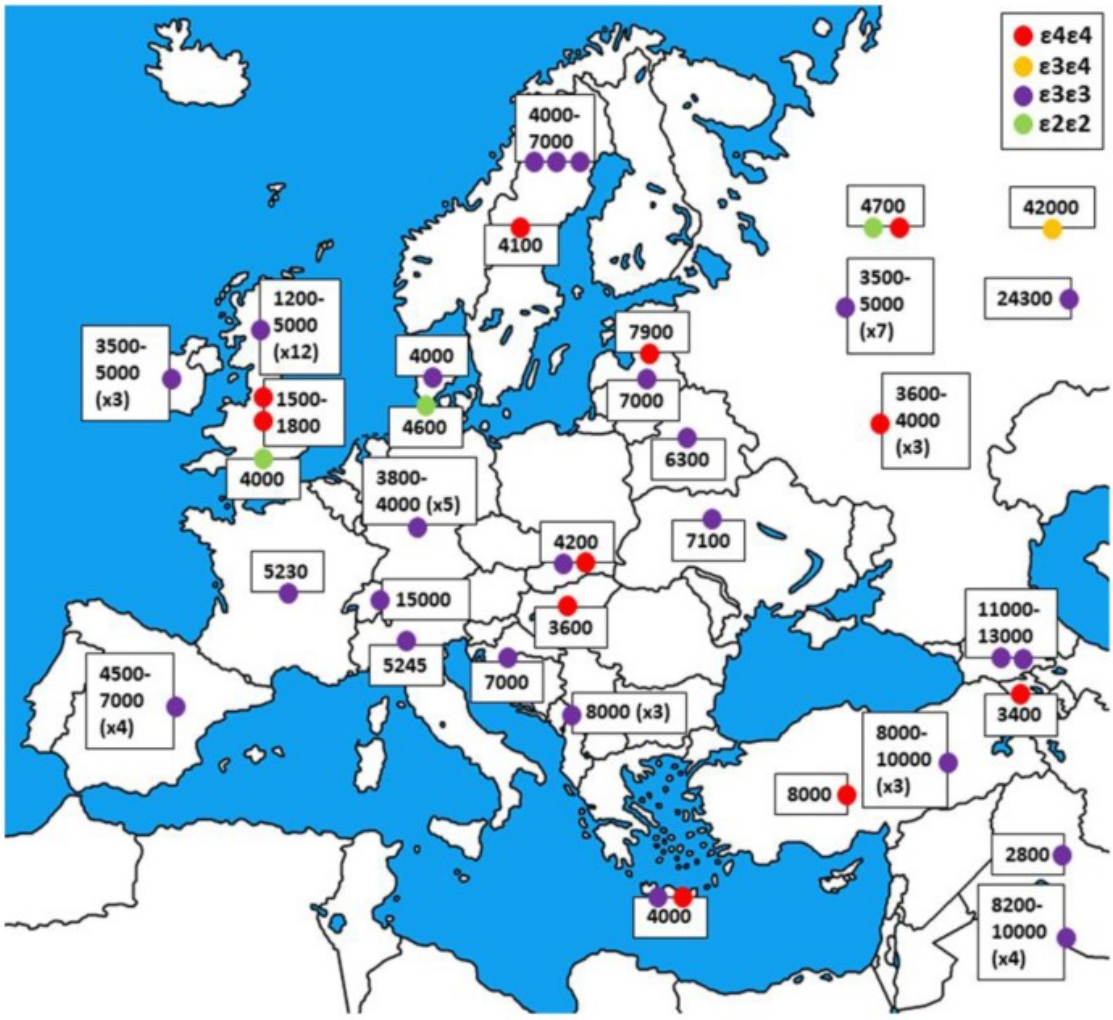


Figure 3.7.1. Distribution and estimated age of the ancient samples. Specimens belonging to the same location and culture have been grouped under the same label. The number of clustered individuals is given in brackets on the label. From Abondio *et al.* (2019).

Chapter 4: Genetics of the linguistic isolates of Aspromonte

4.1. Introduction

As described in Chapter 2, the Italian Peninsula presents an outstanding range of geographical and environmental variation, which has exerted differential selective pressures at the local level on a highly convoluted genomic background, given by numerous pre-historic and historical demo-geographic and admixture processes during the peopling of Europe and the Mediterranean. The genetic structure of the Italian population, in fact, is not defined only by the environmental and climate conditions that induced patterns of local adaptation (Sazzini *et al.* 2014; Sazzini *et al.* 2016), but geographic limitations to mobility, as well as ethnicity, language and cultural dynamics have also been deemed responsible for more subtle local levels of isolation and differentiation (Destro Bisol *et al.* 2008; Coia *et al.* 2013; Boattini *et al.* 2015).

Uniparental markers (Capelli *et al.* 2007; Turchi *et al.* 2008; Brisighelli *et al.* 2012; Boattini *et al.* 2013), autosomal SNP-chip data (Di Gaetano *et al.* 2012; Sazzini *et al.* 2016; Fiorito *et al.* 2016; Raveane *et al.* 2019) and whole genome sequencing studies, as presented in Chapter 2, have helped in disentangling the genetic makeup of the Italian peninsula as a whole. However, the study of geographically, culturally or ethno-linguistically isolated minorities offers notable advantages by limiting the genomic context of the analysis. In fact, trait-associated alleles may be found at much higher frequencies in groups with limited gene flow, smaller effective population size (Ne) and increased homozygosity, supporting and facilitating genome-wide association studies (Peltonen *et al.* 2000; Service *et al.* 2006; Kristiansson *et al.* 2008; Hatzikotoulas *et al.* 2004; Zeggini, 2014; Ayub *et al.* 2015; Cilli *et al.* 2019). In fact, several isolates present on the Italian peninsula have been analyzed by both autosomal and uniparental based population genetic studies (Destro Bisol *et al.* 2008; Coia *et al.* 2013; Esko *et al.* 2013; Capocasa *et al.* 2014; Sarno *et al.* 2016; Anagnostou *et al.* 2017; Anagnostou *et al.* 2019).

In a recent genomic analysis of Southern Italian and Mediterranean admixture (Sarno *et al.* 2017), the Calabrian Greek-speaking groups of the Aspromonte area in Reggio Calabria (Southern Italy) have been judged of peculiar interest, as these communities are characterized linguistically by speaking *Greco* or *Calabrian Greek,* a specific variety of Greek. Two contrasting historical-linguistic hypotheses tried to unravel the possible origin of these Hellenic remnants: the first one maintains that descendants of Byzantine people settled in Southern Italy at the beginning of the Medieval age (Pott, 1856; Comparetti, 1866; Pellegrini, 1880); the other hand, the second theory postulates that the language may have been introduced way before, during the Greek colonization of Southern Italy. In both cases, the Greek language would have been continuously spoken by these ethno-linguistic

enclaves until the present time (Chatzidakis, 1892; Rohlfs, 1932; Tsopanakis, 1977). A more recent hypothesis took into consideration both the influence of intense Greek-Romance linguistic exchanges with coexistence between Greek-speaking Latins and Latin-speaking Greeks, and the presence of lexical elements in common with peripheral Greek dialects, which may imply influences from the Greek of different periods (Trumper, 2016). According to these observations, Greek may have survived in diglossia with Latin for centuries and have been revitalized locally by the Byzantine expansions, as well as multiple waves of Greek speakers in subsequent periods (Mosino, 1988). It is known, in fact, that the area of Greek-influence in the past extended to the whole of Southern Italy, while today the number of Greek-speaking communities limited to those inhabiting the Aspromonte mountain area of Reggio Calabria (Martino, 1980; Violi, 2006; Squillaci, 2017).

The previously cited study on Calabrian Greek-speaking groups hinted at signs of genetic drift possibly involving these communities (Sarno *et al.* 2017). Here, the number of analyzed variants was improved and new communities from the same geographical area have been genotypes to achieve a deeper level of understanding about the impact of geographic isolation. In an effort to double the number of representatives, six additional communities, some still speaking Greco, and some having lost its use in earlier times, were sampled from the Aspromonte area (Amendolea, Africo, San Lorenzo, Cardeto, Samo, San Luca) in an attempt to temporally describe the restriction of the zone of Greek influence in Southern Calabria. New samples were also collected from a less isolated area enclosing four villages from the province of Catanzaro (Girifalco, Jacurso, Pentone, Tiriolo) as well as the non-isolated entities of Castrovillari (Northern Calabria, Southern Italy) and Benevento (Campania, Southern Italy). This dataset is, then, suitable to examine the demographic contribution to the genetic disparity of contemporary Calabrian Greeks by comparing the genetic architecture of the isolated communities to the general Southern Italian population (**Table 4.1.1**).

4.2 Materials and methods

4.2.1. Population samples

Saliva samples from 149 unrelated Italian volunteers, with local ancestry of at least three generations and either Greco speaking or non-speaking, were gathered with the Oragene-DNA Self Collection Kit OG-500 (DNA Genotek, Ottawa, Ontario, Canada) from 11 isolated communities from the Aspromonte area of Reggio Calabria (Southern Calabria), 4 municipalities in the province of Catanzaro (Central Calabria), and to population samples from the provinces of Cosenza (Northern Calabria) and Benevento (Campania) (**Figure 4.2.1.1**).

Table 4.1.1. Southern Italian communities investigated in the present study, with sample numerosity.

Population	Province	Region	N*
Southern Italy			
Benevento	Benevento	Campania	20
Castrovillari	Cosenza	Calabria	26
Pentone	Cosenza	Calabria	7
Tiriolo	Cosenza	Calabria	11
Jacurso	Cosenza	Calabria	6
Girifalco	Cosenza	Calabria	14
San Luca	Reggio Calabria	Calabria	4
Samo	Reggio Calabria	Calabria	2
Cardeto	Reggio Calabria	Calabria	6
San Lorenzo	Reggio Calabria	Calabria	8
Amendolea	Reggio Calabria	Calabria	4
Africo	Reggio Calabria	Calabria	5
Bova	Reggio Calabria	Calabria	10
Condofuri	Reggio Calabria	Calabria	4
Roccaforte del Greco	Reggio Calabria	Calabria	6
Gallicianò	Reggio Calabria	Calabria	3
Roghudi	Reggio Calabria	Calabria	5

Figure 4.2.1.1. Southern Italian communities in their geographic context. On the right panel, the four villages around Cosenza. On the left panel, the communities of Aspromonte (in yellow, the ones already collected; in orange, the ones collected for this study). From Sarno *et al.* (2021).

4.2.2. Genotyping and quality filtering

Following producer's recommendations, genomic DNA was purified from the collection kits and quantified with the Qubit® dsDNA BR Assay Kit (Life Technologies, Carlsbad, CA, USA). The Center for Biomedical Research & Technologies of the Italian Auxologic Institute (Milan, Italy) genotyped the samples with the HumanOmniExpress BeadChip (Illumina, San Diego, CA, USA), which implements 713,014 SNPs.

Similarly to what already presented in the Methods section of Chapter 2, after the exclusion of variants on sex chromosomes, genotyping results were filtered using the following pipeline implemented with PLINK software version 1.9 (Chang *et al.* 2015): only samples with a genotyping success rate over 92% were retained, as well as markers with missing call rates under 2%, variants with a minor allele frequency (MAF) over 1% and SNPs without significant deviations from the Hardy-Weinberg equilibrium after Bonferroni correction (p-value threshold of 1.41×10^{-8}). IBD sharing was estimated to locate pairs with PiHat value higher than 0.125, in which case one of the two individuals was removed.

The "local dataset" achieved after quality control filtering procedures included 141 samples and 621,755 SNPs. To perform genotype-based analyses, the dataset was further trimmed by eliminating variants in linkage disequilibrium ($r2 > 0.1$) within a sliding window of 50 SNPs moving by 10 SNPs at each time (PLINK option *--indep-pairwise 50 10 0.1*). The obtained "pruned local" dataset comprised 64,147 SNPs.

4.2.3. Datasets production

A panel of publicly available genomes from Europe, Caucasus and the Near East was extracted from the Human Genome Diversity Project (HGDP) (Li *et al.* 2008), subjected to the same QC described above and ambiguous polymorphisms (A/T or C/G) were filtered out to avoid strand-flipping issues during merging. This panel, united with the Southern Italian populations under analysis, was used to frame the variability of the "local" dataset into the Euro-Mediterranean genetic landscape. The "modern extended" dataset obtained after merging includes 238 additional individuals from 10 Euro-Mediterranean populations (**Table 4.2.3.1**) and a set of 337,711 SNPs (which became 59,124 SNPs after pruning, as described in the previous paragraph).

Table 4.2.3.1. Euro-Mediterranean panel used to analyze Southern Italian genetic variability.

Comparison populations	
Sardinia	28
North Italy	13
Tuscan	8
French Basque	24
French	28
Orcadian	14
Russia	25
Adygei	17
Druze	40
Palestinians	41

The "modern extended" dataset was also merged with genomic data for 1,059 ancient samples recovered from literature (Lazaridis *et al.* 2014; Haak *et al.* 2015; Mathieson *et al.* 2015; Jones *et al.* 2015; Hofmanova *et al.* 2016; Lazaridis *et al.* 2016; Fu *et al.* 2016; Lazaridis *et al.* 2017; Lipson *et al.* 2017; Mathieson *et al.* 2018; Olalde *et al.* 2018), to test for ancestral contributions of populations having lived in the European continent before. The obtainined "modern-plus-ancient" dataset comprised 326,832 SNPs. A less stringent LD-pruning procedure was applied to prepare this dataset for genotype-based analyses by using a 200-SNPs window, sliding 25 loci the time and a LD threshold of 0.4 (PLINK option *--indep-pairwise 200 25 0.4*). After pruning, 286,656 variants were left.

4.2.4. Population structure and admixture analyses

The *smartpca* function implemented in the EIGENSOFT package (Patterson *et al.* 2006) was used to carry out Principal Component Analysis (PCA) on both the "local" and "extended" datasets with modern populations only. The *lsqproject* option was then used to separately contextualize the ancient samples into the obtained PCA space.

The ADMIXTURE software (Alexander *et al.* 2009) was applied to estimate ancestry proportions in the modern populations for a number of ancestral contributions K from 2 to 10; ten independent runs for each K were performed and the ones with highest log-likelihood were used in the final plot; the number of hypothesized ancestral contributions with the lowest cross-validation error was chosen as the best fit for the input data.

To formally assess relationships between modern populations and ancient individuals we computed *outgroup-f3* statistics in the form of *f3(YRI; Modern, Ancient)*, by using the *qp3pop* function implemented in ADMIXTOOLS package (Reich *et al.* 2009, Patterson *et al.* 2012). The Treemix v1.12 software (Pickrell and Pritchard, 2012) was run to infer patterns of gene flow among

Euro-Mediterranean populations, using the Berber Mozabite population of Algeria as root for a tree without migration events, and then allowing for 1 up to 6 migration instances.

The f4 method proposed by qpWave/*qpAdmix* (Haak *et al.* 2015), with a P-value threshold of 0.01 to accept the tested model, allowed for a finer characterization of the number of independent ancestral gene pools that contributed to modern populations (by verifying that the ancient groups considered as "source" and those used as "outgroups" were genetically distinguishable in the first place) and for modeling the admixture profile of the modern target groups based on the ancient source samples.

4.2.5. Haplotype-based fine-scale structuring

As described in the Methods section of Chapter 2, the "extended modern dataset" was subjected to haplotype estimation using SHAPEIT (Delaneau *et al.* 2013), then the CHROMOPAINTERv2/ fineSTRUCTURE suite (Lawson *et al.* 2012) was applied to estimate clusters of homogeneous individuals. After mutation and recombination rate estimation on four chromosomes (4, 10, 15, 22) with CHROMOPAINTER, however, the full dataset was re-run leaving all individuals as both "recipients" and "donors" (in Chapter 2, N_ITA and S_ITA clusters were recipients, but not donors). fineSTRUCTURE parameters were also significantly different, as 3,000,000 burn-in MCMC iterations were set, followed by 2,000,000 true iterations of clustering estimation and a sampling frequency of 10,000 runs. Finally, 1,000,000 hill-climbing steps were applied to improve probability estimates and merge spurious clusters.

4.2.6. Genetic isolation and population differentiation

Internal genetic variation was estimated for each Southern Italian population by computing the extension of run-of-homozygosity (ROH) segments (Kirin *et al.* 2012; Pemberton *et al.* 2012), with the criterion that longer and numerous homozygous segments shared across individuals denote a genomic footprint of strong bottleneck and closer genetic relationship due to long-standing isolation. Both *Fin* (inbreeding estimate) and *Fhom* (run-of-homozygosity) indexes were computed by applying the the *--homozyg* and *--het* options of PLINK software (Panoutsoupoulou *et al.* 2014).

Signals of genetic differentiation among Southern Italian groups were finally explored by computing single locus Weir and Cockerham Fst for each variant in the high-density "local" dataset. According to the clusters recognized by fineSTRUCTURE, the top 1% Fst values for variants that distinguish the *Aspromonte* group from both *Catanzaro* and *Benevento+Castrovillari* were collected, then a list of genes spanning these SNPs (for which the Biomart data mining tool was used to retrieve chromosome location, start-end positions and nomenclature of the genes in the Ensembl database

GRCh37 human archive version 100 and the HUGO resource for gene nomenclature at www.genenames.org) was compared to those covered by the Illumina HumanOmniExpress BeadChip and an enrichment analysis was performed on these genes, using the PANTHER Gene Ontology (GO) tool (Thomas *et al.* 2006; Smedley *et al.* 2015; Yates *et al.* 2017; Mi *et al.* 2019; Yates *et al.* 2020). This allowed to identify the prevalent pathways that discriminate the *Aspromonte* cluster.

4.3 Results and discussion

4.3.1 Southern Italy within the Euro-Mediterranean genetic context

A PCA was implemented to set the variability of the newly collected Southern Italian populations into the wider Mediterranean and European groups extracted from the HGDP (**Table 4.2.3.1**). Principal components PC1 and PC2 (**Figure 4.3.1.1a**) summarize geographic clines of genetic variation already emerged for Euro-Mediterranean samples (Lao *et al.* 2008; Novembre *et al.* 2008): PC1 is consistent with a SE-NW variation from the Near East to the Orcadian islands; PC2 highlights the clear separation of Sardinian samples with respect to the other European groups (Chiang *et al.* 2018; Marcus *et al.* 2020). Taking into account the Italian peninsula, the usual North-to-South cline of differentiation (Sazzini *et al.* 2016) is clearly highlighted, with Northern Italy genetically closer to Western Europe and Southern Italy gravitating towards the Near East. At PC4, the isolated populations from Aspromonte tend to separate from the rest of Southern Italy (**Figure 4.3.1.1b**).

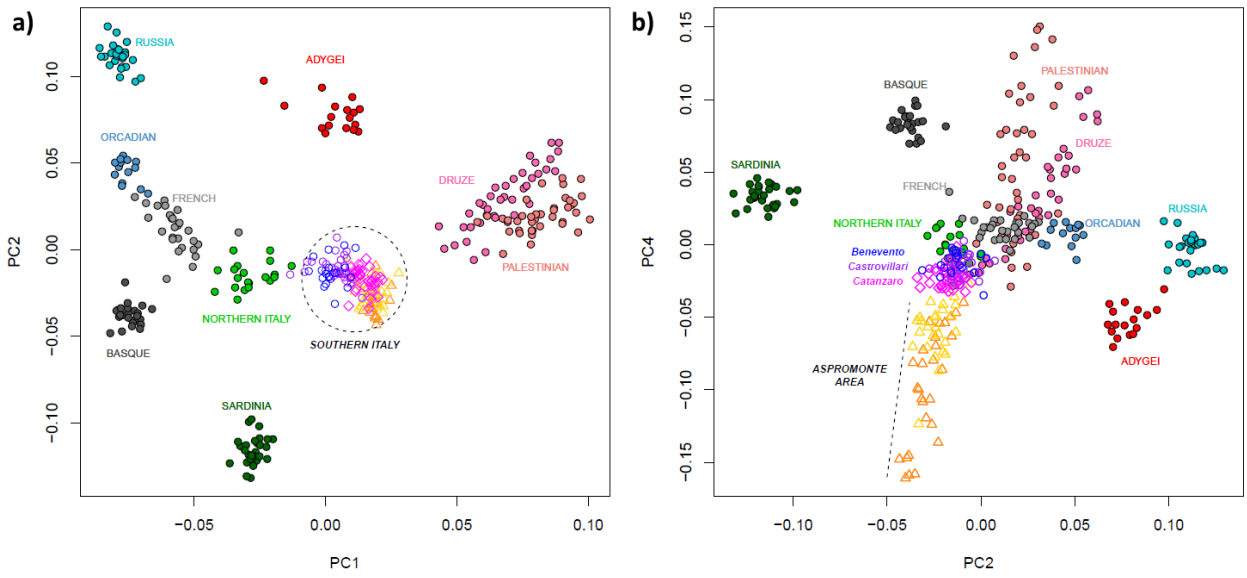


Figure 4.3.1.1. PCA plots for the Euro-Mediterranean dataset. a) Plot of PC1 against PC2, highlights the location of Southern Italy in the Euro-Mediterranean context. b) Plot of PC1 against PC4, highlights the distribution of the Aspromonte samples, with the Greco-speaking communities extending farthest from the rest of Southern Italy. From Sarno *et al.* (2021).

ADMIXTURE analysis identifies the expected genetic ancestries across Euro-Mediterranean (Central European, Levantine, Sardinian Neolithic), but higher values of K pinpoint the emergence of a component peculiar for the Aspromonte individuals (**Figure 4.3.1.2**). Population structure analyses, in line with what presented in Chapter 2, confirm results from previous studies by identifying strong genetic links between Southern Italy and the Caucasus/Middle East (Sarno *et al.* 2017). Regardless of linguistic affiliation, Southern Italian populations showed a common ancestry composed of Sardinian and South-East Mediterranean ADMIXTURE proportions revealed the ancestry of Southern Italian populations, regardless of their linguistic affiliation, to be composed mainly by Sardinian-like and South-Eastern Mediterranean genetic components (**Figure 4.3.1.2**).

Genetic proximity of the Aspromonte groups to the other populations of Southern Italy is confirmed by both PCA and AMIXTURE results, simultaneously showing evident signals of isolation (**Figure 4.3.1.1** and **4.3.1.2**).

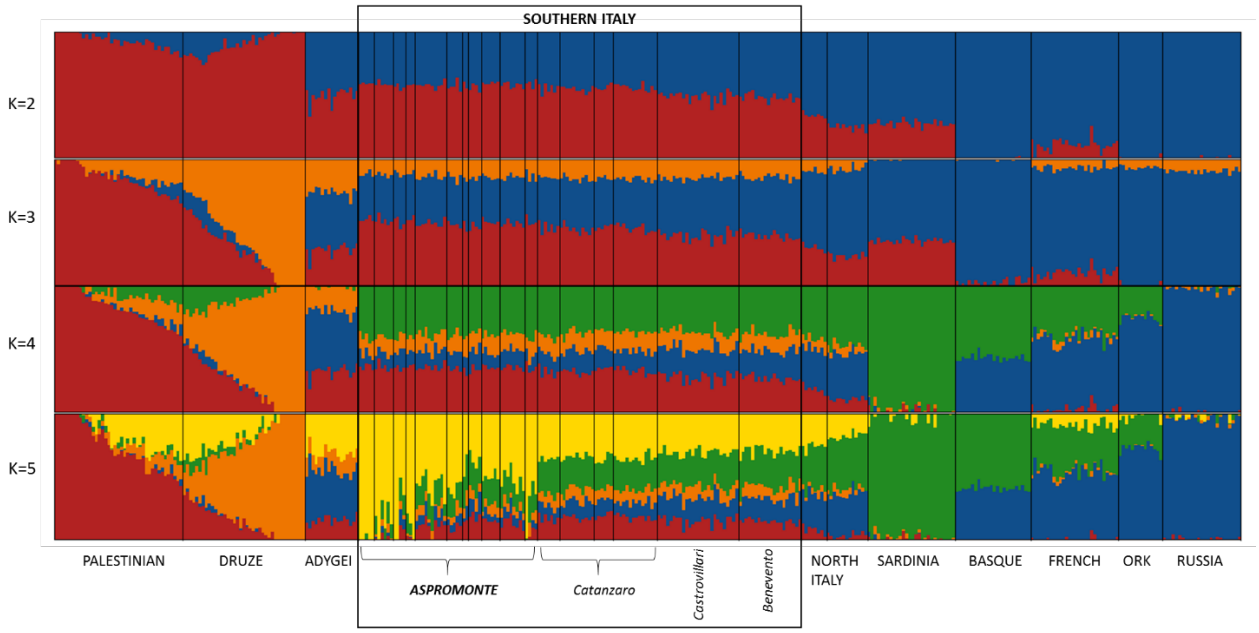


Figure 4.3.1.2 ADMIXTURE plots on the extended modern dataset. A number of putative ancestral components from K=2 to K=5 was tested. The Aspromonte cluster assumes a peculiar combination of ancestral contributions at K=5. From Sarno *et al.* (2021).

The F3 formal test for evaluating admixture into each Southern Italian cluster reveals that the non-isolated populations of Castrovillari, Benevento and Catanzaro may be significantly represented either by a mixture between Caucasus individuals and a Sardinian source or between Central European and Near-East clusters (**Appendix A4**). This is also consistent with the Treemix reconstruction of the Maximum Likelihood tree without migration, which positions the Southern Italian groups in between Caucasian and Near Eastern clusters on one hand, and a continental European cluster with Sardinians in a rooted position, on the other hand. If migration waves are

allowed, edges always indicate a contribution to admixture from a Caucasian/Near-Eastern root to the Southern Italian populations of Benevento, Castrovillari and Catanzaro when they do not split from the other European populations. (**Appendix A4**), supporting the hypothesis of gene flow between Southern Italy and the Middle East. The samples from the Aspromonte mountain never show genetic contributions from other groups (**Appendix A4**), and they always appear to be a basal branch to all other Southern Italian populations in every reconstructed phylogeny. Their branch is also exceedingly long, possibly indicating the presence of ancient isolation and relevant drift effects.

Empirical clustering obtained with the haplotype-based method implemented in CHROMOPAINTER and fineSTRUCTURE (**Figure 4.3.1.3**) corresponds to the geographical distribution of the populations under analysis, as well as corroborating the genetic affiliations that already emerged in genotype-based investigations: the non-isolated Southern Italian clusters, together with the non-Greco speaking communities at the fringe of Aspromonte (Cardeto, Samo and San Luca) form a clade related to the Caucasus group and, the Near Eastern Palestinians and Druze populations. All of the other Aspromonte populations merge into a single cluster which is distinct from that of the other Southern Italian groups (**Figure 4.3.1.3**).

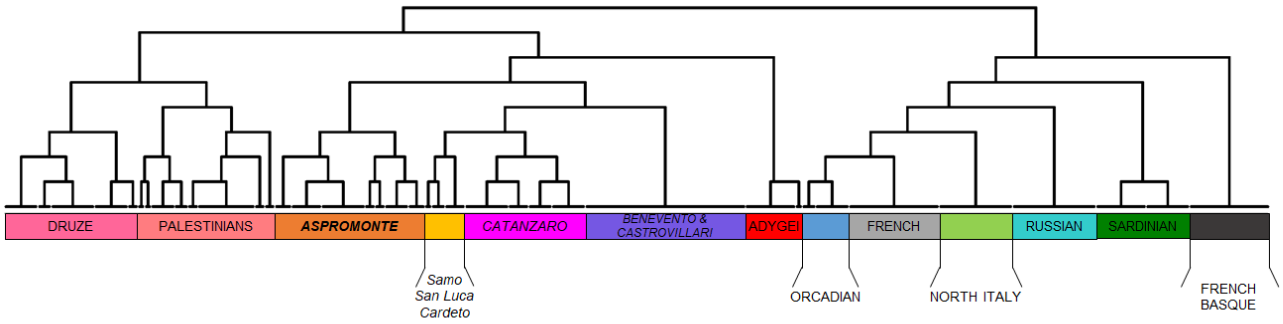


Figure 4.3.1.3 Dendrogram reconstructed by the fineSTRUCTURE algorithm. Each cluster is recognized and color-coded as in Figure 4.3.1.1. The Aspromonte samples fall into a single branch that separates from the rest of the Southern Italian samples, as well as from the municipalities of Cardeto, Samo and San Luca, which are in proximity to the Aspromonte area but never retained Greco as a spoken language. From Sarno *et al.* (2021).

4.3.2. Isolation and genetic differentiation in Southern Italy

PCA limited to the Southern Italian samples highlights the uniqueness previously noted for the Aspromonte groups (**Figure 4.3.2.1a**), with the isolated Greco-speaking communities most differentiated along PC1. ADMIXTURE results for the best value of K=2 (**Figure 4.3.2.1b**) also support this genetic uniqueness, pinpointing an ancestral genetic component which is maximized in almost the Greco-speaking communities (100% Roghudi and Gallicianò, 88% Roccaforte Del Greco, 73% Condofuri) and accounts for at least 30% of ancestry in the other Aspromonte groups. Accordingly, the Greco-speaking communities can be identified as having higher levels of drift, thus replicating patterns of geographic isolation amplified by cultural differences. Therefore, presence of

ancient genetic links between Southern Italy and the populations of Caucasus and Middle East seems to be confirmed by the observed patterns of variation, while the groups from the Aspromonte mountain area, and even more so those that still speak *Greco*, departing from this shared genetic background as a consequence of isolation.

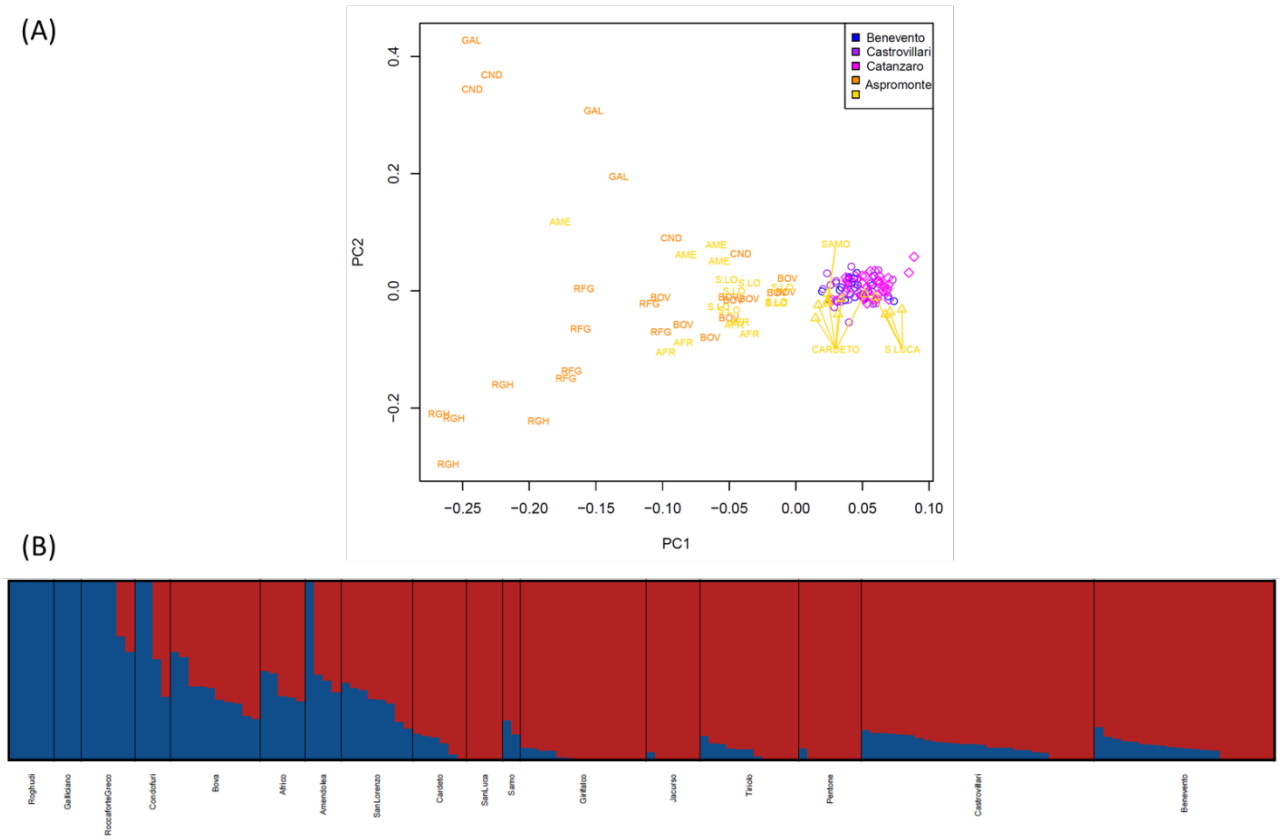


Figure 4.3.2.1 Population structure analyses focused on Southern Italy. A) PCA showing the fan-like scattering of the individuals in Greco-speaking communities, when compared to the open clusters of Benevento, Castrovillari and Catanzaro. B) ADMIXTURE analysis highlights a private ancestral component for the Greco-speaking communities (Roccaforte del Greco, Condofuri, Gallicianò, Roghudi) at K=2. From Sarno *et al.* (2021).

The analysis of genetic variation induced by genetic isolation and drift (through the inbreeding coefficient *Fin*, the genome-wide homozygosity *Fhom* and the number and the extension of genomic runs of homozygosity) reveal that both inbreeding and homozygosity scores are higher for the Aspromonte communities compared to Southern Italy (**Table 4.3.2.1**); moreover, Aspromonte samples reveal a much higher number (NSEG) and length (KB) of runs of homozygosity on average, with respect to other Southern Italian populations (**Figure 4.3.2.2**). These tests indeed confirm that the populations of the Aspromonte area have experienced a higher degree of genetic drift if compared to the other Southern Italian groups.

Table 4.3.2.1. Inbreeding (Fin) and homozygosity (Fhom) scores for the Southern Italian clusters.

	Fin	Fhom	
Benevento	0,00430	0,67333	± 0,00138
Castrovillari	0,00155	0,67243	± 0,00146
Catanzaro	0,00208	0,67261	± 0,00078
Aspromonte Area	*0,01334*	*0,67630*	*± 0,00324*

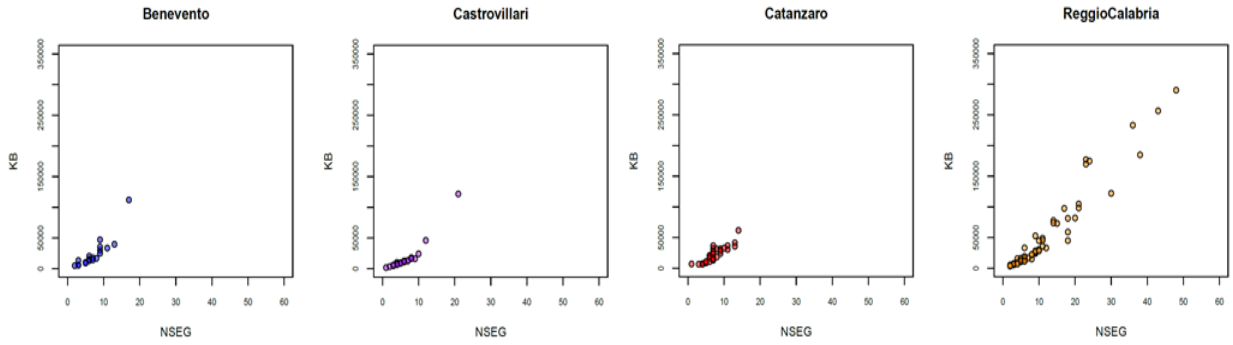


Figure 4.3.2.2. Runs of homozygosity in the Southern Italian clusters. The "ReggioCalabria" group (equivalent to the samples from the Aspromonte area) shows a higher number of longer homozygous genetic segments. From Sarno *et al.* (2021).

Detection of population differentiation in the Aspromonte area and the influence of peculiar variants on biological functions was performed by computing pairwise Fst for the 621,755 SNPs in the "local" dataset, among the three Southern Italian groups defined by fineSTRUCTURE: the Aspromonte cluster (ASPR) as representative of the isolated groups, and the Benevento-Castrovillari (BN+CS) and Catanzaro clusters (CZ) as the non-isolated genomic background of Southern Italy at different geographic distances in the historical area of Greek influence. The top 1% of loci (797 SNPs) in the Fst distribution that distinguished ASPR from both CZ and BN+CS was extracted (**Figure 4.3.2.3**) and an enrichment analysis with Panther on the list of corresponding genes shows a prevalence of genes associated with the Gene Ontology (GO) terms to of "*nervous system development*" for the processes, as well as the cellular components "*neuron part*", "*cell periphery*" and "*plasma membrane*" (**Table 4.3.2.2**). Besides the importance in population history, ethnogenesis and linguistic variation, processes of isolation and genetic drift might have also affected the genetic composition of present-day groups inhabiting these areas of Southern Italy. In fact, the GO analysis showed peculiar biological function of genes related to neurological pathways with higher level of differentiation in the Calabrian area (**Table 4.3.2.2**). Recent studies on hereditary neurodegenerative disorders such as Alzheimer's, Frontotemporal Dementia and Parkinson diseases in Southern Italy were carried out and highlighted that certain areas of the Calabrian region are characterized by low

genetic heterogeneity and high levels of consanguinity due to the geographic isolation over the centuries (Bernardi *et al.* 2012; Anfossi *et al.* 2014; Bernardi *et al.* 2014; Conidi *et al.* 2015; Borrello *et al.* 2016; Cupidi *et al.* 2017; Maletta *et al.* 2018). The observation of same mutations and haplotypes in isolated populations with high rates of consanguinity might be potentially informative for the study of hereditary diseases. Overall, these data more generally remark the importance of population isolates in genetic studies. In fact, due to isolation and drift, coupled with the effects of smaller Ne and higher levels of consanguinity, isolated populations may have modified their genetic architecture through the random amplification or loss of certain genetic variants, thus allowing the study of the role of loci found at higher frequency in these groups. Future studies with phenotypic data could be of extreme value to understand the role of trait-associated variants on health status as recently demonstrated by studies that have linked population genetics and medical genetics (Panoutsopoulou, 2014).

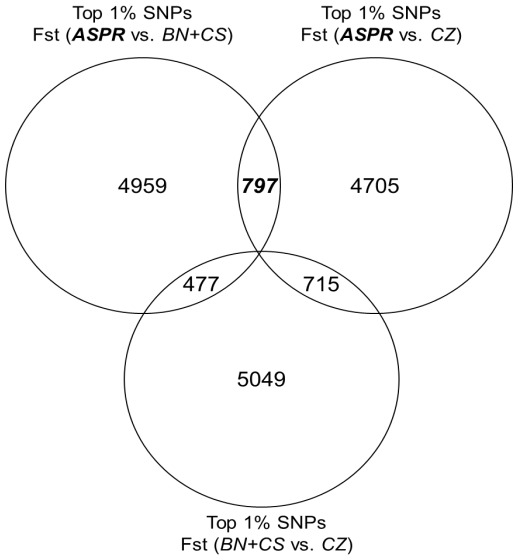


Figure 4.3.2.3 Venn diagram of pairwise Fst comparison. The variants with the top 1% values of Fst for each pairwise comparison between Aspromonte (ASPR), Cosenza (CZ) and Benevento-Castrovillari (BN+CS) are contrasted to find the variants (797) that characterize the Aspromonte cluster. From Sarno *et al.* (2021).

Table 4.3.2.2 Results of the GO enrichment analysis on the genes containing the variants with top 1% Fst values differentiating Aspromonte.

Biological process	Description	# (All)	# (Test)	Exp	Fold Enrichment	+/-	P value	FDR
GO:0007399	nervous system development	2078	55	28,57	1,93	+	1,75E-06	2,73E-02

Cellular component	Description	# (All)	# (Test)	Exp	Fold Enrichment	+/-	P value	FDR
GO:0098590	plasma membrane region	1095	32	15,06	2,13	+	7,51E-05	3,01E-02
GO:0044459	→ plasma membrane part	2522	58	34,68	1,67	+	6,14E-05	3,07E-02
GO:0005886	→ plasma membrane	4614	101	63,44	1,59	+	1,41E-07	1,41E-04
GO:0071944	→ cell periphery	4729	104	65,02	1,6	+	5,16E-08	1,03E-04
GO:0097458	neuron part	1588	45	21,83	2,06	+	4,14E-07	2,76E-04

#(All), number of genes for that GO term; #(Test), number of genes for that GO term in the gene list; Exp, expected number of genes for that GO term in the gene list; FDR, false discovery rate

4.3.3. Ancient genetic legacy of Southern Italian populations

To provide a temporal framework to the patterns of genetic diversity in the analyzed Southern Italian groups, a panel of ancient DNA samples retrieved from literature and covering a period of time from the Mesolithic to the Iron Age (**Appendix A3**) was assembled and merged with the "local" dataset. PCA results (**Figure 4.3.3.1**) are consistent with tendencies previously observed for other Italian datasets (Sarno *et al.* 2017): Southern Italian populations show a closer clustering to Chalcolithic and Bronze Age samples from Anatolia and Greece, as well as with Anatolian and European Neolithic farmers. As presented in Chapter 2, the ancient genetic legacy of Southern Italy points back to migratory events that occurred during the Neolithic and the Bronze Age and involved the eastern coasts of the Mediterranean (Sarno *et al.* 2017; Raveane *et al.* 2019). In particular, differential legacies were observed for Southern Italy during this period, as a non-steppe Caucasian-related ancestry apparently provided an independent genetic contribution with respect to the Yamnaya migration from the Caspian Steppe that occurred through Continental Europe, particularly affecting Northern Italy. Accordingly, genetic analyses performed by comparing modern Eurasian populations with relevant ancestral sources have located the Southern Italian clusters closer to Neolithic and Bronze Age samples from Anatolian and Aegean populations, while affinity of Northern Italy falls within the distribution of continental Europe Late-Neolithic and Bronze-Age samples (**Figure 4.3.3.1**).

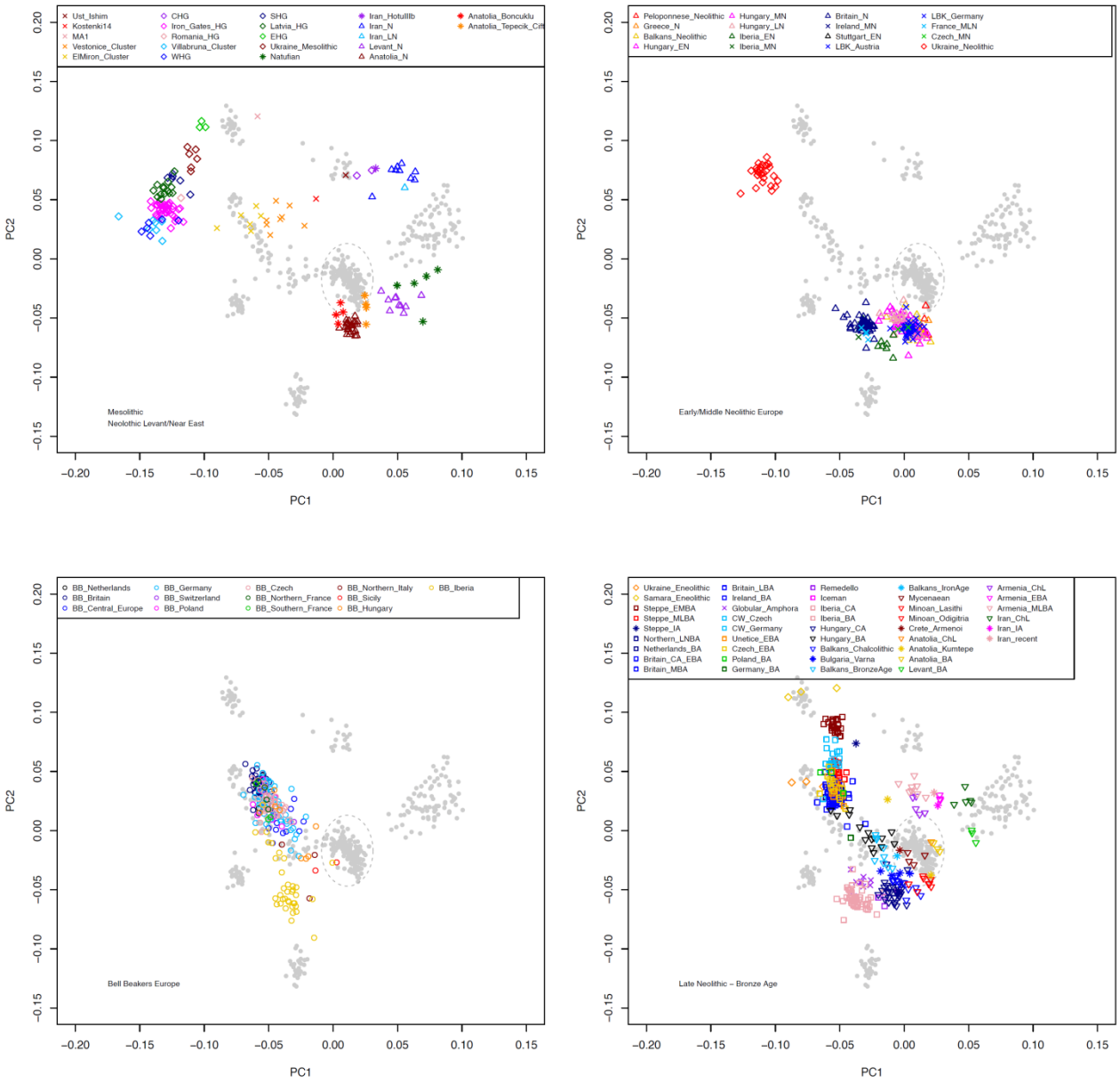


Figure 4.3.3.1 PCA analysis including aDNA samples. In all panels, modern samples are reported in gray dots and Southern Italy is highlighted by a dashed grey circle. Top left panel: Mesolithic/Paleolithic samples, plus Neolithic Levant. Top right panel: Early/middle Neolithic European samples. Bottom left panel: European samples from the Bell Beaker culture. Bottom right panel: Late Neolithic/Bronze Age samples. From Sarno *et al.* (2021).

The admixture proportions computed for the Italian populations with *qpAdm* suggest that all Italian clusters were characterized by a high proportion of Anatolian Neolithic ancestry, with Sardinians being the major recipient (**Figure 4.3.3.2**). In particular, while the Yamnaya source is greater in the Northern Italian cluster (~27%), the Iran_N/CHG ancestry is characteristic of Southern Italy (~24% Benevento, ~25% Castrovillari, ~27% Catanzaro), with a peak value in the Aspromonte mountain groups (~29%). *qpAdmixture* confirms the influence of pastoralists from the Pontic Steppe in the lineage of Northern Italian groups, paralleled in Southern Italy by an analogous Caucasian-related Neolithic source. The observation that both the non-isolated populations of Southern Italy and

the linguistically isolated communities of Aspromonte share the same ancestral sources indicates common genetic contributions, possibly predating the first estimates for the adoption of the Greco language in Southern Italy (**Figure 4.3.3.1**). In fact, a genetic continuity among Southern Italian groups and Eastern Mediterranean populations dating back to the Neolithic could have been easily maintained and even reinforced by subsequent overlapping migration events and gene flow along the same routes, including the more recent interactions with Greek-speaking populations and Byzantines. If these populations were characterized by frequent mobility across the Mediterranean to the coasts of Southern Italy, one could hypothesize a cultural, genetic and linguistic continuum sustained by repeated contacts over time; the presence of Greco-speaking minorities in contemporary Calabria may consequently be interpreted as the remnants of the area of Greek influence in Southern Italy, where the language and culture was preserved via geographic isolation in the most impervious areas of Aspromonte (Violi, 2006; Squillaci, 2017).

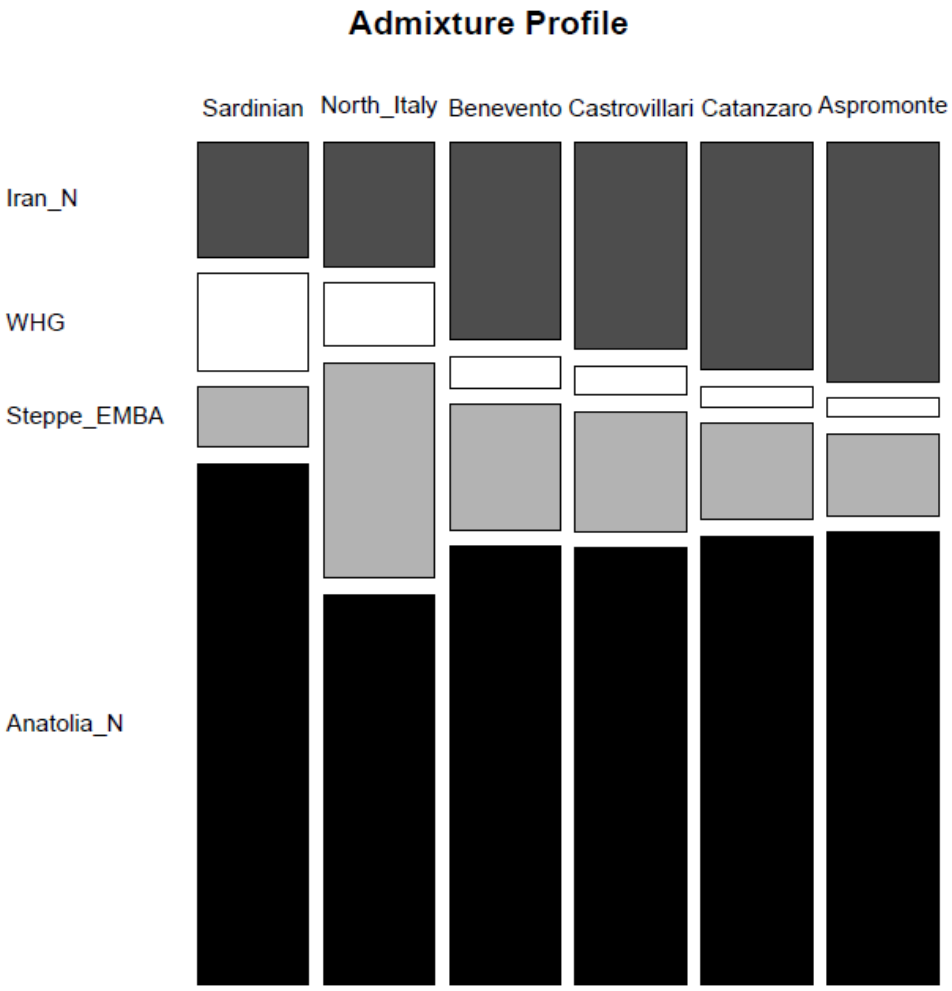


Figure 4.3.3.2 Admixture proportions computed for the Italian populations with *qpAdm.* From Sarno *et al.* (2021).

Chapter 5: Conclusions

The research presented in Chapter 2 revealed how the Italian peninsula presents patterns of population structure that uniquely recapitulate the clinal distribution of human genetic variation across Europe at a micro-geographic scale. Previous publications elucidating the demographic structure of the ancestors of contemporary Italians relied either on uniparental markers or common autosomal variants, which limit the amplitude of the analysis; moreover, attempts at disentangling the effect of differential selective pressures along the peninsula did not include models of polygenic adaptation that would take into account small but diffuse genetic contributions from loci with limited phenotypic effects, insisting on the same biological function. A panel of high-coverage whole-genome sequences provided the opportunity to dissect the demographic and adaptive history of the ancestors of modern Italians. In particular, new evidence for an early differentiation of population clusters at the extremities of the cline of genetic variation after the Late Glacial Maximum was provided, as well as successive differential demographic processes and migrations that exacerbated the genetic distinction of gene pools across the peninsula. Cold, energetically demanding environmental conditions and a high-fat diet are suggested as factors having driven adaptation at insulin-related genes in the ancestors of modern Northern Italian individuals, making the contemporary population less susceptible to T2D and obesity in the context of the modern highly caloric lifestyle. Southern Italian people possibly benefitted from an adaptation towards pathogen exposure and extended UV radiation, that in the modern context translates to reduced susceptibility to IgAN and melanomas. Several genes and biological processes under selection across the peninsula, but more pronounced in Southern Italy, were finally revealed to potentially modulate the achievement of longevity as a healthy phenotype. Therefore, this research was also able to disclose relevant biomedical implications underlying the adaptive processes that characterized the Italian peninsula and highlights the appropriateness of applying an evolutionary approach to uncover the causes of health and disease in the human population, especially by pinpointing the genetic determinants supporting biological adaptation.

The investigation reported in Chapter 3 focused on the pivotal role of APOE in the genetics of human longevity, both by reviewing relevant literature on the subject and adding novel demo-geographic considerations based on modern data form Italian centenarian individuals, as well as publicly available ancient Euro-Mediterranean genomes. It demonstrated that the analysis of the worldwide variability is paramount to understand how the patterns of differentiation for this complex phenotype strongly depends on the pleiotropic actions of the APOE isoforms, as well as the peculiar environmental conditions that each population is subjected to. It is argued that an evolutionary

perspective is always crucial to disentangle the conservation and worldwide distribution of variants profoundly affecting complex traits that have a high impact on health and disease. In this context, an evolutionary approach, such as the one introduced by evolutionary medicine (Nesse, 2008; Nesse, 2010; Nesse *et al.* 2011; Wells *et al.* 2017), may be useful to interpret distant connections for pleiotropic genes, such as APOE, in different populations.

The unpublished work introduced in Chapter 4 tackles the long-lasting debate around the linguistic origins of the Greek-speaking communities of the Aspromonte area of Calabria, from a genetic and demographic viewpoint. Population demography analyses reveal private ancestral components for the Aspromonte communities, closer affinity with modern populations from the Eastern Mediterranean and higher genetic contributions from ancient Neolithic clusters of the Caucasus and the Near East. The results support a reconciled scenario, whereby the existence of long standing contacts with populations of the Eastern Mediterranean, who provided a substantial portion of the ancestral genetic background observable today in Southern Italy, and multiple migrations from subsequent historical periods (*Magna Graecia* and Byzantine contacts in particular) coexist in a stratified multitude of contributions that maintained and reinforced the genetic uniqueness of the communities of the Aspromonte mountain. An investigation of the most differentiated variants that characterize the Aspromonte populations reveals biological functions related to neurological pathways, in agreement with recent studies on Alzheimer's and Parkinson's disease in the Calabrian area that show a higher incidence of rare familiar forms of these diseases in close-knit and isolated communities. In this case, through an evolutionary approach this research was able to uncover relevant biomedical indicators underlying differentiation processes characteristic of linguistically, geographically and culturally isolated genetic enclaves.

Overall, this thesis provides novel significant contributions to the understanding of population demography, ancestral influences, divergence and adaptive evolutionary events that have occurred globally and locally along the Italian peninsula, by describing them in the context of the complex migration patterns that involved the European continent and, in particular, the Mediterranean basin across prehistorical and historical times. It also offers an analysis of local patterns of differentiation by cultural, linguistic and geographic isolation, highlighting the influence of continuous commercial contacts with Euro-Mediterranean groups in recent times on the genomic background of the Italian population. Finally, the application of evolutionary and population genomic methodologies allowed to disclose the biomedical repercussions of ancient advantageous adaptive events on the modern inhabitants of the peninsula, with a particular interest in the multifaceted interactions that define complex phenotypic traits, such as neurological diseases and the achievement of extreme longevity.

Data availability statement

Data presented in this thesis are included in this manuscript and publicly available repositories. In particular, data for the Italian population have been deposited at: https://figshare.com/articles/dataset/Italian_dataset_Sazzini_et_al_2020_/11993202.
Sequencing data for modern European and Mediterranean populations for the Simons Genome Diversity Project are publicly available at the EBI European Nucleotide Archive: https://www.ebi.ac.uk/ena/browser/view/PRJEB9586.
Sequencing data for modern European and Mediterranean populations for the 1000 Genomes Project are publicly available at: http://ftp.1000genomes.ebi.ac.uk
Sequencing data for ancient samples can be recovered at: https://www.ebi.ac.uk/ena/browser/view/PRJEB11450, https://www.ebi.ac.uk/ena/browser/view/PRJEB13123 and https://www.ebi.ac.uk/ena/browser/view/PRJEB14455.

References

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. Nature. 2015 Oct 1;526(7571):68-74. doi: 10.1038/nature15393. PMID: 26432245; PMCID: PMC4750478.

Acharya P, Segall ML, Zaiou M, Morrow J, Weisgraber KH, Phillips MC, Lund-Katz S, Snow J. Comparison of the stabilities and unfolding pathways of human apolipoprotein E isoforms by differential scanning calorimetry and circular dichroism. Biochim Biophys Acta. 2002 Sep 5;1584(1):9-19. doi: 10.1016/s1388-1981(02)00263-9. PMID: 12213488.

Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009 Sep;19(9):1655-64. doi: 10.1101/gr.094052.109. Epub 2009 Jul 31. PMID: 19648217; PMCID: PMC2752134.

Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, Malaspinas AS, Margaryan A, Higham T, Chivall D, Lynnerup N, Harvig L, Baron J, Della Casa P, Dąbrowski P, Duffy PR, Ebel AV, Epimakhov A, Frei K, Furmanek M, Gralak T, Gromov A, Gronkiewicz S, Grupe G, Hajdu T, Jarysz R, Khartanovich V, Khokhlov A, Kiss V, Kolář J, Kriiska A, Lasak I, Longhi C, McGlynn G, Merkevicius A, Merkyte I, Metspalu M, Mkrtchyan R, Moiseyev V, Paja L, Pálfi G, Pokutta D, Pospieszny Ł, Price TD, Saag L, Sablin M, Shishlina N, Smrčka V, Soenov VI, Szeverényi V, Tóth G, Trifanova SV, Varul L, Vicze M, Yepiskoposyan L, Zhitenev V, Orlando L, Sicheritz-Pontén T, Brunak S, Nielsen R, Kristiansen K, Willerslev E. Population genomics of Bronze Age Eurasia. Nature. 2015 Jun 11;522(7555):167-72. doi: 10.1038/nature14507. PMID: 26062507.

Alvergne A., Jenkinson C., Faurie C., editors. Evolutionary Thinking in Medicine: From Research to Policy and Practice. Springer International Publishing; Basel, Switzerland: 2016.

Amorim A, Fernandes T, Taveira N. Mitochondrial DNA in human identification: a review. PeerJ. 2019 Aug 13;7:e7314. doi: 10.7717/peerj.7314. PMID: 31428537; PMCID: PMC6697116.

Anagnostou P, Dominici V, Battaggia C, Lisi A, Sarno S, Boattini A, Calò C, Francalacci P, Vona G, Tofanelli S, Vilar MG, Colonna V, Pagani L, Destro Bisol G. Inter-individual genomic heterogeneity within European population isolates. PLoS One. 2019 Oct 9;14(10):e0214564. doi: 10.1371/journal.pone.0214564. PMID: 31596857; PMCID: PMC6785074.

Anagnostou P, Dominici V, Battaggia C, Pagani L, Vilar M, Wells RS, Pettener D, Sarno S, Boattini A, Francalacci P, Colonna V, Vona G, Calò C, Destro Bisol G, Tofanelli S. Overcoming the dichotomy between open and isolated populations using genomic data from a large European dataset. Sci Rep. 2017 Feb 1;7:41614. doi: 10.1038/srep41614. PMID: 28145502; PMCID: PMC5286425.

Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc. 2010 Sep;5(9):1564-73. doi: 10.1038/nprot.2010.116. Epub 2010 Aug 26. PMID: 21085122; PMCID: PMC3025522.

Anfossi M, Colao R, Gallo M, Bernardi L, Conidi ME, Frangipane F, Vasso F, Puccio G, Clodomiro A, Mirabelli M, Curcio SA, Torchia G, Smirne N, Di Lorenzo R, Maletta R, Bruni AC. Identification of three novel LRRK2 mutations associated with Parkinson's disease in a Calabrian population. J Alzheimers Dis. 2014;38(2):351-7. doi: 10.3233/JAD-130689. PMID: 23963289.

Ang LS, Cruz RP, Hendel A, Granville DJ. Apolipoprotein E, an important player in longevity and age-related diseases. Exp Gerontol. 2008 Jul;43(7):615-22. doi: 10.1016/j.exger.2008.03.010. Epub 2008 Apr 4. PMID: 18468826.

Anselmi CV, Malovini A, Roncarati R, Novelli V, Villa F, Condorelli G, Bellazzi R, Puca AA. Association of the FOXO3A locus with extreme longevity in a southern Italian centenarian study. Rejuvenation Res. 2009 Apr;12(2):95-104. doi: 10.1089/rej.2008.0827. PMID: 19415983.

Antonio ML, Gao Z, Moots HM, Lucci M, Candilio F, Sawyer S, Oberreiter V, Calderon D, Devitofranceschi K, Aikens RC, Aneli S, Bartoli F, Bedini A, Cheronet O, Cotter DJ, Fernandes DM, Gasperetti G, Grifoni R, Guidi A, La

Pastina F, Loreti E, Manacorda D, Matullo G, Morretta S, Nava A, Fiocchi Nicolai V, Nomi F, Pavolini C, Pentiricci M, Pergola P, Piranomonte M, Schmidt R, Spinola G, Sperduti A, Rubini M, Bondioli L, Coppa A, Pinhasi R, Pritchard JK. Ancient Rome: A genetic crossroads of Europe and the Mediterranean. Science. 2019 Nov 8;366(6466):708-714. doi: 10.1126/science.aay6826. PMID: 31699931; PMCID: PMC7093155.

Arauna LR, Hellenthal G, Comas D. Dissecting human North African gene-flow into its western coastal surroundings. Proc Biol Sci. 2019 May 15;286(1902):20190471. doi: 10.1098/rspb.2019.0471. PMID: 31039721; PMCID: PMC6532504.

Argyri L, Dafnis I, Theodossiou TA, Gantz D, Stratikos E, Chroni A. Molecular basis for increased risk for late-ons*et al.*zheimer disease due to the naturally occurring L28P mutation in apolipoprotein E4. J Biol Chem. 2014 May 2;289(18):12931-45. doi: 10.1074/jbc.M113.538124. Epub 2014 Mar 18. PMID: 24644280; PMCID: PMC4007480.

Ayub Q, Mezzavilla M, Pagani L, Haber M, Mohyuddin A, Khaliq S, Mehdi SQ, Tyler-Smith C. The Kalash genetic isolate: ancient divergence, drift, and selection. Am J Hum Genet. 2015 May 7;96(5):775-83. doi: 10.1016/j.ajhg.2015.03.012. Epub 2015 Apr 30. PMID: 25937445; PMCID: PMC4570283.

Bankwitz D, Doepke M, Hueging K, Weller R, Bruening J, Behrendt P, Lee JY, Vondran FWR, Manns MP, Bartenschlager R, Pietschmann T. Maturation of secreted HCV particles by incorporation of secreted ApoE protects from antibodies by enhancing infectivity. J Hepatol. 2017 Sep;67(3):480-489. doi: 10.1016/j.jhep.2017.04.010. Epub 2017 Apr 22. PMID: 28438690.

Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E, Pickrell JK, Bouchier C, Tichit M, Neyrolles O, Gicquel B, Kidd JR, Kidd KK, Alcaïs A, Ragimbeau J, Pellegrini S, Abel L, Casanova JL, Quintana-Murci L. Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. PLoS Genet. 2009 Jul;5(7):e1000562. doi: 10.1371/journal.pgen.1000562. Epub 2009 Jul 17. PMID: 19609346; PMCID: PMC2702086.

Benazzi S, Douka K, Fornai C, Bauer CC, Kullmer O, Svoboda J, Pap I, Mallegni F, Bayle P, Coquerelle M, Condemi S, Ronchitelli A, Harvati K, Weber GW. Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. Nature. 2011 Nov 2;479(7374):525-8. doi: 10.1038/nature10617. PMID: 22048311.

Bernardi L, Cupidi C, Frangipane F, Anfossi M, Gallo M, Conidi ME, Vasso F, Colao R, Puccio G, Curcio SAM, Mirabelli M, Clodomiro A, Di Lorenzo R, Smirne N, Maletta R, Bruni AC. Novel N-terminal domain mutation in prion protein detected in 2 patients diagnosed with frontotemporal lobar degeneration syndrome. Neurobiol Aging. 2014 Nov;35(11):2657.e7-2657.e11. doi: 10.1016/j.neurobiolaging.2014.06.006. Epub 2014 Jun 14. PMID: 25022973.

Bernardi L, Frangipane F, Smirne N, Colao R, Puccio G, Curcio SA, Mirabelli M, Maletta R, Anfossi M, Gallo M, Geracitano S, Conidi ME, Di Lorenzo R, Clodomiro A, Cupidi C, Marzano S, Comito F, Valenti V, Zirilli MA, Ghani M, Xi Z, Sato C, Moreno D, Borelli A, Leone RA, St George-Hyslop P, Rogaeva E, Bruni AC. Epidemiology and genetics of frontotemporal dementia: a door-to-door survey in southern Italy. Neurobiol Aging. 2012 Dec;33(12):2948.e1-2948.e10. doi: 10.1016/j.neurobiolaging.2012.06.017. Epub 2012 Jul 20. PMID: 22819134; PMCID: PMC5019339.

Bevivino A, Dalmastri C, Tabacchioni S, Chiarini L, Belli ML, Piana S, Materazzo A, Vandamme P, Manno G. Burkholderia cepacia complex bacteria from clinical and environmental sources in Italy: genomovar status and distribution of traits related to virulence and transmissibility. J Clin Microbiol. 2002 Mar;40(3):846-51. doi: 10.1128/jcm.40.3.846-851.2002. PMID: 11880403; PMCID: PMC120285.

Blessmann J, Ali IK, Nu PA, Dinh BT, Viet TQ, Van AL, Clark CG, Tannich E. Longitudinal study of intestinal Entamoeba histolytica infections in asymptomatic adult carriers. J Clin Microbiol. 2003 Oct;41(10):4745-50. doi: 10.1128/jcm.41.10.4745-4750.2003. PMID: 14532214; PMCID: PMC294961.

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. Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata. PLoS One. 2013 May 29;8(5):e65441. doi: 10.1371/journal.pone.0065441. PMID: 23734255; PMCID: PMC3666984.

Boattini A, Sarno S, Pedrini P, Medoro C, Carta M, Tucci S, Ferri G, Alù M, Luiselli D, Pettener D. Traces of medieval migrations in a socially stratified population from Northern Italy. Evidence from uniparental markers and deep-rooted pedigrees. Heredity (Edinb). 2015 Feb;114(2):155-62. doi: 10.1038/hdy.2014.77. Epub 2014 Sep 10. PMID: 25204305; PMCID: PMC4815625.

Bonacina F, Coe D, Wang G, Longhi MP, Baragetti A, Moregola A, Garlaschelli K, Uboldi P, Pellegatta F, Grigore L, Da Dalt L, Annoni A, Gregori S, Xiao Q, Caruso D, Mitro N, Catapano AL, Marelli-Berg FM, Norata GD. Myeloid apolipoprotein E controls dendritic cell antigen presentation and T cell activation. Nat Commun. 2018 Aug 6;9(1):3083. doi: 10.1038/s41467-018-05322-1. PMID: 30082772; PMCID: PMC6079066.

Borrello L, Cupidi C, Laganà V, Anfossi M, Conidi ME, Smirne N, Tavernitì M, Guarasci R, Bruni AC. Angela R.: a familial Alzheimer's disease case in the days of Auguste D. J Neurol. 2016 Dec;263(12):2494-2498. doi: 10.1007/s00415-016-8294-x. Epub 2016 Oct 11. PMID: 27730373.

Bour A, Grootendorst J, Vogel E, Kelche C, Dodart JC, Bales K, Moreau PH, Sullivan PM, Mathis C. Middle-aged human apoE4 targeted-replacement mice show retention deficits on a wide range of spatial memory tasks. Behav Brain Res. 2008 Nov 21;193(2):174-82. doi: 10.1016/j.bbr.2008.05.008. Epub 2008 May 18. PMID: 18572260.

Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017 Jun 15;169(7):1177-1186. doi: 10.1016/j.cell.2017.05.038. PMID: 28622505; PMCID: PMC5536862.

Bramble DM, Lieberman DE. Endurance running and the evolution of Homo. Nature. 2004 Nov 18;432(7015):345-52. doi: 10.1038/nature03052. PMID: 15549097.

Bräuer G. The morphological differentiation of anatomically modern man in Africa, with special regard to recent finds from East Africa. Z Morphol Anthropol. 1978 Dec;69(3):266-92. PMID: 113948.

Brisighelli F, Álvarez-Iglesias V, Fondevila M, Blanco-Verea A, Carracedo A, Pascali VL, Capelli C, Salas A. Uniparental markers of contemporary Italian population reveals details on its pre-Roman heritage. PLoS One. 2012;7(12):e50794. doi: 10.1371/journal.pone.0050794. Epub 2012 Dec 10. Erratum in: PLoS One. 2014 Jan 17;9(1): PMID: 23251386; PMCID: PMC3519480.

Broer L, Buchman AS, Deelen J, Evans DS, Faul JD, Lunetta KL, Sebastiani P, Smith JA, Smith AV, Tanaka T, Yu L, Arnold AM, Aspelund T, Benjamin EJ, De Jager PL, Eirkisdottir G, Evans DA, Garcia ME, Hofman A, Kaplan RC, Kardia SL, Kiel DP, Oostra BA, Orwoll ES, Parimi N, Psaty BM, Rivadeneira F, Rotter JI, Seshadri S, Singleton A, Tiemeier H, Uitterlinden AG, Zhao W, Bandinelli S, Bennett DA, Ferrucci L, Gudnason V, Harris TB, Karasik D, Launer LJ, Perls TT, Slagboom PE, Tranah GJ, Weir DR, Newman AB, van Duijn CM, Murabito JM. GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. J Gerontol A Biol Sci Med Sci. 2015 Jan;70(1):110-8. doi: 10.1093/gerona/glu166. Epub 2014 Sep 8. PMID: 25199915; PMCID: PMC4296168.

Buckley MT, Racimo F, Allentoft ME, Jensen MK, Jonsson A, Huang H, Hormozdiari F, Sikora M, Marnetto D, Eskin E, Jørgensen ME, Grarup N, Pedersen O, Hansen T, Kraft P, Willerslev E, Nielsen R. Selection in Europeans on Fatty Acid Desaturases Associated with Dietary Changes. Mol Biol Evol. 2017 Jun 1;34(6):1307-1318. doi: 10.1093/molbev/msx103. PMID: 28333262; PMCID: PMC5435082.

Büntgen U, Tegel W, Nicolussi K, McCormick M, Frank D, Trouet V, Kaplan JO, Herzig F, Heussner KU, Wanner H, Luterbacher J, Esper J. 2500 years of European climate variability and human susceptibility. Science. 2011 Feb 4;331(6017):578-82. doi: 10.1126/science.1197175. Epub 2011 Jan 13. PMID: 21233349.

Busby GB, Brisighelli F, Sánchez-Diz P, Ramos-Luis E, Martinez-Cadenas C, Thomas MG, Bradley DG, 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 JF, Capelli C. The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. Proc Biol Sci. 2012 Mar 7;279(1730):884-92. doi: 10.1098/rspb.2011.1044. Epub 2011 Aug 24. PMID: 21865258; PMCID: PMC3259916.

Capelli C, Brisighelli F, Scarnicci F, Arredi B, Caglia' A, Vetrugno G, Tofanelli S, Onofri V, Tagliabracci A, Paoli G, Pascali VL. Y chromosome genetic variation in the Italian peninsula is clinal and supports an admixture model for the Mesolithic-Neolithic encounter. Mol Phylogenet Evol. 2007 Jul;44(1):228-39. doi: 10.1016/j.ympev.2006.11.030. Epub 2006 Dec 13. PMID: 17275346.

Capocasa M, Anagnostou P, Bachis V, Battaggia C, Bertoncini S, Biondi G, Boattini A, Boschi I, Brisighelli F, Caló CM, Carta M, Coia V, Corrias L, Crivellaro F, De Fanti S, Dominici V, Ferri G, Francalacci P, Franceschi ZA, Luiselli D, Morelli L, Paoli G, Rickards O, Robledo R, Sanna D, Sanna E, Sarno S, Sineo L, Taglioli L, Tagarelli G, Tofanelli S, Vona G, Pettener D, Destro Bisol G. Linguistic, geographic and genetic isolation: a collaborative study of Italian populations. J Anthropol Sci. 2014;92:201-31. doi: 10.4436/JASS.92001. Epub 2014 Mar 3. PMID: 24607994.

Cashdan E. Biogeography of human infectious diseases: a global historical analysis. PLoS One. 2014 Oct 1;9(10):e106752. doi: 10.1371/journal.pone.0106752. PMID: 25271730; PMCID: PMC4182673.

Caspari R, Lee SH. Older age becomes common late in human evolution. Proc Natl Acad Sci U S A. 2004 Jul 27;101(30):10895-900. doi: 10.1073/pnas.0402857101. Epub 2004 Jul 13. PMID: 15252198; PMCID: PMC503716.

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The History and Geography of Human Genes. Princeton University Press, Princeton, New Jersey.

Cavalli-Sforza LL, Piazza A. Human genomic diversity in Europe: a summary of recent research and prospects for the future. Eur J Hum Genet. 1993;1(1):3-18. doi: 10.1159/000472383. PMID: 7520820.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015 Feb 25;4:7. doi: 10.1186/s13742-015-0047-8. PMID: 25722852; PMCID: PMC4342193.

Chatzidakis, G. N. *Einleitung in die neugriechischen Grammatik*. (Breikopf & Hartl, 1892).

Chetty PS, Mayne L, Lund-Katz S, Englander SW, Phillips MC. Helical structure, stability, and dynamics in human apolipoprotein E3 and E4 by hydrogen exchange and mass spectrometry. Proc Natl Acad Sci U S A. 2017 Jan 31;114(5):968-973. doi: 10.1073/pnas.1617523114. Epub 2017 Jan 17. PMID: 28096372; PMCID: PMC5293021.

Chiang CWK, Marcus JH, Sidore C, Biddanda A, Al-Asadi H, Zoledziewska M, Pitzalis M, Busonero F, Maschio A, Pistis G, Steri M, Angius A, Lohmueller KE, Abecasis GR, Schlessinger D, Cucca F, Novembre J. Genomic history of the Sardinian population. Nat Genet. 2018 Oct;50(10):1426-1434. doi: 10.1038/s41588-018-0215-8. Epub 2018 Sep 17. PMID: 30224645; PMCID: PMC6168346.

Chou CY, Lin YL, Huang YC, Sheu SY, Lin TH, Tsay HJ, Chang GG, Shiao MS. Structural variation in human apolipoprotein E3 and E4: secondary structure, tertiary structure, and size distribution. Biophys J. 2005 Jan;88(1):455-66. doi: 10.1529/biophysj.104.046813. Epub 2004 Oct 8. PMID: 15475580; PMCID: PMC1305022.

Cilli E, Sarno S, Gnecchi Ruscone GA, Serventi P, De Fanti S, Delaini P, Ognibene P, Basello GP, Ravegnini G, Angelini S, Ferri G, Gentilini D, Di Blasio AM, Pelotti S, Pettener D, Sazzini M, Panaino A, Luiselli D, Gruppioni G. The genetic legacy of the Yaghnobis: A witness of an ancient Eurasian ancestry in the historically reshuffled central Asian gene pool. Am J Phys Anthropol. 2019 Apr;168(4):717-728. doi: 10.1002/ajpa.23789. Epub 2019 Jan 29. PMID: 30693949.

Coia V, Capocasa M, Anagnostou P, Pascali V, Scarnicci F, Boschi I, Battaggia C, Crivellaro F, Ferri G, Alù M, Brisighelli F, Busby GB, Capelli C, Maixner F, Cipollini G, Viazzo PP, Zink A, Destro Bisol G. Demographic histories, isolation and social factors as determinants of the genetic structure of Alpine linguistic groups. PLoS One. 2013 Dec 2;8(12):e81704. doi: 10.1371/journal.pone.0081704. PMID: 24312576; PMCID: PMC3847036.

Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. Science. 2003 Apr 11;300(5617):286-90. doi: 10.1126/science.1084564. PMID: 12690187.

Comparetti, D. *Saggi sui dialetti greci dell'Italia meridionale*. (Fratelli Nistri, 1866).

Conidi ME, Bernardi L, Puccio G, Smirne N, Muraca MG, Curcio SA, Colao R, Piscopo P, Gallo M, Anfossi M, Frangipane F, Clodomiro A, Mirabelli M, Vasso F, Cupidi C, Torchia G, Di Lorenzo R, Mandich P, Confaloni A, Maletta RG, Bruni AC. Homozygous carriers of APP A713T mutation in an autosomal dominant Alzheimer

disease family. Neurology. 2015 Jun 2;84(22):2266-73. doi: 10.1212/WNL.0000000000001648. Epub 2015 May 6. PMID: 25948718; PMCID: PMC4456653.

Cooper DM. Regulation and organization of adenylyl cyclases and cAMP. Biochem J. 2003 Nov 1;375(Pt 3):517-29. doi: 10.1042/BJ20031061. PMID: 12940771; PMCID: PMC1223734.

Corbo RM, Scacchi R. Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a 'thrifty' allele? Ann Hum Genet. 1999 Jul;63(Pt 4):301-10. doi: 10.1046/j.1469-1809.1999.6340301.x. PMID: 10738542.

Cupidi, C., Laganà, V., Smirne, N., Bruni, C. A. The role of historical medical archives in the genealogical rebuilding of large families affected by neurodegenerative diseases. *J. Neurol. Neuromedicine* **2,** 1-3 (2017).

D'Mello SA, Finlay GJ, Baguley BC, Askarian-Amiri ME. Signaling Pathways in Melanogenesis. Int J Mol Sci. 2016 Jul 15;17(7):1144. doi: 10.3390/ijms17071144. PMID: 27428965; PMCID: PMC4964517.

Dafnis I, Argyri L, Sagnou M, Tzinia A, Tsilibary EC, Stratikos E, Chroni A. The ability of apolipoprotein E fragments to promote intraneuronal accumulation of amyloid beta peptide 42 is both isoform and size-specific. Sci Rep. 2016 Aug 1;6:30654. doi: 10.1038/srep30654. PMID: 27476701; PMCID: PMC4967930.

De Fanti S, Barbieri C, Sarno S, Sevini F, Vianello D, Tamm E, Metspalu E, van Oven M, Hübner A, Sazzini M, Franceschi C, Pettener D, Luiselli D. Fine Dissection of Human Mitochondrial DNA Haplogroup HV Lineages Reveals Paleolithic Signatures from European Glacial Refugia. PLoS One. 2015 Dec 7;10(12):e0144391. doi: 10.1371/journal.pone.0144391. PMID: 26640946; PMCID: PMC4671665.

De Fanti S, Sazzini M, Giuliani C, Frazzoni F, Sarno S, Boattini A, Marasco E, Mantovani V, Franceschi C, Moral P, Garagnani P, Luiselli D. Inferring the genetic history of lactase persistence along the Italian peninsula from a large genomic interval surrounding the LCT gene. Am J Phys Anthropol. 2015 Dec;158(4):708-18. doi: 10.1002/ajpa.22814. Epub 2015 Jul 15. PMID: 26175013.

DeGiorgio M, Lohmueller KE, Nielsen R. A model-based approach for identifying signatures of ancient balancing selection in genetic data. PLoS Genet. 2014 Aug 21;10(8):e1004561. doi: 10.1371/journal.pgen.1004561. PMID: 25144706; PMCID: PMC4140648.

Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013 Jan;10(1):5-6. doi: 10.1038/nmeth.2307. PMID: 23269371.

Deng H, Li W. Monoacylglycerol lipase inhibitors: modulators for lipid metabolism in cancer malignancy, neurological and metabolic disorders. Acta Pharm Sin B. 2020 Apr;10(4):582-602. doi: 10.1016/j.apsb.2019.10.006. Epub 2019 Oct 18. PMID: 32322464; PMCID: PMC7161712.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011 May;43(5):491-8. doi: 10.1038/ng.806. Epub 2011 Apr 10. PMID: 21478889; PMCID: PMC3083463.

Destro Bisol G, Anagnostou P, Batini C, Battaggia C, Bertoncini S, Boattini A, Caciagli L, Caló MC, Capelli C, Capocasa M, Castrí L, Ciani G, Coia V, Corrias L, Crivellaro F, Ghiani ME, 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. Italian isolates today: geographic and linguistic factors shaping human biodiversity. J Anthropol Sci. 2008;86:179-88. PMID: 19934475.

Détroit F, Mijares AS, Corny J, Daver G, Zanolli C, Dizon E, Robles E, Grün R, Piper PJ. A new species of Homo from the Late Pleistocene of the Philippines. Nature. 2019 Apr;568(7751):181-186. doi: 10.1038/s41586-019-1067-9. Epub 2019 Apr 10. PMID: 30971845.

Di Gaetano C, Voglino F, Guarrera S, Fiorito G, Rosa F, Di Blasio AM, Manzini P, Dianzani I, Betti M, Cusi D, Frau F, Barlassina C, Mirabelli D, Magnani C, Glorioso N, Bonassi S, Piazza A, Matullo G. An overview of the genetic structure within the Italian population from genome-wide data. PLoS One. 2012;7(9):e43759. doi: 10.1371/journal.pone.0043759. Epub 2012 Sep 12. PMID: 22984441; PMCID: PMC3440425.

Dong LM, Parkin S, Trakhanov SD, Rupp B, Simmons T, Arnold KS, Newhouse YM, Innerarity TL, Weisgraber KH. Novel mechanism for defective receptor binding of apolipoprotein E2 in type III hyperlipoproteinemia. Nat Struct Biol. 1996 Aug;3(8):718-22. doi: 10.1038/nsb0896-718. PMID: 8756331.

Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, Sulovari A, Ebler J, Zhou W, Serra Mari R, Yilmaz F, Zhao X, Hsieh P, Lee J, Kumar S, Lin J, Rausch T, Chen Y, Ren J, Santamarina M, Höps W, Ashraf H, Chuang NT, Yang X, Munson KM, Lewis AP, Fairley S, Tallon LJ, Clarke WE, Basile AO, Byrska-Bishop M, Corvelo A, Evani US, Lu TY, Chaisson MJP, Chen J, Li C, Brand H, Wenger AM, Ghareghani M, Harvey WT, Raeder B, Hasenfeld P, Regier AA, Abel HJ, Hall IM, Flicek P, Stegle O, Gerstein MB, Tubio JMC, Mu Z, Li YI, Shi X, Hastie AR, Ye K, Chong Z, Sanders AD, Zody MC, Talkowski ME, Mills RE, Devine SE, Lee C, Korbel JO, Marschall T, Eichler EE. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. Science. 2021 Apr 2;372(6537):eabf7117. doi: 10.1126/science.abf7117. Epub 2021 Feb 25. PMID: 33632895.

Eysert F, Kinoshita PF, Mary A, Vaillant-Beuchot L, Checler F, Chami M. Molecular Dysfunctions of Mitochondria-Associated Membranes (MAMs) in Alzheimer's Disease. Int J Mol Sci. 2020 Dec 14;21(24):E9521. doi: 10.3390/ijms21249521. PMID: 33327665.

Fan Y, Silber SJ. Y Chromosome Infertility. 2002 Oct 31 [updated 2019 Aug 1]. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mirzaa G, Amemiya A, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993–2021. PMID: 20301513.

Fagny M, Patin E, Enard D, Barreiro LB, Quintana-Murci L, Laval G. Exploring the occurrence of classic selective sweeps in humans using whole-genome sequencing data sets. Mol Biol Evol. 2014 Jul;31(7):1850-68. doi: 10.1093/molbev/msu118. Epub 2014 Apr 1. PMID: 24694833.

Falk D. The natural endocast of Taung (Australopithecus africanus): insights from the unpublished papers of Raymond Arthur Dart. Am J Phys Anthropol. 2009;140 Suppl 49:49-65. doi: 10.1002/ajpa.21184. PMID: 19890860.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. On detecting incomplete soft or hard selective sweeps using haplotype structure. Mol Biol Evol. 2014 May;31(5):1275-91. doi: 10.1093/molbev/msu077. Epub 2014 Feb 18. PMID: 24554778; PMCID: PMC3995338.

Fiorito G, Di Gaetano C, Guarrera S, Rosa F, Feldman MW, Piazza A, Matullo G. The Italian genome reflects the history of Europe and the Mediterranean basin. Eur J Hum Genet. 2016 Jul;24(7):1056-62. doi: 10.1038/ejhg.2015.233. Epub 2015 Nov 11. PMID: 26554880; PMCID: PMC5070887.

Flachsbart F, Caliebe A, Kleindorp R, Blanché H, von Eller-Eberstein H, Nikolaus S, Schreiber S, Nebel A. Association of FOXO3A variation with human longevity confirmed in German centenarians. Proc Natl Acad Sci U S A. 2009 Feb 24;106(8):2700-5. doi: 10.1073/pnas.0809594106. Epub 2009 Feb 5. PMID: 19196970; PMCID: PMC2650329.

Flachsbart F, Dose J, Gentschew L, Geismann C, Caliebe A, Knecht C, Nygaard M, Badarinarayan N, ElSharawy A, May S, Luzius A, Torres GG, Jentzsch M, Forster M, Häsler R, Pallauf K, Lieb W, Derbois C, Galan P, Drichel D, Arlt A, Till A, Krause-Kyora B, Rimbach G, Blanché H, Deleuze JF, Christiansen L, Christensen K, Nothnagel M, Rosenstiel P, Schreiber S, Franke A, Sebens S, Nebel A. Identification and characterization of two functional variants in the human longevity gene FOXO3. Nat Commun. 2017 Dec 12;8(1):2063. doi: 10.1038/s41467-017-02183-y. Erratum in: Nat Commun. 2018 Jan 17;9(1):320. PMID: 29234056; PMCID: PMC5727304.

Foraker J, Millard SP, Leong L, Thomson Z, Chen S, Keene CD, Bekris LM, Yu CE. The APOE Gene is Differentially Methylated in Alzheimer's Disease. J Alzheimers Dis. 2015;48(3):745-55. doi: 10.3233/JAD-143060. PMID: 26402071; PMCID: PMC6469491.

Fortney K, Dobriban E, Garagnani P, Pirazzini C, Monti D, Mari D, Atzmon G, Barzilai N, Franceschi C, Owen AB, Kim SK. Genome-Wide Scan Informed by Age-Related Disease Identifies Loci for Exceptional Human Longevity. PLoS Genet. 2015 Dec 17;11(12):e1005728. doi: 10.1371/journal.pgen.1005728. PMID: 26677855; PMCID: PMC4683064.

Francalacci P, Morelli L, Angius A, Berutti R, Reinier F, Atzeni R, Pilu R, Busonero F, Maschio A, Zara I, Sanna D, Useli A, Urru MF, Marcelli M, Cusano R, Oppo M, Zoledziewska M, Pitzalis M, Deidda F, Porcu E, Poddie F, Kang HM, Lyons R, Tarrier B, Gresham JB, Li B, Tofanelli S, Alonso S, Dei M, Lai S, Mulas A, Whalen MB,

Uzzau S, Jones C, Schlessinger D, Abecasis GR, Sanna S, Sidore C, Cucca F. Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. Science. 2013 Aug 2;341(6145):565-9. doi: 10.1126/science.1237947. PMID: 23908240; PMCID: PMC5500864.

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013 Jan;41(Database issue):D808-15. doi: 10.1093/nar/gks1094. Epub 2012 Nov 29. PMID: 23203871; PMCID: PMC3531103.

Fu Q, Mittnik A, Johnson PLF, Bos K, Lari M, Bollongino R, Sun C, Giemsch L, Schmitz R, Burger J, Ronchitelli AM, Martini F, Cremonesi RG, Svoboda J, Bauer P, Caramelli D, Castellano S, Reich D, Pääbo S, Krause J. A revised timescale for human evolution based on ancient mitochondrial genomes. Curr Biol. 2013 Apr 8;23(7):553-559. doi: 10.1016/j.cub.2013.02.044. Epub 2013 Mar 21. PMID: 23523248; PMCID: PMC5036973.

Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, Furtwängler A, Haak W, Meyer M, Mittnik A, Nickel B, Peltzer A, Rohland N, Slon V, Talamo S, Lazaridis I, Lipson M, Mathieson I, Schiffels S, Skoglund P, Derevianko AP, Drozdov N, Slavinsky V, Tsybankov A, Cremonesi RG, Mallegni F, Gély B, Vacca E, Morales MR, Straus LG, Neugebauer-Maresch C, Teschler-Nicola M, Constantin S, Moldovan OT, Benazzi S, Peresani M, Coppola D, Lari M, Ricci S, Ronchitelli A, Valentin F, Thevenet C, Wehrberger K, Grigorescu D, Rougier H, Crevecoeur I, Flas D, Semal P, Mannino MA, Cupillard C, Bocherens H, Conard NJ, Harvati K, Moiseyev V, Drucker DG, Svoboda J, Richards MP, Caramelli D, Pinhasi R, Kelso J, Patterson N, Krause J, Pääbo S, Reich D. The genetic history of Ice Age Europe. Nature. 2016 Jun 9;534(7606):200-5. doi: 10.1038/nature17993. Epub 2016 May 2. PMID: 27135931; PMCID: PMC4943878.

Fuku N, Díaz-Peña R, Arai Y, Abe Y, Zempo H, Naito H, Murakami H, Miyachi M, Spuch C, Serra-Rexach JA, Emanuele E, Hirose N, Lucia A. Epistasis, physical capacity-related genes and exceptional longevity: FNDC5 gene interactions with candidate genes FOXOA3 and APOE. BMC Genomics. 2017 Nov 14;18(Suppl 8):803. doi: 10.1186/s12864-017-4194-4. PMID: 29143599; PMCID: PMC5688477.

Fullerton SM, Clark AG, Weiss KM, Nickerson DA, Taylor SL, Stengård JH, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF. Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. Am J Hum Genet. 2000 Oct;67(4):881-900. doi: 10.1086/303070. Epub 2000 Sep 13. PMID: 10986041; PMCID: PMC1287893.

Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admetlla A, Pattini L, Nielsen R. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet. 2011 Nov;7(11):e1002355. doi: 10.1371/journal.pgen.1002355. Epub 2011 Nov 3. Erratum in: PLoS Genet. 2011 Nov;7(11). doi:10.1371/annotation/ca428083-dbcb-476a-956c-d7bb6e317cf7. Ferrer-Admettla, Anna [corrected to Ferrer-Admetlla, Anna]. PMID: 22072984; PMCID: PMC3207877.

Gale DP, Molyneux K, Wimbury D, Higgins P, Levine AP, Caplin B, Ferlin A, Yin P, Nelson CP, Stanescu H, Samani NJ, Kleta R, Yu X, Barratt J. Galactosylation of IgA1 Is Associated with Common Variation in *C1GALT1*. J Am Soc Nephrol. 2017 Jul;28(7):2158-2166. doi: 10.1681/ASN.2016091043. Epub 2017 Feb 16. PMID: 28209808; PMCID: PMC5491291.

Garatachea N, Emanuele E, Calero M, Fuku N, Arai Y, Abe Y, Murakami H, Miyachi M, Yvert T, Verde Z, Zea MA, Venturini L, Santiago C, Santos-Lozano A, Rodríguez-Romo G, Ricevuti G, Hirose N, Rábano A, Lucia A. ApoE gene and exceptional longevity: Insights from three independent cohorts. Exp Gerontol. 2014 May;53:16-23. doi: 10.1016/j.exger.2014.02.004. Epub 2014 Feb 15. PMID: 24534555.

Garatachea N, Marín PJ, Santos-Lozano A, Sanchis-Gomar F, Emanuele E, Lucia A. The ApoE gene is related with exceptional longevity: a systematic review and meta-analysis. Rejuvenation Res. 2015 Feb;18(1):3-13. doi: 10.1089/rej.2014.1605. PMID: 25385258.

Geffin R, McCarthy M. Aging and Apolipoprotein E in HIV Infection. J Neurovirol. 2018 Oct;24(5):529-548. doi: 10.1007/s13365-018-0660-2. Epub 2018 Jul 9. Erratum in: J Neurovirol. 2018 Nov 19;: PMID: 29987582; PMCID: PMC6244718.

Gerdes LU, Jeune B, Ranberg KA, Nybo H, Vaupel JW. Estimation of apolipoprotein E genotype-specific relative mortality risks from the distribution of genotypes in centenarians and middle-aged men: apolipoprotein E gene

is a "frailty gene," not a "longevity gene". Genet Epidemiol. 2000 Oct;19(3):202-10. doi: 10.1002/1098-2272(200010)19:3<202::AID-GEPI2>3.0.CO;2-Q. PMID: 11015124.

Giuliani C, Garagnani P, Franceschi C. Genetics of Human Longevity Within an Eco-Evolutionary Nature-Nurture Framework. Circ Res. 2018 Sep 14;123(7):745-772. doi: 10.1161/CIRCRESAHA.118.312562. PMID: 30355083.

Giuliani C, Sazzini M, Pirazzini C, Bacalini MG, Marasco E, Ruscone GAG, Fang F, Sarno S, Gentilini D, Di Blasio AM, Crocco P, Passarino G, Mari D, Monti D, Nacmias B, Sorbi S, Salvarani C, Catanoso M, Pettener D, Luiselli D, Ukraintseva S, Yashin A, Franceschi C, Garagnani P. Impact of demography and population dynamics on the genetic architecture of human longevity. Aging (Albany NY). 2018 Aug 8;10(8):1947-1963. doi: 10.18632/aging.101515. PMID: 30089705; PMCID: PMC6128422.

Gnecchi-Ruscone GA, Abondio P, De Fanti S, Sarno S, Sherpa MG, Sherpa PT, Marinelli G, Natali L, Di Marcello M, Peluzzi D, Luiselli D, Pettener D, Sazzini M. Evidence of Polygenic Adaptation to High Altitude from Tibetan and Sherpa Genomes. Genome Biol Evol. 2018 Nov 1;10(11):2919-2930. doi: 10.1093/gbe/evy233. PMID: 30335146; PMCID: PMC6239493.

Goodman C, Goodman CS, Hur J, Jeyendran RS, Coulam C. The association of Apoprotein E polymorphisms with recurrent pregnancy loss. Am J Reprod Immunol. 2009 Jan;61(1):34-8. doi: 10.1111/j.1600-0897.2008.00659.x. PMID: 19086990.

Gouy A, Daub JT, Excoffier L. Detecting gene subnetworks under selection in biological pathways. Nucleic Acids Res. 2017 Sep 19;45(16):e149. doi: 10.1093/nar/gkx626. PMID: 28934485; PMCID: PMC5766194.

Grarup N, Moltke I, Andersen MK, Dalby M, Vitting-Seerup K, Kern T, Mahendran Y, Jørsboe E, Larsen CVL, Dahl-Petersen IK, Gilly A, Suveges D, Dedoussis G, Zeggini E, Pedersen O, Andersson R, Bjerregaard P, Jørgensen ME, Albrechtsen A, Hansen T. Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. Nat Genet. 2018 Feb;50(2):172-174. doi: 10.1038/s41588-017-0022-7. Epub 2018 Jan 8. PMID: 29311636; PMCID: PMC5828106.

Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA; 1000 Genomes Project, Bustamante CD. Demographic history and rare allele sharing among human populations. Proc Natl Acad Sci U S A. 2011 Jul 19;108(29):11983-8. doi: 10.1073/pnas.1019276108. Epub 2011 Jul 5. PMID: 21730125; PMCID: PMC3142009.

Guernier V, Hochberg ME, Guégan JF. Ecology drives the worldwide distribution of human diseases. PLoS Biol. 2004 Jun;2(6):e141. doi: 10.1371/journal.pbio.0020141. Epub 2004 Jun 15. PMID: 15208708; PMCID: PMC423130.

Gursky O, Atkinson D. Thermal unfolding of human high-density apolipoprotein A-1: implications for a lipid-free molten globular state. Proc Natl Acad Sci U S A. 1996 Apr 2;93(7):2991-5. doi: 10.1073/pnas.93.7.2991. PMID: 8610156; PMCID: PMC39748.

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K, Fu Q, Mittnik A, Bánffy E, Economou C, Francken M, Friederich S, Pena RG, Hallgren F, Khartanovich V, Khokhlov A, Kunst M, Kuznetsov P, Meller H, Mochalov O, Moiseyev V, Nicklisch N, Pichler SL, Risch R, Rojo Guerra MA, Roth C, Szécsényi-Nagy A, Wahl J, Meyer M, Krause J, Brown D, Anthony D, Cooper A, Alt KW, Reich D. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015 Jun 11;522(7555):207-11. doi: 10.1038/nature14317. Epub 2015 Mar 2. PMID: 25731166; PMCID: PMC5048219.

Haile-Selassie Y, Suwa G, White TD. Late Miocene teeth from Middle Awash, Ethiopia, and early hominid dental evolution. Science. 2004 Mar 5;303(5663):1503-5. doi: 10.1126/science.1092978. PMID: 15001775.

Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, Di Rienzo A. Adaptations to climate in candidate genes for common metabolic disorders. PLoS Genet. 2008 Feb;4(2):e32. doi: 10.1371/journal.pgen.0040032. PMID: 18282109; PMCID: PMC2242814.

Hartman RE, Wozniak DF, Nardi A, Olney JW, Sartorius L, Holtzman DM. Behavioral phenotyping of GFAP-apoE3 and -apoE4 transgenic mice: apoE4 mice show profound working memory impairments in the absence of

Alzheimer's-like neuropathology. Exp Neurol. 2001 Aug;170(2):326-44. doi: 10.1006/exnr.2001.7715. PMID: 11476599.

Hatzikotoulas K, Gilly A, Zeggini E. Using population isolates in genetic association studies. Brief Funct Genomics. 2014 Sep;13(5):371-7. doi: 10.1093/bfgp/elu022. Epub 2014 Jul 9. PMID: 25009120; PMCID: PMC4168662.

Hawkes K. Genomic evidence for the evolution of human postmenopausal longevity. Proc Natl Acad Sci U S A. 2016 Jan 5;113(1):17-8. doi: 10.1073/pnas.1522936113. Epub 2015 Dec 23. PMID: 26699495; PMCID: PMC4711858.

Hawkes K. The Centrality of Ancestral Grandmothering in Human Evolution. Integr Comp Biol. 2020 Sep 1;60(3):765-781. doi: 10.1093/icb/icaa029. PMID: 32386309.

Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, Myers S. A genetic atlas of human admixture history. Science. 2014 Feb 14;343(6172):747-751. doi: 10.1126/science.1243518. PMID: 24531965; PMCID: PMC4209567.

Henry N, Krammer EM, Stengel F, Adams Q, Van Liefferinge F, Hubin E, Chaves R, Efremov R, Aebersold R, Vandenbussche G, Prévost M, Raussens V, Deroo S. Lipidated apolipoprotein E4 structure and its receptor binding mechanism determined by a combined cross-linking coupled to mass spectrometry and molecular dynamics approach. PLoS Comput Biol. 2018 Jun 22;14(6):e1006165. doi: 10.1371/journal.pcbi.1006165. PMID: 29933361; PMCID: PMC6033463.

Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G; 1000 Genomes Project, Sella G, Przeworski M. Classic selective sweeps were rare in recent human evolution. Science. 2011 Feb 18;331(6019):920-4. doi: 10.1126/science.1198878. PMID: 21330547; PMCID: PMC3669691.

Herrero J, Muffato M, Beal K, Fitzgerald S, Gordon L, Pignatelli M, Vilella AJ, Searle SM, Amode R, Brent S, Spooner W, Kulesha E, Yates A, Flicek P. Ensembl comparative genomics resources. Database (Oxford). 2016 Feb 20;2016:bav096. doi: 10.1093/database/bav096. Erratum in: Database (Oxford). 2016;2016. pii: baw053. doi: 10.1093/database/baw053. PMID: 26896847; PMCID: PMC4761110.

Hofmanová Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Díez-Del-Molino D, van Dorp L, López S, Kousathanas A, Link V, Kirsanow K, Cassidy LM, Martiniano R, Strobel M, Scheu A, Kotsakis K, Halstead P, Triantaphyllou S, Kyparissi-Apostolika N, Urem-Kotsou D, Ziota C, Adaktylou F, Gopalan S, Bobo DM, Winkelbach L, Blöcher J, Unterländer M, Leuenberger C, Çilingiroğlu Ç, Horejs B, Gerritsen F, Shennan SJ, Bradley DG, Currat M, Veeramah KR, Wegmann D, Thomas MG, Papageorgopoulou C, Burger J. Early farmers from across Europe directly descended from Neolithic Aegeans. Proc Natl Acad Sci U S A. 2016 Jun 21;113(25):6886-91. doi: 10.1073/pnas.1523951113. Epub 2016 Jun 6. PMID: 27274049; PMCID: PMC4922144.

Hu P, Qin YH, Jing CX, Lu L, Hu B, Du PF. Does the geographical gradient of ApoE4 allele exist in China? A systemic comparison among multiple Chinese populations. Mol Biol Rep. 2011 Jan;38(1):489-94. doi: 10.1007/s11033-010-0132-0. Epub 2010 Mar 31. PMID: 20354905.

Huang RY, Garai K, Frieden C, Gross ML. Hydrogen/deuterium exchange and electron-transfer dissociation mass spectrometry determine the interface and dynamics of apolipoprotein E oligomerization. Biochemistry. 2011 Nov 1;50(43):9273-82. doi: 10.1021/bi2010027. Epub 2011 Oct 5. PMID: 21899263; PMCID: PMC3202646.

Huang Y, Mahley RW. Apolipoprotein E: structure and function in lipid metabolism, neurobiology, and Alzheimer's diseases. Neurobiol Dis. 2014 Dec;72 Pt A:3-12. doi: 10.1016/j.nbd.2014.08.025. Epub 2014 Aug 27. PMID: 25173806; PMCID: PMC4253862.

Hublin JJ, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, Bergmann I, Le Cabec A, Benazzi S, Harvati K, Gunz P. New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. Nature. 2017 Jun 7;546(7657):289-292. doi: 10.1038/nature22336. Erratum in: Nature. 2018 Jun;558(7711):E6. PMID: 28593953.

Huebbe P, Dose J, Schloesser A, Campbell G, Glüer CC, Gupta Y, Ibrahim S, Minihane AM, Baines JF, Nebel A, Rimbach G. Apolipoprotein E (APOE) genotype regulates body weight and fatty acid utilization-Studies in gene-targeted replacement mice. Mol Nutr Food Res. 2015 Feb;59(2):334-43. doi: 10.1002/mnfr.201400636. Epub 2014 Dec 12. PMID: 25381750.

Huebbe P, Lange J, Lietz G, Rimbach G. Dietary beta-carotene and lutein metabolism is modulated by the APOE genotype. Biofactors. 2016 Jul 8;42(4):388-96. doi: 10.1002/biof.1284. Epub 2016 Apr 4. PMID: 27040933.

Huebbe P, Rimbach G. Evolution of human apolipoprotein E (APOE) isoforms: Gene structure, protein function and interaction with dietary factors. Ageing Res Rev. 2017 Aug;37:146-161. doi: 10.1016/j.arr.2017.06.002. Epub 2017 Jun 21. PMID: 28647612.

Mounier A, Marchal F, Condemi S. Is Homo heidelbergensis a distinct species? New insight on the Mauer mandible. J Hum Evol. 2009 Mar;56(3):219-46. doi: 10.1016/j.jhevol.2008.12.006. Epub 2009 Feb 27. PMID: 19249816.

Hyman BT, Gomez-Isla T, West H, Briggs M, Chung H, Growdon JH, Rebeck GW. Clinical and neuropathological correlates of apolipoprotein E genotype in Alzheimer's disease. Window on molecular epidemiology. Ann N Y Acad Sci. 1996 Jan 17;777:158-65. doi: 10.1111/j.1749-6632.1996.tb34414.x. PMID: 8624079.

Itzhaki RF. Corroboration of a Major Role for Herpes Simplex Virus Type 1 in Alzheimer's Disease. Front Aging Neurosci. 2018 Oct 19;10:324. doi: 10.3389/fnagi.2018.00324. PMID: 30405395; PMCID: PMC6202583.

Itzhaki RF. Herpes and Alzheimer's Disease: Subversion in the Central Nervous System and How It Might Be Halted. J Alzheimers Dis. 2016 Oct 18;54(4):1273-1281. doi: 10.3233/JAD-160607. PMID: 27497484.

Itzhaki RF. Herpes simplex virus type 1 and Alzheimer's disease: possible mechanisms and signposts. FASEB J. 2017 Aug;31(8):3216-3226. doi: 10.1096/fj.201700360. PMID: 28765170.

Jablonski NG, Chaplin G. The evolution of human skin coloration. J Hum Evol. 2000 Jul;39(1):57-106. doi: 10.1006/jhev.2000.0403. PMID: 10896812.

Jablonski NG, Chaplin G. Colloquium paper: human skin pigmentation as an adaptation to UV radiation. Proc Natl Acad Sci U S A. 2010 May 11;107 Suppl 2(Suppl 2):8962-8. doi: 10.1073/pnas.0914628107. Epub 2010 May 5. PMID: 20445093; PMCID: PMC3024016.

Ji ZS, Müllendorff K, Cheng IH, Miranda RD, Huang Y, Mahley RW. Reactivity of apolipoprotein E4 and amyloid beta peptide: lysosomal stability and neurodegeneration. J Biol Chem. 2006 Feb 3;281(5):2683-92. doi: 10.1074/jbc.M506646200. Epub 2005 Nov 17. PMID: 16298992.

Jobling M, Hollox E, Hurles M, Kivisild T, Tyler-Smith C. Human Evolutionary Genetics, 2nd edition. New York: Garland Science.

Jobling MA, Tyler-Smith C. Human Y-chromosome variation in the genome-sequencing era. Nat Rev Genet. 2017 Aug;18(8):485-497. doi: 10.1038/nrg.2017.36. Epub 2017 May 30. PMID: 28555659.

Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R, McLaughlin RL, Gallego Llorente M, Cassidy LM, Gamba C, Meshveliani T, Bar-Yosef O, Müller W, Belfer-Cohen A, Matskevich Z, Jakeli N, Higham TFG, Currat M, Lordkipanidze D, Hofreiter M, Manica A, Pinhasi R, Bradley DG. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat Commun. 2015 Nov 16;6:8912. doi: 10.1038/ncomms9912. PMID: 26567969; PMCID: PMC4660371.

Kaniewski D, Van Campo E, Guiot J, Le Burel S, Otto T, Baeteman C. Environmental roots of the late bronze age crisis. PLoS One. 2013 Aug 14;8(8):e71004. doi: 10.1371/journal.pone.0071004. PMID: 23967146; PMCID: PMC3743901.

Kasahara K, Tanoue T, Yamashita T, Yodoi K, Matsumoto T, Emoto T, Mizoguchi T, Hayashi T, Kitano N, Sasaki N, Atarashi K, Honda K, Hirata KI. Commensal bacteria at the crossroad between cholesterol homeostasis and chronic inflammation in atherosclerosis. J Lipid Res. 2017 Mar;58(3):519-528. doi: 10.1194/jlr.M072165. Epub 2017 Jan 27. PMID: 28130274; PMCID: PMC5335582.

Keele GR, Prokop JW, He H, Holl K, Littrell J, Deal A, Francic S, Cui L, Gatti DM, Broman KW, Tschannen M, Tsaih SW, Zagloul M, Kim Y, Baur B, Fox J, Robinson M, Levy S, Flister MJ, Mott R, Valdar W, Solberg Woods LC. Genetic Fine-Mapping and Identification of Candidate Genes and Variants for Adiposity Traits in Outbred Rats. Obesity (Silver Spring). 2018 Jan;26(1):213-222. doi: 10.1002/oby.22075. Epub 2017 Nov 28. PMID: 29193816; PMCID: PMC5740008.

Keinan A, Clark AG. Recent explosive human population growth has resulted in an excess of rare genetic variants. Science. 2012 May 11;336(6082):740-3. doi: 10.1126/science.1217283. PMID: 22582263; PMCID: PMC3586590.

Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. Genomic runs of homozygosity record population history and consanguinity. PLoS One. 2010 Nov 15;5(11):e13996. doi: 10.1371/journal.pone.0013996. PMID: 21085596; PMCID: PMC2981575.

Kiryluk K, Li Y, Sanna-Cherchi S, Rohanizadegan M, Suzuki H, Eitner F, Snyder HJ, Choi M, Hou P, Scolari F, Izzi C, Gigante M, Gesualdo L, Savoldi S, Amoroso A, Cusi D, Zamboli P, Julian BA, Novak J, Wyatt RJ, Mucha K, Perola M, Kristiansson K, Viktorin A, Magnusson PK, Thorleifsson G, Thorsteinsdottir U, Stefansson K, Boland A, Metzger M, Thibaudin L, Wanner C, Jager KJ, Goto S, Maixnerova D, Karnib HH, Nagy J, Panzer U, Xie J, Chen N, Tesar V, Narita I, Berthoux F, Floege J, Stengel B, Zhang H, Lifton RP, Gharavi AG. Geographic differences in genetic susceptibility to IgA nephropathy: GWAS replication study and geospatial risk analysis. PLoS Genet. 2012;8(6):e1002765. doi: 10.1371/journal.pgen.1002765. Epub 2012 Jun 21. PMID: 22737082; PMCID: PMC3380840.

Kivisild T. Maternal ancestry and population history from whole mitochondrial genomes. Investig Genet. 2015 Mar 10;6:3. doi: 10.1186/s13323-015-0022-2. PMID: 25798216; PMCID: PMC4367903.

Kockx M, Traini M, Kritharides L. Cell-specific production, secretion, and function of apolipoprotein E. J Mol Med (Berl). 2018 May;96(5):361-371. doi: 10.1007/s00109-018-1632-y. Epub 2018 Mar 7. PMID: 29516132.

Kregel KC, Zhang HJ. An integrated view of oxidative stress in aging: basic mechanisms, functional effects, and pathological considerations. Am J Physiol Regul Integr Comp Physiol. 2007 Jan;292(1):R18-36. doi: 10.1152/ajpregu.00327.2006. Epub 2006 Aug 17. PMID: 16917020.

Kristiansson K, Naukkarinen J, Peltonen L. Isolated populations and complex disease gene identification. Genome Biol. 2008;9(8):109. doi: 10.1186/gb-2008-9-8-109. Epub 2008 Aug 26. PMID: 18771588; PMCID: PMC2575505.

Kulminski AM, Huang J, Wang J, He L, Loika Y, Culminskaya I. Apolipoprotein E region molecular signatures of Alzheimer's disease. Aging Cell. 2018 Aug;17(4):e12779. doi: 10.1111/acel.12779. Epub 2018 May 23. PMID: 29797398; PMCID: PMC6052488.

Kulminski AM, Loika Y, Culminskaya I, Huang J, Arbeev KG, Bagley O, Feitosa MF, Zmuda JM, Christensen K, Yashin AI; Long Life Family Study Research Group. Independent associations of TOMM40 and APOE variants with body mass index. Aging Cell. 2019 Feb;18(1):e12869. doi: 10.1111/acel.12869. Epub 2018 Nov 21. PMID: 30462377; PMCID: PMC6351823.

Kulminski AM, Raghavachari N, Arbeev KG, Culminskaya I, Arbeeva L, Wu D, Ukraintseva SV, Christensen K, Yashin AI. Protective role of the apolipoprotein E2 allele in age-related disease traits and survival: evidence from the Long Life Family Study. Biogerontology. 2016 Nov;17(5-6):893-905. doi: 10.1007/s10522-016-9659-3. Epub 2016 Jul 22. PMID: 27447179; PMCID: PMC5065761.

Kumar NT, Liestøl K, Løberg EM, Reims HM, Brorson SH, Maehlen J. The apolipoprotein E polymorphism and cardiovascular diseases--an autopsy study. Cardiovasc Pathol. 2012 Nov-Dec;21(6):461-9. doi: 10.1016/j.carpath.2012.02.005. Epub 2012 Mar 21. PMID: 22440829.

Langergraber KE, Prüfer K, Rowney C, Boesch C, Crockford C, Fawcett K, Inoue E, Inoue-Muruyama M, Mitani JC, Muller MN, Robbins MM, Schubert G, Stoinski TS, Viola B, Watts D, Wittig RM, Wrangham RW, Zuberbühler K, Pääbo S, Vigilant L. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. Proc Natl Acad Sci U S A. 2012 Sep 25;109(39):15716-21. doi: 10.1073/pnas.1211740109. Epub 2012 Aug 13. PMID: 22891323; PMCID: PMC3465451.

Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, Balascakova M, Bertranpetit J, Bindoff LA, Comas D, Holmlund G, Kouvatsi A, Macek M, Mollet I, Parson W, Palo J, Ploski R, Sajantila A, Tagliabracci A, Gether U, Werge T, Rivadeneira F, Hofman A, Uitterlinden AG, Gieger C, Wichmann HE, Rüther A, Schreiber S, Becker C, Nürnberg P, Nelson MR, Krawczak M, Kayser M. Correlation between genetic and geographic structure in Europe. Curr Biol. 2008 Aug 26;18(16):1241-8. doi: 10.1016/j.cub.2008.07.049. Epub 2008 Aug 7. PMID: 18691889.

Larmuseau MH, Vanderheyden N, Van Geystelen A, van Oven M, de Knijff P, Decorte R. Recent radiation within Y-chromosomal haplogroup R-M269 resulted in high Y-STR haplotype resemblance. Ann Hum Genet. 2014 Mar;78(2):92-103. doi: 10.1111/ahg.12050. PMID: 24571229.

Lau YC. Y chromosome in health and diseases. Cell Biosci. 2020 Aug 13;10:97. doi: 10.1186/s13578-020-00452-w. PMID: 32817785; PMCID: PMC7427063.

Lawless C, Greaves L, Reeve AK, Turnbull DM, Vincent AE. The rise and rise of mitochondrial DNA mutations. Open Biol. 2020 May;10(5):200061. doi: 10.1098/rsob.200061. Epub 2020 May 20. PMID: 32428418; PMCID: PMC7276526.

Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. PLoS Genet. 2012 Jan;8(1):e1002453. doi: 10.1371/journal.pgen.1002453. Epub 2012 Jan 26. PMID: 22291602; PMCID: PMC3266881.

Lazaridis I, Mittnik A, Patterson N, Mallick S, Rohland N, Pfrengle S, Furtwängler A, Peltzer A, Posth C, Vasilakis A, McGeorge PJP, Konsolaki-Yannopoulou E, Korres G, Martlew H, Michalodimitrakis M, Özsait M, Özsait N, Papathanasiou A, Richards M, Roodenberg SA, Tzedakis Y, Arnott R, Fernandes DM, Hughey JR, Lotakis DM, Navas PA, Maniatis Y, Stamatoyannopoulos JA, Stewardson K, Stockhammer P, Pinhasi R, Reich D, Krause J, Stamatoyannopoulos G. Genetic origins of the Minoans and Mycenaeans. Nature. 2017 Aug 10;548(7666):214-218. doi: 10.1038/nature23310. Epub 2017 Aug 2. PMID: 28783727; PMCID: PMC5565772.

Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, Fernandes D, Novak M, Gamarra B, Sirak K, Connell S, Stewardson K, Harney E, Fu Q, Gonzalez-Fortes G, Jones ER, Roodenberg SA, Lengyel G, Bocquentin F, Gasparian B, Monge JM, Gregg M, Eshed V, Mizrahi AS, Meiklejohn C, Gerritsen F, Bejenaru L, Blüher M, Campbell A, Cavalleri G, Comas D, Froguel P, Gilbert E, Kerr SM, Kovacs P, Krause J, McGettigan D, Merrigan M, Merriwether DA, O'Reilly S, Richards MB, Semino O, Shamoon-Pour M, Stefanescu G, Stumvoll M, Tönjes A, Torroni A, Wilson JF, Yengo L, Hovhannisyan NA, Patterson N, Pinhasi R, Reich D. Genomic insights into the origin of farming in the ancient Near East. Nature. 2016 Aug 25;536(7617):419-24. doi: 10.1038/nature19310. Epub 2016 Jul 25. PMID: 27459054; PMCID: PMC5003663.

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, Berger B, Economou C, Bollongino R, Fu Q, Bos KI, Nordenfelt S, Li H, de Filippo C, Prüfer K, Sawyer S, Posth C, Haak W, Hallgren F, Fornander E, Rohland N, Delsate D, Francken M, Guinet JM, Wahl J, Ayodo G, Babiker HA, Bailliet G, Balanovska E, Balanovsky O, Barrantes R, Bedoya G, Ben-Ami H, Bene J, Berrada F, Bravi CM, Brisighelli F, Busby GB, Cali F, Churnosov M, Cole DE, Corach D, Damba L, van Driem G, Dryomov S, Dugoujon JM, Fedorova SA, Gallego Romero I, Gubina M, Hammer M, Henn BM, Hervig T, Hodoglugil U, Jha AR, Karachanak-Yankova S, Khusainova R, Khusnutdinova E, Kittles R, Kivisild T, Klitz W, Kučinskas V, Kushniarevich A, Laredj L, Litvinov S, Loukidis T, Mahley RW, Melegh B, Metspalu E, Molina J, Mountain J, Näkkäläjärvi K, Nesheva D, Nyambo T, Osipova L, Parik J, Platonov F, Posukh O, Romano V, Rothhammer F, Rudan I, Ruizbakiev R, Sahakyan H, Sajantila A, Salas A, Starikovskaya EB, Tarekegn A, Toncheva D, Turdikulova S, Uktveryte I, Utevska O, Vasquez R, Villena M, Voevoda M, Winkler CA, Yepiskoposyan L, Zalloua P, Zemunik T, Cooper A, Capelli C, Thomas MG, Ruiz-Linares A, Tishkoff SA, Singh L, Thangaraj K, Villems R, Comas D, Sukernik R, Metspalu M, Meyer M, Eichler EE, Burger J, Slatkin M, Pääbo S, Kelso J, Reich D, Krause J. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature. 2014 Sep 18;513(7518):409-13. doi: 10.1038/nature13673. PMID: 25230663; PMCID: PMC4170574.

Levy C, Khaled M, Fisher DE. MITF: master regulator of melanocyte development and melanoma oncogene. Trends Mol Med. 2006 Sep;12(9):406-14. doi: 10.1016/j.molmed.2006.07.008. Epub 2006 Aug 8. PMID: 16899407.

Li G, Bien-Ly N, Andrews-Zwilling Y, Xu Q, Bernardo A, Ring K, Halabisky B, Deng C, Mahley RW, Huang Y. GABAergic interneuron dysfunction impairs hippocampal neurogenesis in adult apolipoprotein E4 knockin mice. Cell Stem Cell. 2009 Dec 4;5(6):634-45. doi: 10.1016/j.stem.2009.10.015. PMID: 19951691; PMCID: PMC2992822.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. Worldwide human relationships inferred from genome-wide patterns of variation. Science. 2008 Feb 22;319(5866):1100-4. doi: 10.1126/science.1153717. PMID: 18292342.

Lipson M, Szécsényi-Nagy A, Mallick S, Pósa A, Stégmár B, Keerl V, Rohland N, Stewardson K, Ferry M, Michel M, Oppenheimer J, Broomandkhoshbacht N, Harney E, Nordenfelt S, Llamas B, Gusztáv Mende B, Köhler K, Oross K, Bondár M, Marton T, Osztás A, Jakucs J, Paluch T, Horváth F, Csengeri P, Koós J, Sebők K, Anders A, Raczky P, Regenye J, Barna JP, Fábián S, Serlegi G, Toldi Z, Gyöngyvér Nagy E, Dani J, Molnár E, Pálfi G, Márk L, Melegh B, Bánfai Z, Domboróczki L, Fernández-Eraso J, Antonio Mujika-Alustiza J, Alonso Fernández C, Jiménez Echevarría J, Bollongino R, Orschiedt J, Schierhold K, Meller H, Cooper A, Burger J, Bánffy E, Alt KW, Lalueza-Fox C, Haak W, Reich D. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature. 2017 Nov 16;551(7680):368-372. doi: 10.1038/nature24476. Epub 2017 Nov 8. PMID: 29144465; PMCID: PMC5973800.

Liu J, Zhao W, Ware EB, Turner ST, Mosley TH, Smith JA. DNA methylation in the APOE genomic region is associated with cognitive function in African Americans. BMC Med Genomics. 2018 May 8;11(1):43. doi: 10.1186/s12920-018-0363-9. PMID: 29739406; PMCID: PMC5941603.

López S, van Dorp L, Hellenthal G. Human Dispersal Out of Africa: A Lasting Debate. Evol Bioinform Online. 2016 Apr 21;11(Suppl 2):57-68. doi: 10.4137/EBO.S33489. PMID: 27127403; PMCID: PMC4844272.

Love JE, Day RJ, Gause JW, Brown RJ, Pu X, Theis DI, Caraway CA, Poon WW, Rahman AA, Morrison BE, Rohn TT. Nuclear uptake of an amino-terminal fragment of apolipoprotein E4 promotes cell death and localizes within microglia of the Alzheimer's disease brain. Int J Physiol Pathophysiol Pharmacol. 2017 Apr 15;9(2):40-57. PMID: 28533891; PMCID: PMC5435672.

Luca F, Perry GH, Di Rienzo A. Evolutionary adaptations to dietary changes. Annu Rev Nutr. 2010 Aug 21;30:291-314. doi: 10.1146/annurev-nutr-080508-141048. PMID: 20420525; PMCID: PMC4163920.

Luis JR, Palencia-Madrid L, Mendoza VC, Garcia-Bertrand R, de Pancorbo MM, Herrera RJ. The Y chromosome of autochthonous Basque populations and the Bronze Age replacement. Sci Rep. 2021 Mar 10;11(1):5607. doi: 10.1038/s41598-021-84915-1. PMID: 33692401; PMCID: PMC7970938.

Mahley RW, Huang Y, Weisgraber KH. Putting cholesterol in its place: apoE and reverse cholesterol transport. J Clin Invest. 2006 May;116(5):1226-9. doi: 10.1172/JCI28632. PMID: 16670767; PMCID: PMC1451226.

Mahley RW, Innerarity TL, Rall SC Jr, Weisgraber KH. Plasma lipoproteins: apolipoprotein structure and function. J Lipid Res. 1984 Dec 1;25(12):1277-94. PMID: 6099394.

Malaspinas AS, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, Bergström A, Athanasiadis G, Cheng JY, Crawford JE, Heupink TH, Macholdt E, Peischl S, Rasmussen S, Schiffels S, Subramanian S, Wright JL, Albrechtsen A, Barbieri C, Dupanloup I, Eriksson A, Margaryan A, Moltke I, Pugach I, Korneliussen TS, Levkivskyi IP, Moreno-Mayar JV, Ni S, Racimo F, Sikora M, Xue Y, Aghakhanian FA, Brucato N, Brunak S, Campos PF, Clark W, Ellingvåg S, Fourmile G, Gerbault P, Injie D, Koki G, Leavesley M, Logan B, Lynch A, Matisoo-Smith EA, McAllister PJ, Mentzer AJ, Metspalu M, Migliano AB, Murgha L, Phipps ME, Pomat W, Reynolds D, Ricaut FX, Siba P, Thomas MG, Wales T, Wall CM, Oppenheimer SJ, Tyler-Smith C, Durbin R, Dortch J, Manica A, Schierup MH, Foley RA, Lahr MM, Bowern C, Wall JD, Mailund T, Stoneking M, Nielsen R, Sandhu MS, Excoffier L, Lambert DM, Willerslev E. A genomic history of Aboriginal Australia. Nature. 2016 Oct 13;538(7624):207-214. doi: 10.1038/nature18299. Epub 2016 Sep 21. PMID: 27654914.

Maletta R, Smirne N, Bernardi L, Anfossi M, Gallo M, Conidi ME, Colao R, Puccio G, Curcio SAM, Laganà V, Frangipane F, Cupidi C, Mirabelli M, Vasso F, Torchia G, Muraca MG, Di Lorenzo R, Rose G, Montesanto A, Passarino G, Bruni AC. Frequency of Cardiovascular Genetic Risk Factors in a Calabrian Population and Their Effects on Dementia. J Alzheimers Dis. 2018;1(3):1179-1187. doi: 10.3233/JAD-170687. PMID: 29332048.

Malina RM, Little BB. Physical activity: the present in the context of the past. Am J Hum Biol. 2008 Jul-Aug;20(4):373-91. doi: 10.1002/ajhb.20772. PMID: 18433002.

Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, Skoglund P, Lazaridis I, Sankararaman S, Fu Q, Rohland N, Renaud G, Erlich Y, Willems T, Gallo C, Spence JP, Song YS, Poletti G, Balloux F, van Driem G, de Knijff P, Romero IG, Jha AR, Behar DM, Bravi CM, Capelli C, Hervig T, Moreno-Estrada A, Posukh OL, Balanovska E, Balanovsky O, Karachanak-Yankova S, Sahakyan H, Toncheva D, Yepiskoposyan L, Tyler-Smith C, Xue Y, Abdullah MS, Ruiz-Linares A, Beall CM, Di Rienzo A, Jeong C, Starikovskaya EB, Metspalu E, Parik J, Villems R, Henn BM, Hodoglugil U, Mahley R, Sajantila A, Stamatoyannopoulos G, Wee JT, Khusainova R, Khusnutdinova E, Litvinov S, Ayodo G, Comas D, Hammer

MF, Kivisild T, Klitz W, Winkler CA, Labuda D, Bamshad M, Jorde LB, Tishkoff SA, Watkins WS, Metspalu M, Dryomov S, Sukernik R, Singh L, Thangaraj K, Pääbo S, Kelso J, Patterson N, Reich D. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 2016 Oct 13;538(7624):201-206. doi: 10.1038/nature18964. Epub 2016 Sep 21. PMID: 27654912; PMCID: PMC5161557.

Marcus JH, Posth C, Ringbauer H, Lai L, Skeates R, Sidore C, Beckett J, Furtwängler A, Olivieri A, Chiang CWK, Al-Asadi H, Dey K, Joseph TA, Liu CC, Der Sarkissian C, Radzevičiūtė R, Michel M, Gradoli MG, Marongiu P, Rubino S, Mazzarello V, Rovina D, La Fragola A, Serra RM, Bandiera P, Bianucci R, Pompianu E, Murgia C, Guirguis M, Orquin RP, Tuross N, van Dommelen P, Haak W, Reich D, Schlessinger D, Cucca F, Krause J, Novembre J. Genetic history from the Middle Neolithic to present on the Mediterranean island of Sardinia. Nat Commun. 2020 Feb 24;11(1):939. doi: 10.1038/s41467-020-14523-6. PMID: 32094358; PMCID: PMC7039977.

Martínez-Morillo E, Hansson O, Atagi Y, Bu G, Minthon L, Diamandis EP, Nielsen HM. Total apolipoprotein E levels and specific isoform composition in cerebrospinal fluid and plasma from Alzheimer's disease patients and controls. Acta Neuropathol. 2014 May;127(5):633-43. doi: 10.1007/s00401-014-1266-2. Epub 2014 Mar 15. PMID: 24633805.

Martino, P. L'Isola grecanica dell'Aspromonte. Aspetti sociolinguistici, in *Atti dell'XI Congresso Nazionale di Studi (Cagliari 27-30 maggio 1977)*, (ed. Leoni, F. A.) 305-341 (Bulzoni, 1980).

Mathieson I, Alpaslan-Roodenberg S, Posth C, Szécsényi-Nagy A, Rohland N, Mallick S, Olalde I, Broomandkhoshbacht N, Candilio F, Cheronet O, Fernandes D, Ferry M, Gamarra B, Fortes GG, Haak W, Harney E, Jones E, Keating D, Krause-Kyora B, Kucukkalipci I, Michel M, Mittnik A, Nägele K, Novak M, Oppenheimer J, Patterson N, Pfrengle S, Sirak K, Stewardson K, Vai S, Alexandrov S, Alt KW, Andreescu R, Antonović D, Ash A, Atanassova N, Bacvarov K, Gusztáv MB, Bocherens H, Bolus M, Boroneanţ A, Boyadzhiev Y, Budnik A, Burmaz J, Chohadzhiev S, Conard NJ, Cottiaux R, Čuka M, Cupillard C, Drucker DG, Elenski N, Francken M, Galabova B, Ganetsovski G, Gély B, Hajdu T, Handzhyiska V, Harvati K, Higham T, Iliev S, Janković I, Karavanić I, Kennett DJ, Komšo D, Kozak A, Labuda D, Lari M, Lazar C, Leppek M, Leshtakov K, Vetro DL, Los D, Lozanov I, Malina M, Martini F, McSweeney K, Meller H, Menđušić M, Mirea P, Moiseyev V, Petrova V, Price TD, Simalcsik A, Sineo L, Šlaus M, Slavchev V, Stanev P, Starović A, Szeniczey T, Talamo S, Teschler-Nicola M, Thevenet C, Valchev I, Valentin F, Vasilyev S, Veljanovska F, Venelinova S, Veselovskaya E, Viola B, Virag C, Zaninović J, Zäuner S, Stockhammer PW, Catalano G, Krauß R, Caramelli D, Zariņa G, Gaydarska B, Lillie M, Nikitin AG, Potekhina I, Papathanasiou A, Borić D, Bonsall C, Krause J, Pinhasi R, Reich D. The genomic history of southeastern Europe. Nature. 2018 Mar 8;555(7695):197-203. doi: 10.1038/nature25778. Epub 2018 Feb 21. PMID: 29466330; PMCID: PMC6091220.

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, Sirak K, Gamba C, Jones ER, Llamas B, Dryomov S, Pickrell J, Arsuaga JL, de Castro JM, Carbonell E, Gerritsen F, Khokhlov A, Kuznetsov P, Lozano M, Meller H, Mochalov O, Moiseyev V, Guerra MA, Roodenberg J, Vergès JM, Krause J, Cooper A, Alt KW, Brown D, Anthony D, Lalueza-Fox C, Haak W, Pinhasi R, Reich D. Genome-wide patterns of selection in 230 ancient Eurasians. Nature. 2015 Dec 24;528(7583):499-503. doi: 10.1038/nature16152. Epub 2015 Nov 23. PMID: 26595274; PMCID: PMC4918750.

Mathieson I, McVean G. Differential confounding of rare and common variants in spatially structured populations. Nat Genet. 2012 Feb 5;44(3):243-6. doi: 10.1038/ng.1074. PMID: 22306651; PMCID: PMC3303124.

McGuckin MA, Lindén SK, Sutton P, Florin TH. Mucin dynamics and enteric pathogens. Nat Rev Microbiol. 2011 Apr;9(4):265-78. doi: 10.1038/nrmicro2538. PMID: 21407243.

Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 2019 Jan 8;47(D1):D419-D426. doi: 10.1093/nar/gky1038. PMID: 30407594; PMCID: PMC6323939.

Moghadasian MH, McManus BM, Nguyen LB, Shefer S, Nadji M, Godin DV, Green TJ, Hill J, Yang Y, Scudamore CH, Frohlich JJ. Pathophysiology of apolipoprotein E deficiency in mice: relevance to apo E-related disorders in humans. FASEB J. 2001 Dec;15(14):2623-30. doi: 10.1096/fj.01-0463com. PMID: 11726538.

Morrow JA, Hatters DM, Lu B, Hochtl P, Oberg KA, Rupp B, Weisgraber KH. Apolipoprotein E4 forms a molten globule. A potential basis for its association with disease. J Biol Chem. 2002 Dec 27;277(52):50380-5. doi: 10.1074/jbc.M204898200. Epub 2002 Oct 21. PMID: 12393895.

Mosino, F. *Minoranze etniche in Calabria e Basilicata*. (Di Mauro ed., 1988).

Mostafavi H, Berisa T, Day FR, Perry JRB, Przeworski M, Pickrell JK. Identifying genetic variants that affect viability in large cohorts. PLoS Biol. 2017 Sep 5;15(9):e2002458. doi: 10.1371/journal.pbio.2002458. PMID: 28873088; PMCID: PMC5584811.

Mounier A, Marchal F, Condemi S. Is Homo heidelbergensis a distinct species? New insight on the Mauer mandible. J Hum Evol. 2009 Mar;56(3):219-46. doi: 10.1016/j.jhevol.2008.12.006. Epub 2009 Feb 27. PMID: 19249816.

Myres NM, Rootsi S, Lin AA, Järve M, King RJ, Kutuev I, Cabrera VM, Khusnutdinova EK, Pshenichnov A, Yunusbayev B, Balanovsky O, Balanovska E, Rudan P, Baldovic M, Herrera RJ, Chiaroni J, Di Cristofaro J, Villems R, Kivisild T, Underhill PA. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. Eur J Hum Genet. 2011 Jan;19(1):95-101. doi: 10.1038/ejhg.2010.146. Epub 2010 Aug 25. PMID: 20736979; PMCID: PMC3039512.

Natali E, Forgia V. The beginning of the Neolithic in Southern Italy and Sicily. Quat. Int 2018; 470(B):253-69. doi: 10.1016/j.quaint.2017.07.004

Nesse RM, Bergstrom CT, Ellison PT, Flier JS, Gluckman P, Govindaraju DR, Niethammer D, Omenn GS, Perlman RL, Schwartz MD, Thomas MG, Stearns SC, Valle D. Evolution in health and medicine Sackler colloquium: Making evolutionary biology a basic science for medicine. Proc Natl Acad Sci U S A. 2010 Jan 26;107 Suppl 1(Suppl 1):1800-7. doi: 10.1073/pnas.0906224106. Epub 2009 Nov 16. PMID: 19918069; PMCID: PMC2868284.

Nesse RM, Stearns SC. The great opportunity: Evolutionary applications to medicine and public health. Evol Appl. 2008 Feb;1(1):28-48. doi: 10.1111/j.1752-4571.2007.00006.x. PMID: 25567489; PMCID: PMC3352398.

Nesse RM. Ten questions for evolutionary studies of disease vulnerability. Evol Appl. 2011 Mar;4(2):264-77. doi: 10.1111/j.1752-4571.2010.00181.x. Epub 2011 Jan 17. PMID: 25567972; PMCID: PMC3352562.

Nguyen D, Dhanasekaran P, Nickel M, Mizuguchi C, Watanabe M, Saito H, Phillips MC, Lund-Katz S. Influence of domain stability on the properties of human apolipoprotein E3 and E4 and mouse apolipoprotein E. Biochemistry. 2014 Jun 24;53(24):4025-33. doi: 10.1021/bi500340z. PMID: 24871385; PMCID: PMC4071092.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD. Genes mirror geography within Europe. Nature. 2008 Nov 6;456(7218):98-101. doi: 10.1038/nature07331. Epub 2008 Aug 31. Erratum in: Nature. 2008 Nov 13;456(7219):274. PMID: 18758442; PMCID: PMC2735096.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 1999 Jan 1;27(1):29-34. doi: 10.1093/nar/27.1.29. PMID: 9847135; PMCID: PMC148090.

Ojeda-Granados C, Panduro A, Gonzalez-Aldaco K, Sepulveda-Villegas M, Rivera-Iñiguez I, Roman S. Tailoring Nutritional Advice for Mexicans Based on Prevalence Profiles of Diet-Related Adaptive Gene Polymorphisms. J Pers Med. 2017 Nov 10;7(4):16. doi: 10.3390/jpm7040016. PMID: 29125573; PMCID: PMC5748628.

Olalde I, Brace S, Allentoft ME, Armit I, Kristiansen K, Booth T, Rohland N, Mallick S, Szécsényi-Nagy A, Mittnik A, Altena E, Lipson M, Lazaridis I, Harper TK, Patterson N, Broomandkhoshbacht N, Diekmann Y, Faltyskova Z, Fernandes D, Ferry M, Harney E, de Knijff P, Michel M, Oppenheimer J, Stewardson K, Barclay A, Alt KW, Liesau C, Ríos P, Blasco C, Miguel JV, García RM, Fernández AA, Bánffy E, Bernabò-Brea M, Billoin D, Bonsall C, Bonsall L, Allen T, Büster L, Carver S, Navarro LC, Craig OE, Cook GT, Cunliffe B, Denaire A, Dinwiddy KE, Dodwell N, Ernée M, Evans C, Kuchařík M, Farré JF, Fowler C, Gazenbeek M, Pena RG, Haber-Uriarte M, Haduch E, Hey G, Jowett N, Knowles T, Massy K, Pfrengle S, Lefranc P, Lemercier O, Lefebvre A, Martínez CH, Olmo VG, Ramírez AB, Maurandi JL, Majó T, McKinley JI, McSweeney K, Mende BG, Modi A, Kulcsár G, Kiss V, Czene A, Patay R, Endrődi A, Köhler K, Hajdu T, Szeniczey T, Dani J, Bernert Z, Hoole M, Cheronet O, Keating D, Velemínský P, Dobeš M, Candilio F, Brown F, Fernández RF, Herrero-Corral AM, Tusa S, Carnieri E, Lentini L, Valenti A, Zanini A, Waddington C, Delibes G, Guerra-Doce E, Neil B, Brittain M, Luke M, Mortimer R, Desideri J, Besse M, Brücken G, Furmanek M, Hałuszko A, Mackiewicz M, Rapiński A, Leach S, Soriano I, Lillios KT, Cardoso JL, Pearson MP, Włodarczak P, Price TD, Prieto P, Rey PJ, Risch R, Rojo Guerra MA, Schmitt A, Serralongue J, Silva AM, Smrčka V, Vergnaud L, Zilhão J, Caramelli D, Higham T, Thomas MG, Kennett DJ, Fokkens H, Heyd V, Sheridan A, Sjögren KG, Stockhammer PW, Krause J, Pinhasi R, Haak W, Barnes I, Lalueza-Fox C, Reich D. The Beaker phenomenon and the genomic transformation of

northwest Europe. Nature. 2018 Mar 8;555(7695):190-196. doi: 10.1038/nature25738. Epub 2018 Feb 21. Erratum in: Nature. 2018 Mar 21;555(7697):543. PMID: 29466337; PMCID: PMC5973796.

Olivieri A, Sidore C, Achilli A, Angius A, Posth C, Furtwängler A, Brandini S, Capodiferro MR, Gandini F, Zoledziewska M, Pitzalis M, Maschio A, Busonero F, Lai L, Skeates R, Gradoli MG, Beckett J, Marongiu M, Mazzarello V, Marongiu P, Rubino S, Rito T, Macaulay V, Semino O, Pala M, Abecasis GR, Schlessinger D, Conde-Sousa E, Soares P, Richards MB, Cucca F, Torroni A. Mitogenome Diversity in Sardinians: A Genetic Window onto an Island's Past. Mol Biol Evol. 2017 May 1;34(5):1230-1239. doi: 10.1093/molbev/msx082. PMID: 28177087; PMCID: PMC5400395.

Omrak A, Günther T, Valdiosera C, Svensson EM, Malmström H, Kiesewetter H, Aylward W, Storå J, Jakobsson M, Götherström A. Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. Curr Biol. 2016 Jan 25;26(2):270-275. doi: 10.1016/j.cub.2015.12.019. Epub 2015 Dec 31. PMID: 26748850.

Ong WY, Farooqui T, Kokotos G, Farooqui AA. Synthetic and natural inhibitors of phospholipases A2: their importance for understanding and treatment of neurological disorders. ACS Chem Neurosci. 2015 Jun 17;6(6):814-31. doi: 10.1021/acschemneuro.5b00073. Epub 2015 May 1. PMID: 25891385.

Otten AB, Smeets HJ. Evolutionary defined role of the mitochondrial DNA in fertility, disease and ageing. Hum Reprod Update. 2015 Sep-Oct;21(5):671-89. doi: 10.1093/humupd/dmv024. Epub 2015 May 14. PMID: 25976758.

Pagani L, Lawson DJ, Jagoda E, Mörseburg A, Eriksson A, Mitt M, Clemente F, Hudjashov G, DeGiorgio M, Saag L, Wall JD, Cardona A, Mägi R, Wilson Sayres MA, Kaewert S, Inchley C, Scheib CL, Järve M, Karmin M, Jacobs GS, Antao T, Iliescu FM, Kushniarevich A, Ayub Q, Tyler-Smith C, Xue Y, Yunusbayev B, Tambets K, Mallick CB, Saag L, Pocheshkhova E, Andriadze G, Muller C, Westaway MC, Lambert DM, Zoraqi G, Turdikulova S, Dalimova D, Sabitov Z, Sultana GNN, Lachance J, Tishkoff S, Momynaliev K, Isakova J, Damba LD, Gubina M, Nymadawa P, Evseeva I, Atramentova L, Utevska O, Ricaut FX, Brucato N, Sudoyo H, Letellier T, Cox MP, Barashkov NA, Skaro V, Mulahasanovic L, Primorac D, Sahakyan H, Mormina M, Eichstaedt CA, Lichman DV, Abdullah S, Chaubey G, Wee JTS, Mihailov E, Karunas A, Litvinov S, Khusainova R, Ekomasova N, Akhmetova V, Khidiyatova I, Marjanović D, Yepiskoposyan L, Behar DM, Balanovska E, Metspalu A, Derenko M, Malyarchuk B, Voevoda M, Fedorova SA, Osipova LP, Lahr MM, Gerbault P, Leavesley M, Migliano AB, Petraglia M, Balanovsky O, Khusnutdinova EK, Metspalu E, Thomas MG, Manica A, Nielsen R, Villems R, Willerslev E, Kivisild T, Metspalu M. Genomic analyses inform on migration events during the peopling of Eurasia. Nature. 2016 Oct 13;538(7624):238-242. doi: 10.1038/nature19792. Epub 2016 Sep 21. PMID: 27654910; PMCID: PMC5164938.

Palmieri G, Colombino M, Casula M, Budroni M, Manca A, Sini MC, Lissia A, Stanganelli I, Ascierto PA, Cossu A. Epidemiological and genetic factors underlying melanoma development in Italy. Melanoma Manag. 2015 May;2(2):149-163. doi: 10.2217/mmt.15.12. Epub 2015 May 18. PMID: 30190844; PMCID: PMC6094587.

Panoutsopoulou K, Hatzikotoulas K, Xifara DK, Colonna V, Farmaki AE, Ritchie GR, Southam L, Gilly A, Tachmazidou I, Fatumo S, Matchan A, Rayner NW, Ntalla I, Mezzavilla M, Chen Y, Kiagiadaki C, Zengini E, Mamakou V, Athanasiadis A, Giannakopoulou M, Kariakli VE, Nsubuga RN, Karabarinde A, Sandhu M, McVean G, Tyler-Smith C, Tsafantakis E, Karaleftheri M, Xue Y, Dedoussis G, Zeggini E. Genetic characterization of Greek population isolates reveals strong genetic drift at missense and trait-associated variants. Nat Commun. 2014 Nov 6;5:5345. doi: 10.1038/ncomms6345. PMID: 25373335; PMCID: PMC4242463.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. Ancient admixture in human history. Genetics. 2012 Nov;192(3):1065-93. doi: 10.1534/genetics.112.145037. Epub 2012 Sep 7. PMID: 22960212; PMCID: PMC3522152.

Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006 Dec;2(12):e190. doi: 10.1371/journal.pgen.0020190. PMID: 17194218; PMCID: PMC1713260.

Pellegrini, A. *Il dialetto greco-calabro di Bova.*(Loescher, 1880).

Peltonen L, Palotie A, Lange K. Use of population isolates for mapping complex traits. Nat Rev Genet. 2000 Dec;1(3):182-90. doi: 10.1038/35042049. PMID: 11252747.

Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. Am J Hum Genet. 2012 Aug 10;91(2):275-92. doi: 10.1016/j.ajhg.2012.06.014. PMID: 22883143; PMCID: PMC3415543.

Perićić M, Lauc LB, Klarić IM, Rootsi S, Janićijevic B, Rudan I, Terzić R, Colak I, Kvesić A, Popović D, Sijacki A, Behluli I, Dordevic D, Efremovska L, Bajec DD, Stefanović BD, Villems R, Rudan P. High-resolution phylogenetic analysis of southeastern Europe traces major episodes of paternal gene flow among Slavic populations. Mol Biol Evol. 2005 Oct;22(10):1964-75. doi: 10.1093/molbev/msi185. Epub 2005 Jun 8. PMID: 15944443.

Petraglia MD, Haslam M, Fuller DQ, Boivin N, Clarkson C. Out of Africa: new hypotheses and evidence for the dispersal of Homo sapiens along the Indian Ocean rim. Ann Hum Biol. 2010 Jun;37(3):288-311. doi: 10.3109/03014461003639249. PMID: 20334598.

Picardi A, Gentilucci UV, Bambacioni F, Galati G, Spataro S, Mazzarelli C, D'Avola D, Fiori E, Riva E. Lower schooling, higher hepatitis C virus prevalence in Italy: an association dependent on age. J Clin Virol. 2007 Oct;40(2):168-70. doi: 10.1016/j.jcv.2007.07.010. Epub 2007 Sep 4. PMID: 17804289.

Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 2012;8(11):e1002967. doi: 10.1371/journal.pgen.1002967. Epub 2012 Nov 15. PMID: 23166502; PMCID: PMC3499260.

Pilling LC, Kuo CL, Sicinski K, Tamosauskaite J, Kuchel GA, Harries LW, Herd P, Wallace R, Ferrucci L, Melzer D. Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. Aging (Albany NY). 2017 Dec 6;9(12):2504-2520. doi: 10.18632/aging.101334. PMID: 29227965; PMCID: PMC5764389.

Piras IS, De Montis A, Calò CM, Marini M, Atzori M, Corrias L, Sazzini M, Boattini A, Vona G, Contu L. Genome-wide scan with nearly 700,000 SNPs in two Sardinian sub-populations suggests some regions as candidate targets for positive selection. Eur J Hum Genet. 2012 Nov;20(11):1155-61. doi: 10.1038/ejhg.2012.65. Epub 2012 Apr 25. PMID: 22535185; PMCID: PMC3476709.

Pott, F.A. Altgriechisch in heutigen Kalabrien? in *Philologus*, 244-269 (Dieterich, 1856).

Poznik GD, Henn BM, Yee MC, Sliwerska E, Euskirchen GM, Lin AA, Snyder M, Quintana-Murci L, Kidd JM, Underhill PA, Bustamante CD. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. Science. 2013 Aug 2;341(6145):562-5. doi: 10.1126/science.1237619. PMID: 23908239; PMCID: PMC4032117.

Pritchard JK, Di Rienzo A. Adaptation - not by sweeps alone. Nat Rev Genet. 2010 Oct;11(10):665-7. doi: 10.1038/nrg2880. Epub 2010 Sep 14. PMID: 20838407; PMCID: PMC4652788.

Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol. 2010 Feb 23;20(4):R208-15. doi: 10.1016/j.cub.2009.11.055. PMID: 20178769; PMCID: PMC2994553.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559-75. doi: 10.1086/519795. Epub 2007 Jul 25. PMID: 17701901; PMCID: PMC1950838.

Quagliariello A, De Fanti S, Giuliani C, Abondio P, Serventi P, Sarno S, Sazzini M, Luiselli D. Multiple selective events at the PRDM16 functional pathway shaped adaptation of western European populations to different climate conditions. J Anthropol Sci. 2017 Dec 30;95:235-247. doi: 10.4436/JASS.95011. Epub 2017 Jul 10. PMID: 28696940.

Raber J, Huang Y, Ashford JW. ApoE genotype accounts for the vast majority of AD risk and AD pathology. Neurobiol Aging. 2004 May-Jun;25(5):641-50. doi: 10.1016/j.neurobiolaging.2003.12.023. PMID: 15172743.

Raczy C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, Chuang HY, Källberg M, Kumar SA, Liao A, Little KM, Strömberg MP, Tanner SW. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. Bioinformatics. 2013 Aug 15;29(16):2041-3. doi: 10.1093/bioinformatics/btt314. Epub 2013 Jun 4. PMID: 23736529.

Raichlen DA, Alexander GE. Exercise, APOE genotype, and the evolution of the human lifespan. Trends Neurosci. 2014 May;37(5):247-55. doi: 10.1016/j.tins.2014.03.001. Epub 2014 Mar 30. PMID: 24690272; PMCID: PMC4066890.

Raveane A, Aneli S, Montinaro F, Athanasiadis G, Barlera S, Birolo G, Boncoraglio G, Di Blasio AM, Di Gaetano C, Pagani L, Parolo S, Paschou P, Piazza A, Stamatoyannopoulos G, Angius A, Brucato N, Cucca F, Hellenthal G, Mulas A, Peyret-Guzzon M, Zoledziewska M, Baali A, Bycroft C, Cherkaoui M, Chiaroni J, Di Cristofaro J, Dina C, Dugoujon JM, Galan P, Giemza J, Kivisild T, Mazieres S, Melhaoui M, Metspalu M, Myers S, Pereira L, Ricaut FX, Brisighelli F, Cardinali I, Grugni V, Lancioni H, Pascali VL, Torroni A, Semino O, Matullo G, Achilli A, Olivieri A, Capelli C. Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe. Sci Adv. 2019 Sep 4;5(9):eaaw3492. doi: 10.1126/sciadv.aaw3492. PMID: 31517044; PMCID: PMC6726452.

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PL, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin JJ, Kelso J, Slatkin M, Pääbo S. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 2010 Dec 23;468(7327):1053-60. doi: 10.1038/nature09710. PMID: 21179161; PMCID: PMC4306417.

Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. Nature. 2009 Sep 24;461(7263):489-94. doi: 10.1038/nature08365. PMID: 19779445; PMCID: PMC2842210.

Reinbothe TM, Alkayyali S, Ahlqvist E, Tuomi T, Isomaa B, Lyssenko V, Renström E. The human L-type calcium channel Cav1.3 regulates insulin release and polymorphisms in CACNA1D associate with type 2 diabetes. Diabetologia. 2013 Feb;56(2):340-9. doi: 10.1007/s00125-012-2758-z. Epub 2012 Nov 15. PMID: 23229155.

Revelas M, Thalamuthu A, Oldmeadow C, Evans TJ, Armstrong NJ, Kwok JB, Brodaty H, Schofield PR, Scott RJ, Sachdev PS, Attia JR, Mather KA. Review and meta-analysis of genetic polymorphisms associated with exceptional human longevity. Mech Ageing Dev. 2018 Oct;175:24-34. doi: 10.1016/j.mad.2018.06.002. Epub 2018 Jun 8. PMID: 29890178.

Richards MP, Trinkaus E. Out of Africa: modern human origins special feature: isotopic evidence for the diets of European Neanderthals and early modern humans. Proc Natl Acad Sci U S A. 2009 Sep 22;106(38):16034-9. doi: 10.1073/pnas.0903821106. Epub 2009 Aug 11. PMID: 19706482; PMCID: PMC2752538.

Richter D, Grün R, Joannes-Boyau R, Steele TE, Amani F, Rué M, Fernandes P, Raynal JP, Geraads D, Ben-Ncer A, Hublin JJ, McPherron SP. The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. Nature. 2017 Jun 7;546(7657):293-296. doi: 10.1038/nature22335. PMID: 28593967.

Robertson TA, Dutton NS, Martins RN, Taddei K, Papadimitriou JM. Comparison of astrocytic and myocytic metabolic dysregulation in apolipoprotein E deficient and human apolipoprotein E transgenic mice. Neuroscience. 2000;98(2):353-9. doi: 10.1016/s0306-4522(00)00126-3. PMID: 10854768.

Rohlfs, G. *Scavi linguistici nella Magna Grecia*. (Congedo ed., 1932).

Rosenberg NA, Nordborg M. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat Rev Genet. 2002 May;3(5):380-90. doi: 10.1038/nrg795. PMID: 11988763.

Roses A, Sundseth S, Saunders A, Gottschalk W, Burns D, Lutz M. Understanding the genetics of APOE and TOMM40 and role of mitochondrial structure and function in clinical pharmacology of Alzheimer's disease. Alzheimers Dement. 2016 Jun;12(6):687-94. doi: 10.1016/j.jalz.2016.03.015. Epub 2016 May 4. PMID: 27154058.

Saita D, Ferrarese R, Foglieni C, Esposito A, Canu T, Perani L, Ceresola ER, Visconti L, Burioni R, Clementi M, Canducci F. Adaptive immunity against gut microbiota enhances apoE-mediated immune regulation and reduces atherosclerosis and western-diet-related inflammation. Sci Rep. 2016 Jul 7;6:29353. doi: 10.1038/srep29353. PMID: 27383250; PMCID: PMC4935993.

Sarno S, Boattini A, Carta M, Ferri G, Alù M, Yao DY, Ciani G, Pettener D, Luiselli D. An ancient Mediterranean melting pot: investigating the uniparental genetic structure and population history of sicily and southern Italy. PLoS One. 2014 Apr 30;9(4):e96074. doi: 10.1371/journal.pone.0096074. PMID: 24788788; PMCID: PMC4005757.

Sarno S, Boattini A, Pagani L, Sazzini M, De Fanti S, Quagliariello A, Gnecchi Ruscone GA, Guichard E, Ciani G, Bortolini E, Barbieri C, Cilli E, Petrilli R, Mikerezi I, Sineo L, Vilar M, Wells S, Luiselli D, Pettener D. Ancient and recent admixture layers in Sicily and Southern Italy trace multiple migration routes along the Mediterranean. Sci Rep. 2017 May 16;7(1):1984. doi: 10.1038/s41598-017-01802-4. PMID: 28512355; PMCID: PMC5434004.

Sarno S, Tofanelli S, De Fanti S, Quagliariello A, Bortolini E, Ferri G, Anagnostou P, Brisighelli F, Capelli C, Tagarelli G, Sineo L, Luiselli D, Boattini A, Pettener D. Shared language, diverging genetic histories: high-resolution analysis of Y-chromosome variability in Calabrian and Sicilian Arbereshe. Eur J Hum Genet. 2016 Apr;24(4):600-6. doi: 10.1038/ejhg.2015.138. Epub 2015 Jul 1. PMID: 26130483; PMCID: PMC4929864.

Satizabal CL, Samieri C, Davis-Plourde KL, Voetsch B, Aparicio HJ, Pase MP, Romero JR, Helmer C, Vasan RS, Kase CS, Debette S, Beiser AS, Seshadri S. APOE and the Association of Fatty Acids With the Risk of Stroke, Coronary Heart Disease, and Mortality. Stroke. 2018 Dec;49(12):2822-2829. doi: 10.1161/STROKEAHA.118.022132. PMID: 30571417; PMCID: PMC6310220.

Sazzini M, Gnecchi Ruscone GA, Giuliani C, Sarno S, Quagliariello A, De Fanti S, Boattini A, Gentilini D, Fiorito G, Catanoso M, Boiardi L, Croci S, Macchioni P, Mantovani V, Di Blasio AM, Matullo G, Salvarani C, Franceschi C, Pettener D, Garagnani P, Luiselli D. Complex interplay between neutral and adaptive evolution shaped differential genomic background and disease susceptibility along the Italian peninsula. Sci Rep. 2016 Sep 1;6:32513. doi: 10.1038/srep32513. PMID: 27582244; PMCID: PMC5007512.

Sazzini M, Sarno S, Luiselli D. The Mediterranean human population: an anthropological genetics perspective. In: Goffredo S, Dubinsky Z, editors. The Mediterranean Sea: its history and present challenges. New York: Springer; 2014. pp. 529–551.

Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. Nat Genet. 2014 Aug;46(8):919-25. doi: 10.1038/ng.3015. Epub 2014 Jun 22. PMID: 24952747; PMCID: PMC4116295.

Schrider DR, Kern AD. Soft Sweeps Are the Dominant Mode of Adaptation in the Human Genome. Mol Biol Evol. 2017 Aug 1;34(8):1863-1877. doi: 10.1093/molbev/msx154. PMID: 28482049; PMCID: PMC5850737.

Sebastiani P, Gurinovich A, Nygaard M, Sasaki T, Sweigart B, Bae H, Andersen SL, Villa F, Atzmon G, Christensen K, Arai Y, Barzilai N, Puca A, Christiansen L, Hirose N, Perls TT. APOE Alleles and Extreme Human Longevity. J Gerontol A Biol Sci Med Sci. 2019 Jan 1;74(1):44-51. doi: 10.1093/gerona/gly174. PMID: 30060062; PMCID: PMC6298189.

Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorious H, Bedoya G, Ospina J, Ruiz-Linares A, Macedo A, Palha JA, Heutink P, Aulchenko Y, Oostra B, van Duijn C, Jarvelin MR, Varilo T, Peddle L, Rahman P, Piras G, Monne M, Murray S, Galver L, Peltonen L, Sabatti C, Collins A, Freimer N. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. Nat Genet. 2006 May;38(5):556-60. doi: 10.1038/ng1770. Epub 2006 Apr 2. PMID: 16582909.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498-504. doi: 10.1101/gr.1239303. PMID: 14597658; PMCID: PMC403769.

Shtolz N, Mishmar D. The Mitochondrial Genome–on Selective Constraints and Signatures at the Organism, Cell, and Single Mitochondrion Levels. Front Ecol Evol. 2019 Sep 18;7:342. doi: 10.3389/fevo.2019.00342.

Sikora M, Carpenter ML, Moreno-Estrada A, Henn BM, Underhill PA, Sánchez-Quinto F, Zara I, Pitzalis M, Sidore C, Busonero F, Maschio A, Angius A, Jones C, Mendoza-Revilla J, Nekhrizov G, Dimitrova D, Theodossiev N, Harkins TT, Keller A, Maixner F, Zink A, Abecasis G, Sanna S, Cucca F, Bustamante CD. Population genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe. PLoS Genet. 2014 May 8;10(5):e1004353. doi: 10.1371/journal.pgen.1004353. PMID: 24809476; PMCID: PMC4014435.

Silva-Sena GG, Camporez D, Santos LRD, Silva ASD, Sagrillo Pimassoni LH, Tieppo A, Pimentel Batitucci MDC, Morelato RL, Paula F. An association study of FOXO3 variant and longevity. Genet Mol Biol. 2018 Apr./Jun;41(2):386-396. doi: 10.1590/1678-4685-GMB-2017-0169. Epub 2018 Jun 11. PMID: 30088610; PMCID: PMC6082243.

Singh PP, Singh M, Mastana SS. APOE distribution in world populations with new data from India and the UK. Ann Hum Biol. 2006 May-Jun;33(3):279-308. doi: 10.1080/03014460600594513. PMID: 17092867.

Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, Arnaiz O, Awedh MH, Baldock R, Barbiera G, Bardou P, Beck T, Blake A, Bonierbale M, Brookes AJ, Bucci G, Buetti I, Burge S, Cabau C, Carlson JW, Chelala C, Chrysostomou C, Cittaro D, Collin O, Cordova R, Cutts RJ, Dassi E, Di Genova A, Djari A, Esposito A, Estrella H, Eyras E, Fernandez-Banet J, Forbes S, Free RC, Fujisawa T, Gadaleta E, Garcia-Manteiga JM, Goodstein D, Gray K, Guerra-Assunção JA, Haggarty B, Han DJ, Han BW, Harris T, Harshbarger J, Hastings RK, Hayes RD, Hoede C, Hu S, Hu ZL, Hutchins L, Kan Z, Kawaji H, Keliet A, Kerhornou A, Kim S, Kinsella R, Klopp C, Kong L, Lawson D, Lazarevic D, Lee JH, Letellier T, Li CY, Lio P, Liu CJ, Luo J, Maass A, Mariette J, Maurel T, Merella S, Mohamed AM, Moreews F, Nabihoudine I, Ndegwa N, Noirot C, Perez-Llamas C, Primig M, Quattrone A, Quesneville H, Rambaldi D, Reecy J, Riba M, Rosanoff S, Saddiq AA, Salas E, Sallou O, Shepherd R, Simon R, Sperling L, Spooner W, Staines DM, Steinbach D, Stone K, Stupka E, Teague JW, Dayem Ullah AZ, Wang J, Ware D, Wong-Erasmus M, Youens-Clark K, Zadissa A, Zhang SJ, Kasprzyk A. The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic Acids Res. 2015 Jul 1;43(W1):W589-98. doi: 10.1093/nar/gkv350. Epub 2015 Apr 20. PMID: 25897122; PMCID: PMC4489294.

Solé-Morata N, Bertranpetit J, Comas D, Calafell F. Recent radiation of R-M269 and high Y-STR haplotype resemblance confirmed. Ann Hum Genet. 2014 Jul;78(4):253-4. doi: 10.1111/ahg.12066. Epub 2014 May 13. PMID: 24820547.

Solé-Morata N, Villaescusa P, García-Fernández C, Font-Porterias N, Illescas MJ, Valverde L, Tassi F, Ghirotto S, Férec C, Rouault K, Jiménez-Moreno S, Martínez-Jarreta B, Pinheiro MF, Zarrabeitia MT, Carracedo Á, de Pancorbo MM, Calafell F. Analysis of the R1b-DF27 haplogroup shows that a large fraction of Iberian Y-chromosome lineages originated recently in situ. Sci Rep. 2017 Aug 4;7(1):7341. doi: 10.1038/s41598-017-07710-x. PMID: 28779148; PMCID: PMC5544771.

Spinney L. Alzheimer's disease: The forgetting gene. Nature. 2014 Jun 5;510(7503):26-8. doi: 10.1038/510026a. PMID: 24899289.

Squillaci, M. O. When a language becomes old. The case of Calabrian Greek, in *Selected papers from the XV International Conference on Minority Languages* (University of Belgrade, 2017).

Stengård JH, Weiss KM, Sing CF. An ecological study of association between coronary heart disease mortality rates in men and the relative frequencies of common allelic variations in the gene coding for apolipoprotein E. Hum Genet. 1998 Aug;103(2):234-41. doi: 10.1007/s004390050811. PMID: 9760210.

Stewart JB, Chinnery PF. Extreme heterogeneity of human mitochondrial DNA from organelles to populations. Nat Rev Genet. 2021 Feb;22(2):106-118. doi: 10.1038/s41576-020-00284-x. Epub 2020 Sep 28. PMID: 32989265.

Stoneking M, Krause J. Learning about human population history from ancient and modern genomes. Nat Rev Genet. 2011 Aug 18;12(9):603-14. doi: 10.1038/nrg3029. PMID: 21850041.

Subramanian S, Gottschalk WK, Kim SY, Roses AD, Chiba-Falek O. The effects of PPARγ on the regulation of the TOMM40-APOE-C1 genes cluster. Biochim Biophys Acta Mol Basis Dis. 2017 Mar;1863(3):810-816. doi: 10.1016/j.bbadis.2017.01.004. Epub 2017 Jan 6. PMID: 28065845; PMCID: PMC5285471.

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MH, Konkel MK, Malhotra A, Stütz AM, Shi X, Casale FP, Chen J, Hormozdiari F, Dayama G, Chen K, Malig M, Chaisson MJP, Walter K, Meiers S, Kashin S, Garrison E, Auton A, Lam HYK, Mu XJ, Alkan C, Antaki D, Bae T, Cerveira E, Chines P, Chong Z, Clarke L, Dal E, Ding L, Emery S, Fan X, Gujral M, Kahveci F, Kidd JM, Kong Y, Lameijer EW, McCarthy S, Flicek P, Gibbs RA, Marth G, Mason CE, Menelaou A, Muzny DM, Nelson BJ, Noor A, Parrish NF, Pendleton M, Quitadamo A, Raeder B, Schadt EE, Romanovitch M, Schlattl A, Sebra R, Shabalin AA, Untergasser A, Walker JA, Wang M, Yu F, Zhang C, Zhang J, Zheng-Bradley X, Zhou W, Zichner T, Sebat J, Batzer MA, McCarroll SA; 1000 Genomes Project Consortium, Mills RE, Gerstein MB, Bashir A, Stegle O, Devine SE, Lee C, Eichler EE, Korbel JO. An integrated map of structural variation in 2,504 human genomes. Nature. 2015 Oct 1;526(7571):75-81. doi: 10.1038/nature15394. PMID: 26432246; PMCID: PMC4617611.

Suwa G, Asfaw B, Kono RT, Kubo D, Lovejoy CO, White TD. The Ardipithecus ramidus skull and its implications for hominid origins. Science. 2009 Oct 2;326(5949):68e1-7. PMID: 19810194.

Suwalak T, Srisawasdi P, Puangpetch A, Santon S, Koomdee N, Chamnanphon M, Charoenyingwattana A, Chantratita W, Sukasem C. Polymorphisms of the ApoE (Apolipoprotein E) gene and their influence on dyslipidemia in HIV-1-infected individuals. Jpn J Infect Dis. 2015;68(1):5-12. doi: 10.7883/yoken.JJID.2013.190. Epub 2014 Nov 25. PMID: 25420659.

Szpiech ZA, Hernandez RD. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol. 2014 Oct;31(10):2824-7. doi: 10.1093/molbev/msu211. Epub 2014 Jul 10. PMID: 25015648; PMCID: PMC4166924.

Tanaka M, Vedhachalam C, Sakamoto T, Dhanasekaran P, Phillips MC, Lund-Katz S, Saito H. Effect of carboxyl-terminal truncation on structure and lipid interaction of human apolipoprotein E4. Biochemistry. 2006 Apr 4;45(13):4240-7. doi: 10.1021/bi060023b. PMID: 16566598.

Tedla N, Glaros EN, Brunk UT, Jessup W, Garner B. Heterogeneous expression of apolipoprotein-E by human macrophages. Immunology. 2004 Nov;113(3):338-47. doi: 10.1111/j.1365-2567.2004.01972.x. PMID: 15500620; PMCID: PMC1782582.

Teixeira JC, Jacobs GS, Stringer C, Tuke J, Hudjashov G, Purnomo GA, Sudoyo H, Cox MP, Tobler R, Turney CSM, Cooper A, Helgen KM. Widespread Denisovan ancestry in Island Southeast Asia but no evidence of substantial super-archaic hominin admixture. Nat Ecol Evol. 2021 Mar 22. doi: 10.1038/s41559-021-01408-0. Epub ahead of print. PMID: 33753899.

Terhorst J, Kamm JA, Song YS. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat Genet. 2017 Feb;49(2):303-309. doi: 10.1038/ng.3748. Epub 2016 Dec 26. PMID: 28024154; PMCID: PMC5470542.

Thomas PD, Kejariwal A, Guo N, Mi H, Campbell MJ, Muruganujan A, Lazareva-Ulitsky B. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. Nucleic Acids Res. 2006 Jul 1;34(Web Server issue):W645-50. doi: 10.1093/nar/gkl229. PMID: 16912992; PMCID: PMC1538848.

Tian Y, Peng B, Fu X. New ADCY3 Variants Dance in Obesity Etiology. Trends Endocrinol Metab. 2018 Jun;29(6):361-363. doi: 10.1016/j.tem.2018.02.004. Epub 2018 Feb 14. PMID: 29454745.

Tournebize R, Poncet V, Jakobsson M, Vigouroux Y, Manel S. McSwan: A joint site frequency spectrum method to detect and date selective sweeps across multiple population genomes. Mol Ecol Resour. 2019 Jan;19(1):283-295. doi: 10.1111/1755-0998.12957. Epub 2018 Nov 22. PMID: 30358170.

Trombetta B, Cruciani F, Sellitto D, Scozzari R. A new topology of the human Y chromosome haplogroup E1b1 (E-P2) revealed through the use of newly characterized binary polymorphisms. PLoS One. 2011 Jan 6;6(1):e16073. doi: 10.1371/journal.pone.0016073. PMID: 21253605; PMCID: PMC3017091.

Trumper, J. *Geostoria linguistica della Calabria*. (Aracne ed., 2016).

Tsopanakis, A. G. Echi classici nel greco della Magna Grecia, in *Magna Grecia bizantina e tradizione classica: atti del XVII convegno di studi sulla Magna Grecia* (Ist. Storia Arch. Magna Grecia, 1977).

Turchi C, Buscemi L, Previderè C, Grignani P, Brandstätter A, Achilli A, Parson W, Tagliabracci A; Ge.F.I. Group. Italian mitochondrial DNA database: results of a collaborative exercise and proficiency testing. Int J Legal Med. 2008 May;122(3):199-204. doi: 10.1007/s00414-007-0207-1. Epub 2007 Oct 20. PMID: 17952451.

Underhill PA, Kivisild T. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. Annu Rev Genet. 2007;41:539-64. doi: 10.1146/annurev.genet.41.110306.130407. PMID: 18076332.

Underhill PA, Poznik GD, Rootsi S, Järve M, Lin AA, Wang J, Passarelli B, Kanbar J, Myres NM, King RJ, Di Cristofaro J, Sahakyan H, Behar DM, Kushniarevich A, Sarac J, Saric T, Rudan P, Pathak AK, Chaubey G, Grugni V, Semino O, Yepiskoposyan L, Bahmanimehr A, Farjadian S, Balanovsky O, Khusnutdinova EK, Herrera RJ, Chiaroni J, Bustamante CD, Quake SR, Kivisild T, Villems R. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. Eur J Hum Genet. 2015 Jan;23(1):124-31. doi: 10.1038/ejhg.2014.50. Epub 2014 Mar 26. PMID: 24667786; PMCID: PMC4266736.

Valverde L, Illescas MJ, Villaescusa P, Gotor AM, García A, Cardoso S, Algorta J, Catarino S, Rouault K, Férec C, Hardiman O, Zarrabeitia M, Jiménez S, Pinheiro MF, Jarreta BM, Olofsson J, Morling N, de Pancorbo MM. New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia. Eur J Hum Genet. 2016 Mar;24(3):437-41. doi: 10.1038/ejhg.2015.114. Epub 2015 Jun 17. PMID: 26081640; PMCID: PMC4755366.

van Exel E, Koopman JJE, Bodegom DV, Meij JJ, Knijff P, Ziem JB, Finch CE, Westendorp RGJ. Effect of APOE ε4 allele on survival and fertility in an adverse environment. PLoS One. 2017 Jul 6;12(7):e0179497. doi: 10.1371/journal.pone.0179497. PMID: 28683096; PMCID: PMC5500260.

Van Oosten M, Rensen PC, Van Amersfoort ES, Van Eck M, Van Dam AM, Breve JJ, Vogel T, Panet A, Van Berkel TJ, Kuiper J. Apolipoprotein E protects against bacterial lipopolysaccharide-induced lethality. A new therapeutic approach to treat gram-negative sepsis. J Biol Chem. 2001 Mar 23;276(12):8820-4. doi: 10.1074/jbc.M009915200. Epub 2001 Jan 2. PMID: 11136731.

Vasto S, Candore G, Balistreri CR, Caruso M, Colonna-Romano G, Grimaldi MP, Listi F, Nuzzo D, Lio D, Caruso C. Inflammatory networks in ageing, age-related diseases and longevity. Mech Ageing Dev. 2007 Jan;128(1):83-91. doi: 10.1016/j.mad.2006.11.015. Epub 2006 Nov 21. PMID: 17118425.

Vercellotti G, Alciati G, Richards MP, Formicola V. The Late Upper Paleolithic skeleton Villabruna 1 (Italy): a source of data on biology and behavior of a 14,000 year-old hunter. J Anthropol Sci. 2008;86:143-63. PMID: 19934473.

Violi, F. *Storia della Calabria Greca con particolare riguardo all'odierna isola ellenofona*. (Kaleidon ed., 2006).

Wang B, Bao S, Zhang Z, Zhou X, Wang J, Fan Y, Zhang Y, Li Y, Chen L, Jia Y, Li J, Li M, Zheng W, Mu N, Wang L, Yu Z, Wong DSM, Zhang Y, Kwan J, Ka-Fung Mak H, Ambalavanan A, Zhou S, Cai W, Zheng J, Huang S, Rouleau GA, Yang W, Rogaeva E, Ma X, St George-Hyslop P, Chu LW, Song YQ. A rare variant in MLKL confers susceptibility to ApoE ε4-negative Alzheimer's disease in Hong Kong Chinese population. Neurobiol Aging. 2018 Aug;68:160.e1-160.e7. doi: 10.1016/j.neurobiolaging.2018.03.006. Epub 2018 Mar 10. PMID: 29656768.

Wang C, Szpiech ZA, Degnan JH, Jakobsson M, Pemberton TJ, Hardy JA, Singleton AB, Rosenberg NA. Comparing spatial maps of human population-genetic variation using Procrustes analysis. Stat Appl Genet Mol Biol. 2010;9(1):Article 13. doi: 10.2202/1544-6115.1493. Epub 2010 Jan 27. PMID: 20196748; PMCID: PMC2861313.

Wang CC, Jin L, Li H. Natural selection on human Y chromosomes. J Genet Genomics. 2014 Feb 20;41(2):47-52. doi: 10.1016/j.jgg.2014.01.006. Epub 2014 Feb 4. PMID: 24576455.

Weichhart T. mTOR as Regulator of Lifespan, Aging, and Cellular Senescence: A Mini-Review. Gerontology. 2018;64(2):127-134. doi: 10.1159/000484629. Epub 2017 Dec 1. PMID: 29190625; PMCID: PMC6089343.

Weisgraber KH, Innerarity TL, Mahley RW. Abnormal lipoprotein receptor-binding activity of the human E apoprotein due to cysteine-arginine interchange at a single site. J Biol Chem. 1982 Mar 10;257(5):2518-21. PMID: 6277903.

Weisgraber KH. Apolipoprotein E distribution among human plasma lipoproteins: role of the cysteine-arginine interchange at residue 112. J Lipid Res. 1990 Aug;31(8):1503-11. PMID: 2280190.

Wells JCK, Nesse RM, Sear R, Johnstone RA, Stearns SC. Evolutionary public health: introducing the concept. Lancet. 2017 Jul 29;390(10093):500-509. doi: 10.1016/S0140-6736(17)30572-X. Epub 2017 Jul 27. PMID: 28792412.

Wetterau JR, Aggerbeck LP, Rall SC Jr, Weisgraber KH. Human apolipoprotein E3 in aqueous solution. I. Evidence for two structural domains. J Biol Chem. 1988 May 5;263(13):6240-8. PMID: 3360781.

Wilkins JF. Unraveling male and female histories from human genetic data. Curr Opin Genet Dev. 2006 Dec;16(6):611-7. doi: 10.1016/j.gde.2006.10.004. Epub 2006 Oct 25. PMID: 17067791.

Williams B 2nd, Convertino M, Das J, Dokholyan NV. ApoE4-specific Misfolded Intermediate Identified by Molecular Dynamics Simulations. PLoS Comput Biol. 2015 Oct 27;11(10):e1004359. doi: 10.1371/journal.pcbi.1004359. PMID: 26506597; PMCID: PMC4623519.

Wilson C, Wardell MR, Weisgraber KH, Mahley RW, Agard DA. Three-dimensional structure of the LDL receptor-binding domain of human apolipoprotein E. Science. 1991 Jun 28;252(5014):1817-22. doi: 10.1126/science.2063194. PMID: 2063194.

Wilson Sayres MA, Lohmueller KE, Nielsen R. Natural selection reduced diversity on human y chromosomes. PLoS Genet. 2014 Jan;10(1):e1004064. doi: 10.1371/journal.pgen.1004064. Epub 2014 Jan 9. PMID: 24415951; PMCID: PMC3886894.

Wood B. Who is the 'real' *Homo habilis*? Nature. 1987 May 21-27;327(6119):187-8. doi: 10.1038/327187a0. PMID: 3106828.

Wu L, Shen C, Seed Ahmed M, Östenson CG, Gu HF. Adenylate cyclase 3: a new target for anti-obesity drug development. Obes Rev. 2016 Sep;17(9):907-14. doi: 10.1111/obr.12430. Epub 2016 Jun 3. PMID: 27256589.

Yagami T, Yamamoto Y, Koma H. The role of secretory phospholipase A_2 in the central nervous system and neurological diseases. Mol Neurobiol. 2014 Apr;49(2):863-76. doi: 10.1007/s12035-013-8565-9. Epub 2013 Oct 10. PMID: 24113843.

Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, Bhai J, Billis K, Boddu S, Marugán JC, Cummins C, Davidson C, Dodiya K, Fatima R, Gall A, Giron CG, Gil L, Grego T, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, Kay M, Lavidas I, Le T, Lemos D, Martinez JG, Maurel T, McDowall M, McMahon A, Mohanan S, Moore B, Nuhn M, Oheh DN, Parker A, Parton A, Patricio M, Sakthivel MP, Abdul Salam AI, Schmitt BM, Schuilenburg H, Sheppard D, Sycheva M, Szuba M, Taylor K, Thormann A, Threadgold G, Vullo A, Walts B, Winterbottom A, Zadissa A, Chakiachvili M, Flint B, Frankish A, Hunt SE, IIsley G, Kostadima M, Langridge N, Loveland JE, Martin FJ, Morales J, Mudge JM, Muffato M, Perry E, Ruffier M, Trevanion SJ, Cunningham F, Howe KL, Zerbino DR, Flicek P. Ensembl 2020. Nucleic Acids Res. 2020 Jan 8;48(D1):D682-D688. doi: 10.1093/nar/gkz966. PMID: 31691826; PMCID: PMC7145704.

Yates B, Braschi B, Gray KA, Seal RL, Tweedie S, Bruford EA. Genenames.org: the HGNC and VGNC resources in 2017. Nucleic Acids Res. 2017 Jan 4;45(D1):D619-D625. doi: 10.1093/nar/gkw1033. Epub 2016 Oct 30. PMID: 27799471; PMCID: PMC5210531.

Ye K, Gao F, Wang D, Bar-Yosef O, Keinan A. Dietary adaptation of FADS genes in Europe varied across time and geography. Nat Ecol Evol. 2017 May 26;1:167. doi: 10.1038/s41559-017-0167. PMID: 29094686; PMCID: PMC5672832.

Yenicesu GI, Cetin M, Ozdemir O, Cetin A, Ozen F, Yenicesu C, Yildiz C, Kocak N. A prospective case-control study analyzes 12 thrombophilic gene mutations in Turkish couples with recurrent pregnancy loss. Am J Reprod Immunol. 2010 Feb;63(2):126-36. doi: 10.1111/j.1600-0897.2009.00770.x. Epub 2009 Nov 10. PMID: 19906129.

Zeggini E. Using genetically isolated populations to understand the genomic basis of disease. Genome Med. 2014 Oct 17;6(10):83. doi: 10.1186/s13073-014-0083-5. PMID: 25473423; PMCID: PMC4254420.

Zeder MA. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. Proc Natl Acad Sci U S A. 2008 Aug 19;105(33):11597-604. doi: 10.1073/pnas.0801317105. Epub 2008 Aug 12. PMID: 18697943; PMCID: PMC2575338.

Zekraoui L, Lagarde JP, Raisonnier A, Gérard N, Aouizérate A, Lucotte G. High frequency of the apolipoprotein E *4 allele in African pygmies and most of the African populations in sub-Saharan Africa. Hum Biol. 1997 Aug;69(4):575-81. PMID: 9198315.

Zhao S, Liu W, Wang J, Shi J, Sun Y, Wang W, Ning G, Liu R, Hong J. *Akkermansia muciniphila* improves metabolic profiles by reducing inflammation in chow diet-fed mice. J Mol Endocrinol. 2017 Jan;58(1):1-14. doi: 10.1530/JME-16-0054. Epub 2016 Nov 7. PMID: 27821438.

Appendix

Appendix A1. Proportions of APOE isoforms by country.

COUNTRY	f(ε2)	f(ε3)	f(ε4)
Australia	0,00	0,74	0,26
Austria	0,08	0,80	0,12
Bangladesh	0,05	0,86	0,09
Barbados	0,13	0,62	0,26
Belgium	0,08	0,75	0,17
Benin	0,07	0,73	0,20
Brazil	0,00	0,70	0,30
Burkina Faso	0,38	0,50	0,13
Burundi	0,00	0,62	0,39
Cameroun	0,11	0,67	0,22
Canada	0,08	0,80	0,12
Central African Republic	0,06	0,54	0,41
Chad	0,05	0,68	0,27
Chile	0,15	0,70	0,15
China	0,10	0,80	0,10
Colombia	0,09	0,76	0,15
Costa Rica	0,03	0,91	0,07
Cuba	0,07	0,80	0,13
Cyprus	0,05	0,87	0,08
Czech Republic	0,08	0,82	0,10
Denmark	0,08	0,75	0,17
Ecuador	0,00	0,72	0,28
Ethiopia	0,03	0,85	0,12
Finland	0,04	0,73	0,23
France	0,10	0,80	0,10
Gabon	0,12	0,81	0,16
Gambia	0,13	0,59	0,27
Germany	0,07	0,76	0,18
Great Britain	0,08	0,75	0,18
Greece	0,05	0,88	0,07
Greenland	0,02	0,77	0,21
Guinea	0,23	0,60	0,17
Hungary	0,06	0,81	0,13
Iceland	0,07	0,77	0,16
India	0,05	0,87	0,08
Ireland	0,12	0,72	0,16
Israel	0,06	0,80	0,14
Italy	0,05	0,85	0,10
Japan	0,05	0,87	0,08
Java	0,06	0,77	0,17

Kenya	0,09	0,59	0,32
Korea	0,05	0,85	0,10
Madagascar	0,23	0,59	0,18
Malaysia	0,12	0,67	0,23
Mali	0,13	0,72	0,15
Mauritania	0,08	0,83	0,10
Mexico	0,05	0,87	0,09
Mongolia	0,03	0,84	0,13
Morocco	0,07	0,80	0,09
Netherlands	0,10	0,75	0,15
New Guinea	0,14	0,48	0,38
New Zealand	0,12	0,74	0,14
Niger	0,04	0,79	0,17
Nigeria	0,03	0,67	0,03
Norway	0,05	0,75	0,20
Pakistan	0,04	0,88	0,08
Peru	0,01	0,94	0,06
Poland	0,07	0,80	0,13
Polynesia	0,11	0,63	0,26
Portugal	0,06	0,74	0,20
Puerto Rico	0,05	0,85	0,11
Russia	0,02	0,81	0,17
Rwanda	0,10	0,67	0,24
Sao-Tome	0,10	0,65	0,25
Saudi Arabia	0,05	0,89	0,06
Senegal	0,03	0,94	0,03
Sierra Leone	0,14	0,60	0,26
Singapore	0,10	0,83	0,07
South Africa	0,13	0,54	0,32
Spain	0,06	0,80	0,14
Sri Lanka	0,04	0,82	0,13
Sudan	0,08	0,62	0,29
Switzerland	0,07	0,82	0,11
Taiwan	0,09	0,84	0,06
Tanzania	0,14	0,63	0,23
Thailand	0,10	0,82	0,08
Togo	0,31	0,48	0,21
Turkey	0,07	0,87	0,06
Uganda	0,15	0,60	0,25
United States	0,10	0,78	0,12
Vietnam	0,17	0,74	0,09
Zaire	0,04	0,63	0,33

Appendix A2. Identities, archaeological context, country of origin and estimated age of the ancient samples.

Code	Haplotype	Context	Country	Age	Publication
RISE423	E4E4	Middle to Late Bronze Age	Armenia	3400	Allentoft *et al.* 2015
I2424	E3E3	Chalcolithic	Bulgaria	6304	Mathieson *et al.* 2018
I5077	E3E3	Neolithic	Croatia	7026	Mathieson *et al.* 2018
RISE150	E3E3	European Early Bronze Age	Czech Republic	3739	Allentoft *et al.* 2015
I7211	E3E3	Bell Beaker Culture	Czech Republic	4200	Olalde *et al.* 2018
I7249	E3E3	Bell Beaker Culture	Czech Republic	4200	Olalde *et al.* 2018
I7278	E4E4	Bell Beaker Culture	Czech Republic	4200	Olalde *et al.* 2018
I7282	E3E3	Bell Beaker Culture	Czech Republic	4200	Olalde *et al.* 2018
I7286	E3E3	Bell Beaker Culture	Czech Republic	4200	Olalde *et al.* 2018
I7272	E3E3	Corded Ware Culture	Czech Republic	4500	Olalde *et al.* 2018
RISE47	E3E3	Late Neolithic to Bronze Age	Denmark	3362	Allentoft *et al.* 2015
RISE71	E3E3	Late Neolithic to Bronze Age	Denmark	4060	Allentoft *et al.* 2015
RISE61	E2E2	Late Neolithic to Bronze Age	Denmark	4622	Allentoft *et al.* 2015
RISE00	E3E3	Corded Ware Culture	Estonia	4412	Allentoft *et al.* 2015
I1381	E3E3	Bell Beaker Culture	France	3800	Olalde *et al.* 2018
I4304	E3E3	Neolithic	France	5830	Olalde *et al.* 2018
KK1	E3E3	Paleolithic	Georgia	11000	Jones *et al.* 2015
SATP	E3E3	Paleolithic	Georgia	13255	Jones *et al.* 2015
I3601	E3E3	Bell Beaker Culture	Germany	3800	Olalde *et al.* 2018
E09569	E3E3	Bell Beaker Culture	Germany	3819	Olalde *et al.* 2018
I5834	E3E3	Bell Beaker Culture	Germany	4000	Olalde *et al.* 2018
I5836	E3E3	Bell Beaker Culture	Germany	4000	Olalde *et al.* 2018
I0172	E3E3	Neolithic	Germany	5000	Mathieson *et al.* 2015
I0018	E3E3	Linear Pottery Neolithic	Germany	6200	Lipson *et al.* 2017
Stuttgart	E3E3	European Early Neolithic	Germany	9000	Lazaridis *et al.* 2014
I0161.SG	E3E3	Iron Age	Great Britain	1165	Schiffels *et al.* 2016
I0157.SG	E3E3	Iron Age	Great Britain	1232	Schiffels *et al.* 2016
I0769.SG	E4E4	Iron Age	Great Britain	1455	Schiffels *et al.* 2016
I0777.SG	E3E3	Iron Age	Great Britain	1478	Schiffels *et al.* 2016
I0774.SG	E3E3	Iron Age	Great Britain	1483	Schiffels *et al.* 2016
I0773.SG	E3E3	Iron Age	Great Britain	1490	Schiffels *et al.* 2016
6DT18.SG	E3E3	Iron Age	Great Britain	1800	Martiniano *et al.* 2016
6DT22.SG	E3E3	Iron Age	Great Britain	1800	Martiniano *et al.* 2016
6DT3.SG	E4E4	Iron Age	Great Britain	1800	Martiniano *et al.* 2016
I0160.SG	E3E3	Iron Age	Great Britain	1995	Schiffels *et al.* 2016
I2860	E3E3	Bronze Age	Great Britain	2738	Olalde *et al.* 2018
I7579	E3E3	Bronze Age	Great Britain	3135	Olalde *et al.* 2018
I6777	E2E2	Bell Beaker Culture	Great Britain	4000	Olalde *et al.* 2018
I2933	E3E3	Neolithic	Great Britain	4309	Olalde *et al.* 2018
I2691	E3E3	Neolithic	Great Britain	4881	Olalde *et al.* 2018
I6759	E3E3	Neolithic	Great Britain	5500	Olalde *et al.* 2018
I0071	E3E3	Neolithic	Greece	4000	Lazaridis *et al.* 2017
I9005	E4E4	Neolithic	Greece	4000	Lazaridis *et al.* 2017

RISE247	E4E4	Bronze Age	Hungary	3629	Allentoft *et al.* 2015
RISE479	E3E3	Bronze Age	Hungary	3700	Allentoft *et al.* 2015
RISE483	E3E3	Bronze Age	Hungary	3700	Allentoft *et al.* 2015
I7045	E3E3	Bell Beaker Culture	Hungary	4200	Olalde *et al.* 2018
F38.SG	E3E3	Iron Age	Iran	2800	Broushaki *et al.* 2016
WC1.SG	E3E3	Neolithic	Iran	8200	Broushaki *et al.* 2016
AH1.SG	E3E3	Neolithic	Iran	10000	Broushaki *et al.* 2016
AH2.SG	E3E3	Neolithic	Iran	10000	Broushaki *et al.* 2016
AH4.SG	E3E3	Neolithic	Iran	10000	Broushaki *et al.* 2016
rath3.SG	E3E3	Bronze Age	Ireland	3585	Cassidy *et al.* 2016
rath1.SG	E3E3	Bronze Age	Ireland	3906	Cassidy *et al.* 2016
bally.SG	E3E3	Neolithic	Ireland	5000	Cassidy *et al.* 2016
Iceman	E3E3	European Chalcolithic	Italy	5244	Keller *et al.* 2012
RISE487	E3E3	European Neolithic	Italy	5245	Allentoft *et al.* 2015
I4437	E3E3	Neolithic	Latvia	6148	Mathieson *et al.* 2018
I4441	E3E3	Neolithic	Latvia	6725	Mathieson *et al.* 2018
I4552	E3E3	Neolithic	Latvia	7300	Mathieson *et al.* 2018
I4596	E4E4	Neolithic	Latvia	7900	Mathieson *et al.* 2018
Loschbour	E3E3	Mesolithic	Luxembourg	8000	Lazaridis *et al.* 2014
I6531	E3E3	Bronze Age	Poland	3755	Olalde *et al.* 2018
RISE109	E3E3	European Early Bronze Age	Poland	3813	Allentoft *et al.* 2015
RISE577	E3E3	European Early Bronze Age	Poland	4000	Allentoft *et al.* 2015
RISE500	E3E3	Middle to Late Bronze Age	Russia	3550	Allentoft *et al.* 2015
RISE505	E4E4	Middle to Late Bronze Age	Russia	3636	Allentoft *et al.* 2015
RISE395	E4E4	Middle to Late Bronze Age	Russia	3808	Allentoft *et al.* 2015
RISE392	E3E3	Middle to Late Bronze Age	Russia	3961	Allentoft *et al.* 2015
RISE386	E4E4	Middle to Late Bronze Age	Russia	4122	Allentoft *et al.* 2015
RISE552	E3E3	Early to Middle Bronze Age	Russia	4446	Allentoft *et al.* 2015
RISE548	E3E3	Early to Middle Bronze Age	Russia	4650	Allentoft *et al.* 2015
RISE547	E2E2	Early to Middle Bronze Age	Russia	4711	Allentoft *et al.* 2015
RISE509	E4E4	Early to Middle Bronze Age	Russia	4732	Allentoft *et al.* 2015
RISE511	E3E3	Early to Middle Bronze Age	Russia	4744	Allentoft *et al.* 2015
I0231	E3E3	Early to Middle Bronze Age	Russia	4792	Mathieson *et al.* 2015
RISE550	E3E3	Early to Middle Bronze Age	Russia	4935	Allentoft *et al.* 2015
MA1	E3E3	Paleolithic	Russia	24300	Raghavan *et al.* 2013
I4873	E3E3	Neolithic	Serbia	7872	Mathieson *et al.* 2018
I4876	E3E3	Neolithic	Serbia	8455	Mathieson *et al.* 2018
I4874	E3E3	Neolithic	Serbia	8506	Mathieson *et al.* 2018
Ust_Ishim	E3E4	Paleolithic	Siberia	42000	Fu *et al.* 2014
I5838	E3E3	Copper Age	Spain	4500	Lipson *et al.* 2017
ATP2	E3E3	Chalcolithic	Spain	4800	Gunther *et al.* 2014
CB13	E3E3	Neolithic	Spain	7000	Olalde *et al.* 2015
LaBrana1	E3E3	Mesolithic	Spain	7000	Olalde *et al.* 2014
I0412	E3E3	European Neolithic	Spain	7144	Mathieson *et al.* 2015
RISE97	E3E3	Late Neolithic to Bronze Age	Sweden	3905	Allentoft *et al.* 2015
RISE98	E4E4	Late Neolithic to Bronze Age	Sweden	4104	Allentoft *et al.* 2015

RISE94	E3E3	Late Neolithic to Bronze Age	Sweden	4497	Allentoft *et al.* 2015
Motala12	E3E3	Mesolithic	Sweden	7000	Mathieson *et al.* 2015
Bichon	E3E3	Paleolithic	Switzerland	15000	Hofmanova *et al.* 2016
Bar8.SG	E4E4	Neolithic	Turkey	8000	Hofmanova *et al.* 2016
I0708	E3E3	Anatolian Neolithic	Turkey	8000	Mathieson *et al.* 2015
I0745	E3E3	Anatolian Neolithic	Turkey	8000	Mathieson *et al.* 2015
Bon002	E3E3	Neolithic	Turkey	10000	Kilinc *et al.* 2016
I5875	E3E3	European Neolithic	Ukraine	7100	Mathieson *et al.* 2018

Appendix A3. Top 1% variants characterizing the ASPR cluster.

SNP	CHR	POS	SNP(cont.)	CHR(cont.)	POS(cont.)	SNP(cont.)	CHR(cont.)	POS(cont.)
rs982510	1	14160754	rs6882483	5	147382111	rs2501682	10	6959853
rs11578558	1	31267335	rs11135102	5	159865579	rs2580894	10	12995888
rs3886437	1	32424756	rs158851	5	163085666	rs11258082	10	12999115
rs17556209	1	39163072	rs158854	5	163086195	rs10508452	10	13005288
rs6665159	1	39172288	rs1422418	5	167847974	rs10906276	10	13028374
rs17579464	1	58369662	rs1433021	5	171960073	rs7087980	10	28469074
rs7541833	1	69184038	rs11966212	6	1056155	rs10740784	10	28492334
rs417181	1	79710545	rs2317906	6	1373145	rs7919947	10	28500786
rs11161464	1	85298111	rs2765368	6	7013556	rs3121313	10	43440869
rs1409215	1	85310832	rs1286005	6	7046452	rs3004255	10	43448012
rs1325272	1	85312366	rs2294425	6	11745652	rs10824829	10	54642295
rs11161574	1	85727044	rs12211506	6	11762311	rs6479995	10	66407750
rs969788	1	86916856	rs2235396	6	11762961	rs3858156	10	71617387
rs548649	1	87380631	rs10484363	6	16449304	rs4076957	10	71617584
rs11577139	1	98666125	rs9477384	6	17318116	rs4746065	10	72513957
rs992708	1	99209727	rs12526082	6	17330855	rs749128	10	78155142
rs11166108	1	99220941	rs13197759	6	17343655	rs1873465	10	78189959
rs11166113	1	99230383	rs6456825	6	28522695	rs7076791	10	80066149
rs10493914	1	99960670	rs450630	6	28542424	rs1936216	10	91716028
rs6577380	1	99963251	rs380914	6	28569949	rs10882520	10	96799688
rs12060427	1	102922025	rs9501180	6	28579471	rs12413444	10	1,04E+08
rs3806233	1	153600778	rs732395	6	28616327	rs10883701	10	1,04E+08
rs7532286	1	154750816	rs422331	6	28810817	rs11244948	10	1,28E+08
rs6426878	1	164649075	rs377743	6	32211065	rs431690	10	1,32E+08
rs10919325	1	170171083	rs6910071	6	32282854	rs410411	10	1,32E+08
rs10919326	1	170171258	rs9277378	6	33050279	rs3858307	10	1,32E+08
rs10158534	1	170172417	rs9277386	6	33050499	rs4750838	10	1,33E+08
rs1330385	1	170174665	rs9277396	6	33051139	rs1703688	10	1,33E+08
rs4656206	1	170179064	rs9277426	6	33051910	rs1703687	10	1,33E+08
rs7545084	1	170186610	rs9277464	6	33053352	rs1560797	10	1,33E+08
rs6657002	1	170187309	rs9277469	6	33053468	rs12778246	10	1,34E+08
rs10913625	1	171732378	rs9277471	6	33053682	rs2492643	10	1,34E+08
rs500894	1	175837332	rs9277489	6	33053942	rs10833460	11	3225964
rs17186722	1	207572294	rs9277533	6	33054721	rs7930238	11	6324061
rs7413508	1	210839705	rs9277546	6	33055346	rs4757851	11	19843968
rs10863850	1	210848224	rs9277554	6	33055538	rs10833482	11	21274683
rs10494953	1	212764085	rs9277555	6	33055605	rs769149	11	35532452
rs6682064	1	212767840	rs3117229	6	33056069	rs2291444	11	47179143
rs10746435	1	212776328	rs3117228	6	33056435	rs3740690	11	47188592
rs7521489	1	215935660	rs9277557	6	33056694	rs4522197	11	69549801
rs10779438	1	221677672	rs9277565	6	33056897	rs10796842	11	69559418
rs10863628	1	221682007	rs3130188	6	33057176	rs1112008	11	69566488
rs1858566	1	221690507	rs3091282	6	33057198	rs11263551	11	69570033
rs10863632	1	221696457	rs3117226	6	33057659	rs11233261	11	82401684

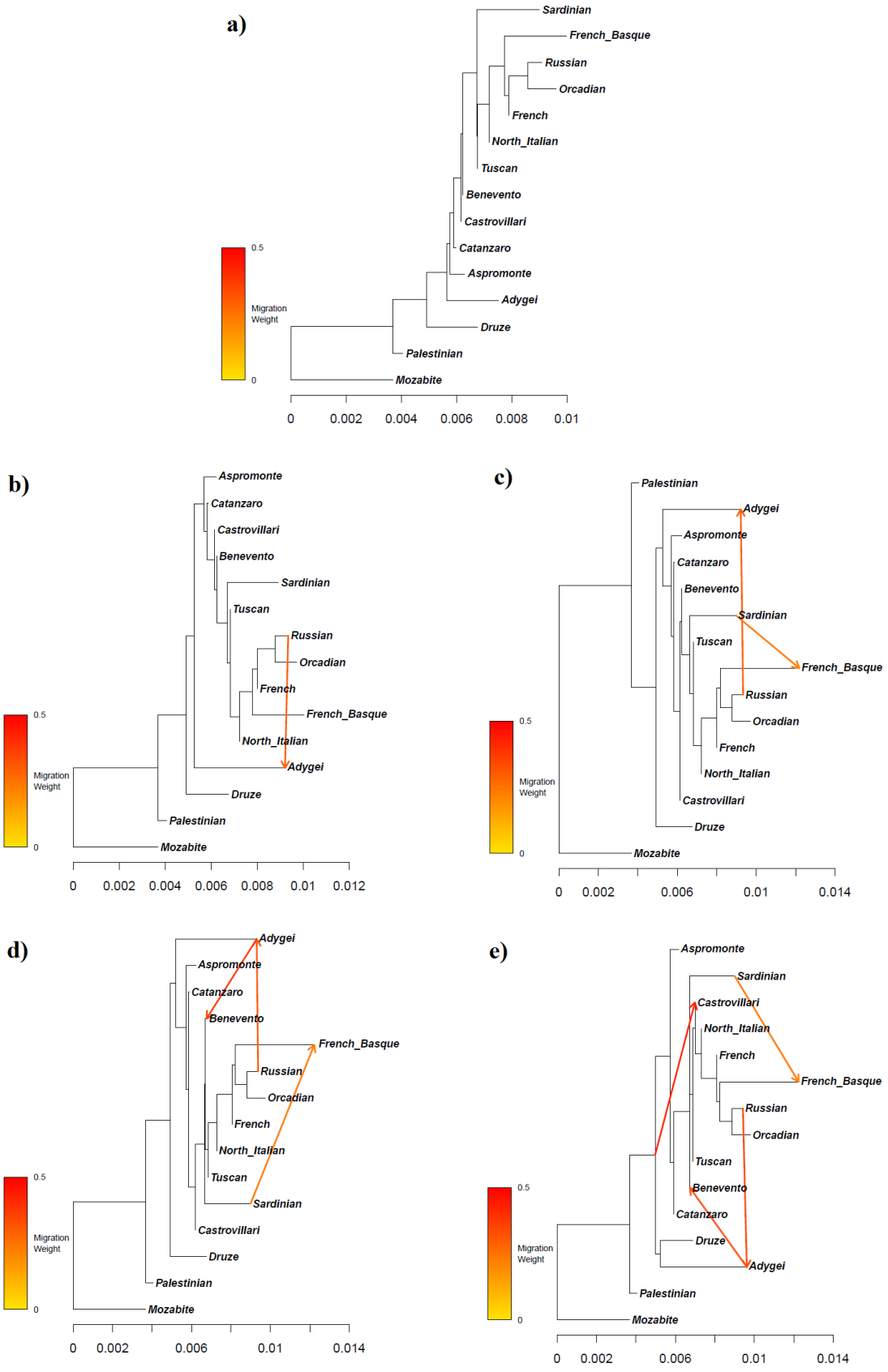
rs2889911	1	221710721	rs3117225	6	33057711	rs7940613	11	91714466
rs1026514	1	221715719	rs3097652	6	33057835	rs2726355	11	99455061
rs10799576	1	224644549	rs1367730	6	33058114	rs7938043	11	1,02E+08
rs4653939	1	228494696	rs3128972	6	33058774	rs2029302	11	1,07E+08
rs4653942	1	228505204	rs2179919	6	33059262	rs3802871	11	1,18E+08
rs728727	1	228643447	rs3128917	6	33059996	rs10892246	11	1,18E+08
rs3213495	1	230416038	rs3117222	6	33060949	rs3133224	11	1,21E+08
rs7541947	1	233161822	rs3130190	6	33061690	rs11218347	11	1,21E+08
rs2463198	1	236533247	rs3130191	6	33061871	rs11218349	11	1,21E+08
rs2066380	1	240901508	rs3130198	6	33063931	rs11224046	11	1,35E+08
rs1591217	1	242927370	rs3117213	6	33064605	rs987181	12	15220406
rs1320334	2	692784	rs2395319	6	33067211	rs12424045	12	15233065
rs11684944	2	3633917	rs3117239	6	33071777	rs2160532	12	15233495
rs4849953	2	3642165	rs2064479	6	33072240	rs10846185	12	15659500
rs13019382	2	3763411	rs2064478	6	33072266	rs2111173	12	15666517
rs7558456	2	3767991	rs2064476	6	33073322	rs17418255	12	21510541
rs2603148	2	3772338	rs3117234	6	33073984	rs1392196	12	25024151
rs1465323	2	10567895	rs3128928	6	33074612	rs2220130	12	30015289
rs4669783	2	12028293	rs3117231	6	33074908	rs813784	12	32438757
rs10181834	2	14702349	rs3117230	6	33075635	rs12814239	12	57569478
rs2675899	2	14705390	rs3128930	6	33075666	rs1504465	12	60676912
rs6734939	2	14708545	rs2295261	6	47236880	rs12826072	12	62701966
rs13414125	2	14711241	rs11758366	6	47253631	rs3825243	12	69750839
rs17348835	2	14712421	rs621387	6	56289743	rs10506670	12	75656510
rs13420936	2	14712895	rs608383	6	56290289	rs7300004	12	96109803
rs3769598	2	32679732	rs594008	6	56291176	rs10777721	12	96116953
rs3731579	2	32735881	rs16900290	6	62588444	rs4762610	12	96120071
rs11683905	2	32751182	rs2061578	6	66669464	rs4128893	12	96123609
rs12466713	2	32751688	rs1037882	6	66672165	rs11108225	12	96130713
rs7587877	2	36902056	rs6455065	6	66701759	rs17026787	12	97729032
rs7602835	2	36902070	rs6455069	6	66722969	rs1148408	12	1,07E+08
rs17490768	2	36926371	rs6919455	6	67621368	rs233714	12	1,13E+08
rs7571216	2	40504788	rs12198920	6	71307347	rs17619048	12	1,18E+08
rs11683295	2	44316109	rs6916197	6	74678421	rs7321901	13	20847138
rs17425376	2	44327351	rs4421160	6	74682273	rs9552555	13	22523977
rs1947195	2	46085325	rs7763484	6	74734641	rs3117951	13	22525954
rs12713280	2	55081376	rs10806180	6	80618523	rs11148372	13	22788665
rs6758100	2	59114739	rs10455125	6	80683626	rs9580631	13	24030514
rs2058563	2	59119018	rs770898	6	83065888	rs6563133	13	36318869
rs1362479	2	67167431	rs770911	6	83074365	rs4943543	13	38474387
rs4854502	2	68965078	rs2875533	6	88030587	rs9526192	13	47109885
rs1558899	2	71309586	rs9294379	6	88152535	rs9534424	13	47118711
rs6744843	2	71320364	rs157695	6	91218473	rs9301571	13	87227991
rs6546670	2	71327146	rs6454818	6	91319984	rs9522944	13	91015462
rs7558412	2	71907930	rs9362761	6	91345657	rs4512968	13	91020626
rs6746541	2	86001628	rs9372328	6	112711218	rs9524955	13	96032939
rs1048084	2	86002303	rs7756462	6	112755156	rs1995372	13	97764447

rs1356626	2	86022788	rs9487914	6	112757968	rs4771983	13	97798733
rs1878902	2	86023425	rs1276639	6	112760200	rs9556683	13	97805817
rs12621362	2	88803885	rs1276628	6	112773200	rs17182592	14	22032003
rs4449134	2	88846961	rs6934797	6	112775377	rs1959179	14	33530555
rs2889323	2	105039687	rs4256466	6	112781952	rs4243636	14	40110065
rs6743263	2	105105148	rs4945516	6	114791568	rs10483520	14	40183285
rs12620665	2	105114879	rs6568841	6	114800340	rs2753404	14	68912686
rs1439671	2	105116184	rs5002415	6	118281878	rs7147798	14	98230632
rs10192506	2	127319936	rs4946312	6	118310478	rs1462268	14	98237766
rs10196438	2	127321457	rs9387532	6	118320108	rs7155566	14	98247629
rs1025788	2	127329197	rs2327534	6	135071826	rs7160746	14	98249964
rs12621600	2	144266489	rs4895561	6	139808000	rs4906321	14	1,04E+08
rs13419857	2	158802021	rs9495497	6	139811689	rs12900257	15	22784095
rs1869543	2	165804660	rs661704	6	141174222	rs2341140	15	25844124
rs10205696	2	165805055	rs365515	6	145802833	rs2076744	15	25926179
rs2390732	2	169605967	rs398060	6	145841515	rs4906619	15	25946781
rs10210408	2	170046850	rs9322447	6	154424321	rs2220545	15	29339499
rs2389589	2	170048929	rs184074	6	157216039	rs4603522	15	29339947
rs6433602	2	177519429	rs877676	6	158102714	rs657473	15	33880432
rs6740315	2	177520106	rs16900421	6	158193244	rs688939	15	33880493
rs4893911	2	177522377	rs6914575	6	160898831	rs12592542	15	33985818
rs1562611	2	177532328	rs9346817	6	160902269	rs11072635	15	34006652
rs13019019	2	177536629	rs9355803	6	160903178	rs8042848	15	34008989
rs1554011	2	177551886	rs9355804	6	160903622	rs12910679	15	34021963
rs10490684	2	182451737	rs10945673	6	160910899	rs12906810	15	34022225
rs833125	2	183182776	rs2322278	6	164523901	rs8036377	15	34027993
rs288316	2	183695449	rs526725	6	165496918	rs1017843	15	40528155
rs17800281	2	191193097	rs544603	6	165501693	rs2617235	15	42758738
rs17626732	2	195808307	rs12703010	7	1680165	rs2617237	15	42765624
rs7578216	2	195832688	rs10271818	7	10137638	rs17766258	15	42785137
rs1304358	2	198969583	rs10488475	7	17813444	rs2899051	15	42786160
rs13012540	2	200056608	rs4291154	7	17929655	rs11636209	15	42788107
rs4675285	2	203465357	rs1229680	7	26462580	rs17709867	15	42800767
rs6760289	2	206346002	rs2525764	7	26478596	rs1060291	15	42820589
rs13007621	2	207393424	rs2698727	7	26482974	rs8033412	15	42822289
rs13004168	2	207492886	rs1561308	7	38582081	rs10454039	15	42832534
rs1558440	2	210618870	rs2686817	7	47968927	rs10851410	15	42844788
rs4585026	2	215275625	rs6955144	7	67889919	rs12324717	15	42846553
rs12472818	2	218358312	rs7785025	7	75628101	rs10851411	15	42859235
rs2373067	2	218360362	rs3801472	7	75629778	rs950027	15	45801035
rs2373072	2	218374066	rs6978677	7	75636240	rs2136897	15	47921514
rs12473089	2	220656956	rs1104878	7	75685593	rs1224674	15	47944152
rs10193131	2	221016421	rs10476	7	75696420	rs1369637	15	47944827
rs1463991	2	221703746	rs6953665	7	75769049	rs2445762	15	51617708
rs4663609	2	236468798	rs10952840	7	75772878	rs7163193	15	51644775
rs12471077	2	238351337	rs10277259	7	75802674	rs17370987	15	69776684
rs11124234	2	239841003	rs2009836	7	75931160	rs7176919	15	69778446

rs2729139	3	1128074	rs3779320	7	77691587	rs1504632	15	69778982
rs2729319	3	2064647	rs7458619	7	79857872	rs9920407	15	69781764
rs13078449	3	9162915	rs2888556	7	86875826	rs5003866	15	69782205
rs6781236	3	9163206	rs1990171	7	86885119	rs6494798	15	69790976
rs7635873	3	16667968	rs1358064	7	86941669	rs12903497	15	69793953
rs7647846	3	16675127	rs1527152	7	113956930	rs11634551	15	69808890
rs11709636	3	19341221	rs10486026	7	114080435	rs6494801	15	69822411
rs7640623	3	19408549	rs36929	7	123505744	rs12593658	15	69830008
rs4470447	3	24836454	rs339092	7	128357300	rs4777151	15	69830696
rs9310820	3	26895022	rs339072	7	128364080	rs7183850	15	69858562
rs12493245	3	27797810	rs731770	7	128366485	rs7403706	15	80861365
rs4680845	3	27799034	rs157925	7	130580165	rs7166858	15	87701586
rs17020288	3	27801959	rs2774960	7	138602417	rs7170640	15	96014096
rs2371150	3	27806147	rs975494	7	142222582	rs10162686	15	96015083
rs9872426	3	27818453	rs2855918	7	142237788	rs12595031	15	96024055
rs2371996	3	30200658	rs2855896	7	142258330	rs1383146	15	98678205
rs17054246	3	54693752	rs1881902	7	143825428	rs3784606	15	99458481
rs11914780	3	54781631	rs2961160	7	143826603	rs10048024	15	99460862
rs1997463	3	54807240	rs2951324	7	143834586	rs11855311	15	1E+08
rs4679524	3	60292606	rs1208355	7	143858276	rs12598611	16	7454984
rs12633258	3	60664643	rs1635884	7	143864670	rs7192216	16	10814546
rs4688298	3	62460883	rs7804359	7	143866532	rs737008	16	11374866
rs12489309	3	68565190	rs2600487	8	1016903	rs949429	16	11376029
rs606028	3	107308977	rs2627283	8	2781277	rs398364	16	11387164
rs658595	3	107371482	rs7013590	8	3969537	rs12448960	16	11419378
rs11710737	3	107464170	rs1946161	8	5165806	rs11074961	16	11422245
rs12637512	3	111028725	rs10092857	8	5165936	rs12446881	16	11435855
rs10804486	3	111039820	rs11985327	8	5172635	rs10852276	16	25040158
rs1873105	3	111042752	rs6992382	8	5177400	rs7498149	16	25043569
rs1552540	3	111054726	rs1895011	8	5182239	rs4638587	16	26962996
rs7633973	3	111087564	rs7840725	8	5186047	rs7204135	16	50626195
rs13072691	3	136425600	rs1592623	8	5189722	rs7184292	16	52919566
rs323618	3	151960936	rs4875157	8	5189771	rs12708901	16	53119733
rs17501198	3	161587855	rs1420832	8	5192203	rs1008400	16	54119892
rs9878522	3	165125159	rs1030080	8	5194048	rs10492872	16	54120330
rs5018958	3	174062834	rs7845635	8	5201210	rs11646512	16	54120952
rs7610632	3	174088909	rs7815392	8	5201718	rs11863548	16	54123512
rs9863476	3	174142441	rs10104026	8	5203682	rs8049247	16	68712730
rs9290501	3	174198101	rs7829940	8	5204568	rs100912	16	75519917
rs16831521	3	180076400	rs7843555	8	5206751	rs2013847	16	84587960
rs6771207	3	182896412	rs11988885	8	5207530	rs34152738	16	86611391
rs35070271	3	182897920	rs6559005	8	5209980	rs34142458	17	2792530
rs11708313	3	182903072	rs11777077	8	5213257	rs2953452	17	10070772
rs17217776	3	185295704	rs10107399	8	5214904	rs2692334	17	10824615
rs2886928	3	192835072	rs10111616	8	5216264	rs11653983	17	18859611
rs9842842	3	192870449	rs2816522	8	5782377	rs429276	17	29984108
rs9837828	3	192886861	rs1383412	8	14807636	rs9901111	17	31046830

rs4677758	3	194551215	rs13256965	8	19962962	rs7207449	17	52619915
rs2857844	4	3029588	rs11204099	8	20071503	rs8065148	17	52640841
rs7441080	4	13788219	rs952149	8	20091363	rs2362726	17	66259285
rs6847559	4	25837310	rs2198011	8	20492735	rs17699266	17	67976326
rs7664033	4	29691666	rs3176300	8	21904891	rs12938367	17	69080285
rs16885694	4	31700291	rs12549516	8	21911601	rs4793289	17	70378730
rs12509470	4	45192865	rs4732844	8	28104473	rs8073890	17	73256533
rs335286	4	62368762	rs11996657	8	29451004	rs9302876	17	76479389
rs1542096	4	77204500	rs4732986	8	29485927	rs7225663	17	77391145
rs6816852	4	85199286	rs4732987	8	29494941	rs7405583	17	80938565
rs1442164	4	85274948	rs12544701	8	35375058	rs9911830	17	80943685
rs7658584	4	89096641	rs602382	8	36104136	rs628229	18	4249393
rs12641788	4	89222164	rs1413875	8	36114239	rs206452	18	10437860
rs10001661	4	89249372	rs17662607	8	41674525	rs1380085	18	30028877
rs4699773	4	101016395	rs2385320	8	50111746	rs958127	18	31497346
rs6532939	4	102399209	rs1519835	8	50252962	rs505601	18	34280131
rs1863575	4	102407109	rs2919308	8	62195086	rs12961192	18	37485412
rs7695470	4	102701262	rs2919303	8	62214493	rs8099208	18	40516148
rs2522515	4	105619927	rs2978558	8	62214626	rs9947826	18	43899499
rs2725762	4	105700542	rs10086214	8	69755610	rs2586776	18	48331524
rs12108563	4	109684830	rs4738080	8	71129813	rs9947544	18	50334201
rs1523504	4	125318885	rs2926702	8	71167994	rs11877576	18	57084974
rs7655016	4	125361767	rs16936880	8	71200109	rs949292	18	58168807
rs11099001	4	130077663	rs16937001	8	71340989	rs4419155	18	65105996
rs13132155	4	130083679	rs16937019	8	71351158	rs7577	18	72188347
rs11726808	4	130089052	rs13259607	8	71352047	rs2849782	18	73341372
rs13109471	4	131062487	rs13259910	8	71352148	rs1789250	18	76715856
rs17050881	4	131103736	rs2383902	8	74268575	rs9967005	18	77882827
rs4610390	4	136778939	rs13269099	8	75016199	rs1064059	18	77893683
rs4318703	4	136910271	rs7817382	8	89292303	rs3087701	18	77915906
rs11723419	4	137213250	rs7835845	8	89338155	rs12201	18	77916015
rs11726692	4	141920037	rs1504801	8	89383727	rs3816322	19	335946
rs12643170	4	148928372	rs821109	8	89499611	rs12460780	19	3645947
rs156500	4	155677315	rs6469448	8	89514775	rs12984273	19	3647901
rs6831272	4	158115848	rs16884419	8	89579649	rs1613070	19	6412929
rs9993365	4	158186342	rs1076030	8	103676345	rs2189784	19	15959200
rs10008950	4	158221234	rs2099408	8	103680416	rs7258885	19	28424804
rs4475186	4	158242540	rs930266	8	103683191	rs3786527	19	41109456
rs10428456	4	160186284	rs6981913	8	128775635	rs2303728	19	41114065
rs4691749	4	162255173	rs17713284	8	134230644	rs2604913	19	41218611
rs7687480	4	166734530	rs7822022	8	138810922	rs2561530	19	41224314
rs736591	4	179374122	rs6577842	8	138813360	rs3745290	19	41890003
rs6815916	4	182553043	rs1511852	8	138843225	rs17318596	19	41937095
rs17212276	4	182571325	rs12545671	8	138844671	rs10409222	19	46379647
rs11727308	4	182572180	rs1551806	8	140922699	rs1152239	19	50420671
rs12502711	4	184866656	rs6583595	8	143249252	rs10407445	19	57649962
rs4862535	4	186397351	rs7010259	8	145846570	rs6053635	20	625267

rs4861680	4	186721885	rs10975792	9	680323	rs4815597	20	3697618
rs10155470	4	186882290	rs10975803	9	680585	rs6052038	20	3707715
rs4866550	5	3308312	rs10815984	9	8884844	rs7270603	20	12007033
rs7722069	5	3623182	rs10815986	9	8885169	rs6043433	20	15659486
rs6883836	5	3670973	rs10815989	9	8891463	rs6129359	20	38168977
rs6889720	5	13817881	rs12235745	9	12540358	rs7267810	20	41061083
rs1445815	5	13829042	rs10810127	9	14294350	rs2868860	20	45674409
rs4547927	5	13834445	rs10961730	9	14819813	rs11086191	20	45680445
rs17259353	5	21699126	rs10810584	9	16658915	rs1206816	20	45704309
rs1879480	5	23730041	rs11794832	9	18910581	rs2298002	20	45726192
rs4701393	5	23755699	rs11788139	9	21651620	rs6012188	20	46077309
rs1520985	5	23758878	rs7846858	9	22836549	rs2426407	20	50652937
rs6871540	5	23861064	rs7857348	9	22856899	rs6089291	20	60467860
rs6862995	5	27824645	rs9969855	9	22857606	rs2822589	21	15711540
rs10075707	5	31179728	rs996367	9	22906349	rs10439652	21	15711772
rs6890687	5	31184402	rs10965520	9	22916961	rs2178907	21	15713544
rs12055267	5	31204843	rs10965529	9	22943509	rs9983082	21	15715197
rs10038069	5	31215770	rs10118162	9	26572537	rs2822602	21	15716285
rs4596398	5	31223681	rs7864522	9	38741712	rs1888397	21	16059028
rs4866808	5	43180535	rs9314845	9	71643131	rs1571691	21	32444209
rs7730004	5	43191033	rs815845	9	84216090	rs1892695	21	32444581
rs10074873	5	43191823	rs2780982	9	86232809	rs2833558	21	33286076
rs1303763	5	51877090	rs1990422	9	122244046	rs1005694	21	37443658
rs702531	5	58267723	rs10491524	9	122257254	rs2835862	21	39020935
rs16894413	5	65127219	rs7043386	9	134866354	rs13052909	21	42224120
rs7701312	5	67012900	rs7025089	9	134881443	rs2269125	21	44072445
rs2406905	5	81260121	rs4363310	9	134890430	rs855072	22	20221844
rs2560266	5	94726879	rs7851262	9	134944428	rs2207361	22	27825804
rs13155917	5	127856494	rs28380074	9	139225139	rs6006267	22	30216982
rs4976524	5	136035580	rs865289	10	595504	rs5997787	22	31254954
rs31746	5	140419991	rs11814112	10	1694750	rs5753350	22	31256066
rs6887887	5	142535913	rs9663489	10	2544480	rs9614325	22	44498134
rs2080085	5	147360904	rs17158495	10	2577639	rs136033	22	46255004



Appendix A4. Splitting patterns among modern population as assessed by TreeMix. Trees are shown to allow for *m=0* through *m=4* migrations (panels *a* to *e*). Branch lengths are proportional to the genetic drift of each population.