bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Assessing temporal and geographic contacts across the Adriatic Sea through the analysis of genome-wide data from Southern Italy

Alessandro Raveane[1,2]*, Ludovica Molinaro[3], Serena Aneli[4], Marco Rosario Capodiferro[1,5], Linda Ongaro[3,5], Nicola Rambaldi Migliore[1], Sara Soffiati[1], Teodoro Scarano[6], Antonio Torroni[1], Alessandro Achilli[1], Mario Ventura[7], Luca Pagani[3,4], Cristian Capelli[8,9], Anna Olivieri[1], Francesco Bertolini[2], Ornella Semino[1,10]*, Francesco Montinaro[3,7,8,10]*

[1] Department of Biology and Biotechnology "L. Spallanzani", University of Pavia, 27100 Pavia, Italy

[2] Laboratory of Hematology-Oncology, European Institute of Oncology IRCCS, 20141 Milan, Italy

[3] Estonian Biocentre, Institute of Genomics, University of Tartu, 51010 Tartu, Estonia

[4] Department of Biology, University of Padua, 35121 Padua, Italy

[5] Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland

[6]Department of Cultural Heritage, University of Salento, 73100 Lecce, Italy

[7] Department of Biology and Genetics, University of Bari, 70126 Bari, Italy

[8] Department of Department of Zoology, University of Oxford, Oxford, UK

[9] Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, 43124 Parma, Italy

[10] These authors contributed equally

*Correspondence to: alessandro.raveane01@universitadipavia.it; ornella.semino@unipv.it; francesco.montinaro@gmail.com

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Abstract

Southern Italy was characterised by a complex prehistory that started with different Palaeolithic cultures, later followed by the Neolithic transition and the demic dispersal from the Pontic-Caspian Steppe during the Bronze Age. Archaeological and historical evidence points to demic and cultural influences between Southern Italians and the Balkans, starting with the initial Palaeolithic occupation until historical and modern times. To shed light on the dynamics of these contacts, we analysed a genome-wide SNP dataset of more than 700 individuals from the South Mediterranean area (102 from Southern Italy), combined with ancient DNA from neighbouring areas. Our findings revealed high affinities of South-Eastern Italians with modern Eastern Peloponnesians, and a closer affinity of ancient Greek genomes with those from specific regions of South Italy than modern Greek genomes. The higher similarity could be associated with the presence of a Bronze Age component ultimately originating from the Caucasus and characterised by high frequencies of Iranian and Anatolian Neolithic ancestries. Furthermore, to reveal possible signals of natural selection, we looked for extremely differentiated allele frequencies among Northern and Southern Italy, uncovering putatively adapted SNPs in genes involved in alcohol metabolism, nevi features and immunological traits, such as ALDH2, *NID1* and *CBLB*.

## Introduction

Southern Italy was one of the first European regions to be inhabited by our species. One of the oldest archaeological remains attributed to *Homo sapiens*, has been in fact excavated in Apulia (Grotta del Cavallo ~45 thousand years ago - kya) and recently associated with the Uluzzian culture (1), also reported in the Middle and Upper Palaeolithic in Greece (2–4), although other interpretations have been proposed (5).

After the Last Glacial Maximum, approximately 20 kya, populations from South-Eastern Europe and West Asia partially replaced continental European Hunter Gatherers (HG) (6,7). Recently, ancient DNA (aDNA) retrieved from remains in North-Eastern Italy post-dated this transition of ~3 ky, and associated it to the demographic transition occurring between Early to Late Epigravettian in Southern Europe around 17 kya (8,9), as confirmed by archaeological data (10). These observations highlighted the existence of a connection between Western and Eastern Europe well before Neolithic times, with Italy playing the role either of a bridge or a refuge (8).

Neolithization revolutionised the culture and the demography of continental Europe, with Southern Italy being the first colonised place West of the Adriatic Sea (11–13). Recently, the availability of ancient Southern European genomes helped in disentangling the dynamics of the early stages of the Pontic Steppe populations diffusion that occurred in the Bronze Age period (14,15). Differently from the rest of Europe, Greece and Southern Italy appear to have been less impacted by this demic dispersal, being characterised by an additional Iranian-related ancestry (16–19). However, the lack of Southern Italian ancient genomes from the Neolithic period keeps open essential questions regarding this major cultural and demographic transition in the region. Starting from the mid 3$^{rd}$ millennium Before Current Era (BCE) archaeological evidence allows to outline a network of cultural connections interacting along the Adriatic-Ionian axis, operating between two or more different core areas and radiating across trajectories of link and expansion which likely triggered small human groups movements (20,21). As a matter of fact, from about 4.3 to 4 kya, the well-know Cetina-type cultural elements, also related to the Bell Beaker phenomenon in the North-Western Balkans, played an active role spreading from the Dalmatian core area Southwards across

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the Adriatic in Northern Apulia and South-Eastern Italy also influencing the Ionian Islands and Western Greece (22).

Later on, during the 2nd millennium BCE a flourishing and continuous cultural relationship was established between Southern Italy and Aegean communities especially from the Recent Bronze Age (3.3 to 3.2 kya) onward (23,24). Although the demographic extent of these contacts is not clear, some valuable insights on mobility could be inferred from ceramic crafts. The most recent analytical evidence, relating to the Aegean-type pottery from the core sites of Punta di Zambrone (Tyrrhenian Calabria) and Roca Vecchia (Southern Adriatic Apulia), allow to highlight a strong connection with the Western Greek regions (Ionian Islands, Acarnania, Achaea and Elis) and, to a lesser extent, with Western Crete (25).

Precisely, following the Bronze Age, in a period between the so-called Greek “Dark Ages” (26) and the Archaic Greece (2.7 to 2.5 kya), Southern Italy became a hotspot for the foundation of Greek colonies. The first colonies were set in Campania and Sicily, possibly by settlers arriving from Eastern Greece (Euboea island) around 780 BCE, soon followed by colonies in Calabria, Basilicata and Apulia (27–30), encompassing an area later to be known as *Magna Grecia.* In the second half of the 7th century BCE, some of these settlements, specifically in South-Eastern Sicily (Siracusa and Megara Iblea) and Apulia (Taranto), were attributed to Eastern Peloponnesian founders (31,32).

The nature of the early settlements, the scale of their demographic impact and genetic legacy are still a matter of debate. Some genetic studies (33,34) have tried characterise the demographic impact of these processes in Southern Italy using present-day Italian populations, but none of them had the intent of finely dissecting these ancient components. Furthermore, a recent aDNA study (35) showed that Iron Age Apulians were not yet superimposable to contemporary Southern Italians, pointing to later processes as keys for the understanding of present-day genetic diversity in Italy.

As a consequence of this complex demographic scenario, Italy harbours the largest degree of genetic population structure identified in Europe so far (19,36), making its population a valuable asset for adaptation studies (36–38).

In this study, we performed a genome-wide analysis, unveiling the contemporary genetic structure of present-day South and Eastern European populations, and highlighting recent and past interactions between Southern Italy and Greece. Furthermore, integrating available ancient genomes, we finely characterised the genetic similarities between modern day Southern Italians and different Eurasian ancient groups. Finally, we searched for putatively selected genomic regions by evaluating differences in allele frequencies between Southern and Northern Italians, identifying SNPs potentially under selection that are linked with immunological and dietary traits.

## Results

### Southern Italian genetic structure

Genetic stratification of Eurasian individuals, inferred by *ADMIXTURE* (K = 9), revealed the existence of three main components virtually shared by all the Europeans included in our dataset (Fig. 1A, Fig. S1). These were structured as follows: one (depicted in salmon in Fig. 1A), modal in Sardinia, was also present at high percentages in some Italian and Greek present-day populations; another (yellow), most common in Germany and Slovakia, occurred at similar proportions across Greece, Kosovo and the Italian Peninsula. An

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

additional component (green) is modal in the Caucasus area and also largely present in Italy and Greece.

*f4* analysis showed a higher allele sharing of modern Caucasus individuals with present-day Central Europeans than with South-Eastern Europeans, possibly as the result of Bronze Age migrations (Table S2).

The first two PCs in the PCA highlight genetic similarities between Southern Italians and the majority of the analysed Greek populations, with the only exception of the Peloponnesian groups (for sample location see Fig. S2) previously identified as outliers (39), which form a tight cluster also when the third PC is evaluated (Fig. S3).

In order to further dissect the degree of affinity between populations across the Adriatic Sea, we applied a clustering approach on the first ten PCs of a PCA performed on a subset composed by 682 individuals (Fig. 1B, Fig. S4). We detected 12 clusters summarising the genetic similarity among these populations (Fig. S4). Present-day Apulians and Calabrians are characterised by similar profiles, with three clusters (cl. 1, 4, 6) grouping most of their individuals. Western Sicilians mainly fall in cl. 4, which is virtually absent in Eastern Sicily, but it is modal in the Western Peloponnese. Cl. 6 is the second most represented cluster in Apulia, Calabria and Sicily and is found also in Western Peloponnese. Notably, cl. 1 is found at moderate frequencies (from ~ 16% to ~ 23%) in the whole Southern Italy, is rare in the Western part of Peloponnese (absent in all Western Peloponnese with the exception of Messenia), but it is present at high proportions on its Eastern part (i.e. Laconia ~ 36%), Greece-Macedonia (~ 33%) and Kosovo (~ 42%). Interestingly, a substantial proportion of Sicilians (between ~ 16 to ~ 30%), are part of a cluster which includes Northern and Central Italian areas (cl. 9), and is absent in Greece and Peloponnese. East Sicily and Apulia are the only Italian groups with individuals assigned to a specific cluster observed in drifted Tsakonians (North and South) (cl. 3). Similarly, one individual (out of 17) from Calabria and two Sardinians (out of 24) were assigned to a cluster (cl. 5) composed mainly of the Deep Mani and Taygetos samples, which were previously described as populations that experienced genetic drift (39).

**Modelling the relationship between modern Southern Italian and ancient Eurasians**

We evaluated the genetic variation in Southern Italy with respect to ancient groups assembling a dataset composed of 138 ancient Eurasian individuals dating between the Palaeolithic and the Iron Age.

Ancient genetic profiles were preliminarily investigated through PCA by projecting ancient genotypes onto the first two eigenvectors inferred on modern individuals (Fig. S5). Hunter Gatherers individuals from West (WHG) and East (EHG) Europe cluster on the right side of the PCA delineating a West to East cline along the second PC. Most of the individuals enriched in “Neolithic ancestry” are placed close to the genetic variability of present-day Sardinians. This group of samples also includes more recent individuals (from Late Neolithic to Copper Age) who in turn are scattered close to present-day inhabitants of the island (Fig. S5A). Interestingly, the only Greek Neolithic individual is close to Sardinians and European Early Neolithic samples and is part of a tight cluster that includes Anatolians and two Peloponnesians, from the Neolithic Age. On the other hand, three out of the five Neolithic Peloponnesians, together with the totality of Minoans and Mycenaeans included in our dataset plot towards the genetic variability of people currently inhabiting Southern Peloponnese (Maniots and Tsakonians) and Southern Italian regions (Sicily, Calabria and Apulia) (Fig. S5B). Modern Southern Italians are closer to Southern European Neolithic and Bronze Age samples (Neolithic Peloponnesians and Minoans) than most modern

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Peloponnesian groups, with the exception of the Deep Mani and Taygetos individuals (Fig. S5B). The affinity between Southern Italians and ancient samples was also investigated by the *f-statistics*. First, we tested the affinity symmetry of Northern Italians and other Italian groups (OIGs) with respect to Anatolia Neolithic (AN) samples (OIGs). All the *f4* (Mbuti, AN, OIG, Lombardy) show a significantly higher affinity between Lombardy and AN, with the only exception of Sardinia (Table 1). This pattern is also evident when Bronze Age Greeks (Table 1) or ancient Steppe groups (Table S3) are used in place of AN. These observations suggest that populations from Central and Southern Italy had a lower contribution from AN than Lombardy, or alternatively, that Central and Southern Italians received contributions from other different groups, possibly associated with present-day Middle Eastern or African regions (19,40,41).

However, when the affinity of Italian groups with African and Middle Eastern populations was tested, Southern Italians resulted not significantly closer to any of the two (Table S4).

Except for East Sicily and Calabria, no significant results were detected for the combination *f4* (Mbuti, Steppe, OIG, DeepMani), confirming a similar genetic composition for Maniots and Southern Italian groups. In contrast, replacing Deep Mani with other groups from different Peloponnesian areas resulted in the latter showing a higher affinity to Steppe ancestry than present-day Southern Italians (Table S3).

We further explored the admixture model of Southern Italian groups using *qpAdm*, which fits a vector of *f4* values by summarising the relationships of test and source groups (left populations) to a set of right populations. First, we used the main ancestral populations of present-day Europeans (Anatolia_N, WHG, Iran ChL and EHG) (14,42) as putative sources (Fig. 2A). Notably, all the tested populations, with the exception of Sardinians and Greece Macedonians, show remarkably similar proportions of Anatolia Neolithic source, suggesting that any difference in ancestry observed with the *f4-outgroup* statistics is very subtle. Moreover, most of the tested populations but Sardinians show a relatively high proportion (from ~ 28 to 54 %) of ancestry related to samples inhabiting Iran in the Chalcolithic period (Fig. 2A). Generally, similar AN ancestral proportions were observed when a model including AN, Steppe_EMBA and Iran ChL was considered (Fig. 2B). In this setting, the Northern Italian group (Lombardy) presents a higher contribution of both AN and Steppe EMBA than the Southern Italians, where a remarkable contribution of Iran Chalcolithic was detected (Fig. 2B).

The admixture graph (Fig. 3) helped in delineating ancestral genetic population differences and similarities between present-day Southern Italians and Greeks in whom an overall higher Steppe-related ancestry was detected (Fig. 2B, Table S3). In particular, we were able to fit a graph admixture model in which 63% of the Minoan genome was found in a population that contributed 97% to the genetic pool of present-day Apulians (Fig. 3A). Contrary, modern Greeks were successfully modelled as an admixture of populations related to present-day Balkan populations (Serbian, (43)) and Bronze Age Greek groups, here represented by Minoans (Fig. 3B). This could potentially explain the differences found in Steppe-related ancestries between the two present-day populations. Additionally, these results suggest a more recent gene exchange between Greeks and neighbouring Slavic-related populations. Alternatively, they could be explained by different ancient legacies between the two sides of the Adriatic Sea, as the result of pre-Bronze Age genetic structure in the Balkans.

**Investigating putative signature of adaptation among Northern and Southern Italians**

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Following the differences in ancient ancestry contributions across the Italian Peninsula, we applied the Population Branch Statistics (PBS) analysis in order to identify markers showing significant allele frequency differences among pairs of Italian populations using CHB as an outgroup. We estimated the average PBS across windows of five consecutive SNPs and focused on the top ten windows with higher frequency in Northern (positive PBS values) and Southern Italy (negative PBS values), respectively (Fig. S6). We finally annotated the variants with the most extreme PBS values (above the 99.5 percentile and below 0.05) and evaluated if they were enriched for any biological or functional feature.

Among the top 20 windows, seven do not overlap with any known gene (Fig. 4, Table S5). Of those showing the highest allele frequencies in Northern Italy, one encompasses *PABPC4L* (*H. sapiens* poly(A) binding protein, cytoplasmic 4-like), a gene associated with venous thromboembolism (44), birth weight (45), blood protein level and alcohol consumption (46); another, located on chromosome 1, encompasses the gene *NID1* (Nidogen 1), which is associated with blood proteins (47) and platelet levels (48,49), and a third on chromosome 3 maps within the *CBLB* (Cbl proto-oncogene, E3 ubiquitin protein ligase B) gene, involved in the immune response by regulating T- and B- cell receptors (50).

Of the windows with extreme negative PBS values, the two most divergent are located within chromosome 12 and encompass 15 genes (Fig. 4, Table S5). They include *ALDH2*, known to affect the metabolism of alcohol and correlated to different behavioural traits associated with alcohol consumption (51). We note here that our dataset does not contain SNPs within this gene, and it is thus possible that this signal is led by variants in other regions. Finally, we found extreme divergence in a region within the *EDAR* (Ectodysplasin A receptor) gene, associated with variations in head hair morphology and facial hair thickness (52). Although the SNP rs3827760 (chr2:109513601), affecting hair follicle thickness and shape, is not present in our dataset, it is unlikely that it is driving the adaptation signal, as also observed in a previous exome analysis (37).

## Discussion

Since prehistoric times Southern Italy has been a cross-roads of many human groups and cultures, and their genetic signatures are still present in the DNA of individuals currently living in this area (19,53–55).

In this study, we highlighted a high similarity between Southern Italy and the Peloponnese. In fact, our cluster analysis showed that present-day South-Eastern Peloponnesian populations have high genetic affinity with modern Apulians, Calabrians and South-Eastern Sicilians, all characterised by a cluster composition different from those displayed by other Greek groups (Fig. 1B, Fig. S3). Additionally, individuals from Western Sicily show similarities with populations inhabiting the Western part of Peloponnese (Fig. 1B, Fig. S4). Although establishing the chronological context for this affinity using present-day genomes might be challenging, our results are in accordance with archaeological and historical sources that attributed the origin of Greek colonies in South-Eastern Sicily and Apulia from populations inhabiting the southern and Eastern parts of the Peloponnese (31,32). Uniparental Y-chromosome findings are also in agreement with these observations revealing Eastern Peloponnesian ancestries in East Sicily (34) and shared haplogroups among modern-day Greeks and populations living in Southern Italian areas colonised by Greeks such as the Salento (Apulia) and the Ionian coast of Calabria (56). The lower affinity with other Balkan populations could be attributed to a lower influence by inland populations, such as Slavic-related people (57) and/or genetic drift in Tsakones and Maniots as suggested by

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

historical sources (39). Therefore, our results imply a high affinity between Southern Italians and Peloponnesians possibly abrupted very recently by major events of migrations and/or admixture as the one recorded during the Middle Age period (58). However the observation that, in some analyses, Southern Italians and ancient Greeks share more alleles than modern and ancient Peloponnesians, may suggest a scenario including the preservation of an ancient population signal in the genome of Southern Italians that was likely diluted by inland migrations in Greece (Fig. 3).

Notably, a substantial genetic affinity was highlighted between Peloponnesian individuals from the Neolithic and the Bronze Age, whose signatures are also recorded in Italy. Recent genomic studies (19,33,38) detected a contribution ultimately derived from the Caucasus (CHG) in modern Southern Italians and likely brought in Italy as early as the Bronze Age, although their demographic dynamics are still not known. Overall, these results are in agreement with the detection of a small proportion of Iranian-related ancestry in Sicilian Middle Bronze Age samples (17), which could be tentatively linked to the spread of the Mycenaean culture (59). Interestingly, our results modelled the source of this contribution as a mixture of AN and Iran Chalcolithic ancestries (14). The latter was found consistently across Southern Italy and the Peloponnese, confirming again common genetic sources shared between these two regions (Fig. 2).

The largest degree of genetic heterogeneity across the European continent has been so far recorded in the Italian Peninsula (19), an important aspect to consider when epidemiological and translational studies in Italy are planned. Recent researches taking advantage of high coverage whole-genome and whole-exome sequences linked variants under selection in Italians to genes that were related to insulin secretion, obesity, thermogenesis, alcohol consumption, pathogenic-response, skin colour and cancer (37,38). Here we assessed the occurrence of extreme differences in allele frequencies between Northern and Southern Italy close to the *ALDH2, NID1,* and *CBLB* genes that may have a role in alcohol metabolism, nevi features and immunological traits, respectively. Although the analysed dataset contained only ~100k SNPs, this replication in the possibly affected phenotypes may suggest an actual important difference among groups on the two parts of the peninsula. Nevertheless, genomic scans are sometimes controversial and difficult to interpret, and a deeper investigation, possibly involving multiple tests on denser datasets, is highly desirable.

In conclusion, we provided new insights on the composition of modern genomes of individuals collected from Southern Italy, a still poorly sampled region of Europe. The fine characterisation of the time of arrival and routes of migrations related to the detected ancient signals is expected to be addressed only by the analysis of Apulian ancient genomes covering a period ranging from the Late Neolithic to the Early and Middle Bronze Age periods. The differences identified in modern-day Southern and Northern Italians have significant phenotypic implications and call for more extensive investigations on a larger number of markers and individuals.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Material and Methods

## Dataset

We assembled a dataset comprehensive of 2,662 present-day individuals from the Mediterranean basin, Europe and Africa genotyped with different Illumina SNP arrays (39,43,57,60–68) (Table S1).

Only SNPs and individuals with a missingness rate lower than 1% were retained (*--geno* and *--mind* flags equal to 0.01). Pairs of related individuals (PI_HAT > 0.05) were identified using PLINK1.9 and within each pair one was randomly removed (69). After merging and pruning, we retained 93,647 linked and 56,956 unlinked (*--indep-pairwise* 100 25 0.2) SNPs. The modern-day dataset was merged with 138 publicly available ancient individuals (6,14,15,18,42,70–82), for whom genotypes of at least 20,000 markers were available (Table S1).

## Genetic structure analyses

*Admixture analysis*

The present-day dataset was processed using *ADMIXTURE* (83). Ten independent runs of unsupervised *ADMIXTURE* for different numbers of clusters (K), from 2 to 20, were performed. Runs were summarised, ordered and aligned using *CLUMPAK* (84). The lowest *Cross Validation* (CV) median error across all the independent runs was observed for K= 9.

*Principal component analysis (PCA)*

Principal Component Analysis was performed to evaluate the genetic structure of Southern Italians in the context of other present-day and ancient Eurasian individuals. All the PCAs were performed with the *smartpca EIGENSOFT* package (85).

First, we considered a dataset including 826 present-day Europeans. Three iterations were initially used to remove outliers resulting in a final dataset of 804 individuals. Secondly, we projected 138 ancient samples on the PCs estimated from the genetic variability of 1,165 present-day Western Eurasians using the *lsqproject* and *autoshrink* options set to *YES* and five iterations to remove outliers.

*Genetic clustering of Mediterranean populations*

In order to group homogenous individual at a more detailed resolution that that offered by *ADMIXTURE* or by the first two PCAs, we applied the clustering approach to finite mixture models implemented in the *mclust* R package (86) using 36 as the number of mixture components for which the BIC is calculated. The approach was performed on the first 10 principal components estimated on 682 present-day individuals from Eastern regions of the Mediterranean basin using *smartpca* and five iterations to remove outliers. The highest BIC value was reported for 12 clusters.

## F-statistics

*f3* and *f4* outgroup analyses were performed using *admixr* v0.7.1 R package (87) which exploits *ADMIXTOOLS v4.2* package (88).

We preliminary explored both Denisovan (89) and the Mbuti, a rainforest African group of hunter-gatherers, as outgroups; since no differences were encountered, we selected Mbuti as the outgroup. A |Z| score higher than 3 was considered significant.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Ancient admixture analyses**

*QpGraph* models were generated using *ADMIXTOOLS v6.5* (88) with Mbuti or Japanese set as outgroups. We used the default settings and filtered out results with a maximum fit for f-statistics (|Z| score) lower than 3.

*QpWave* and *qpAdm* in *ADMIXTOOLSv4.1* package were used as in (19,90) by testing all the possible combinations of two, three and four sources on a set of target populations. We firstly used *qpWave* to establish, using *f-statistics* and a set of outgroups, whether the tested sources were significantly different and could be used to reconstruct the ancestry of a target population. Subsequently, *qpAdm* was employed to infer the proportions of the tested sources.

The threshold used for *qpWave* to establish the significance was set at 0.01, while the set of sources included: Minoan, Mycenaean, Iran Neolithic (Iran_N), Anatolia Bronze Age (Anatolia_BA), Anatolia Neolithic (Anatolia_N), Steppe Early Middle Bronze Age (Steppe_EMBA), Western Hunter Gatherer (WHG), Eastern Hunter Gatherer (EHG), Caucasus Hunter Gatherer (CHG), Peloponnese_Neolithic, Late Western Caucasus Bronze Age culture (LateMaykop), Iran_ChL (Chalcholitic), SteppeMaykopoutlier, Morocco (Table S1).

The set of outgroups comprised: Mota, AfontovaGora3, EHG, CHG, WHG, Goyet, Vestonice16, Malta1, ElMiron, Iran Neolithic (Iran_N), Levant Neolithic (Levant_N), Anatolia Neolithic (Anatolia_N), Kostenki14, Natufian, Ust_Ishim (Table S1).

**Selection scan**

The Population Branching Statistic (PBS), which harnesses the differences in pairwise $F_{st}$ values between three populations, was estimated with the function *allel.pbs* of the *scikit-allel python* package (https://zenodo.org/badge/latestdoi/7890/cggh/scikit-allel) using a window_size of 1 and setting the *normed* option as *True*, which normalised the PBS as in (91). We focused on differences between Southern (Abruzzo, Calabria, East Sicily, West Sicily, Sicily, Apulia) and Northern Italy (from Lombardy, Tuscany) using the Han Chinese in Beijing (CHB) population as an outgroup. We averaged the PBS score across windows comprising five consecutive SNPs. Gene annotation was performed with CADD (Combined Annotation Dependent Depletion) (92) and known genetic association for each variant was assessed interrogating the GWAS catalogue, considaering only associations with a *p-value* $< 5E^{-08}$ (retrieved on November 2021; (93,94)).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## References


1. Benazzi, S., Douka, K., Fornai, C., Bauer, C.C., Kullmer, O., Svoboda, J., Pap, I., Mallegni, F., Bayle, P., Coquerelle, M., *et al.* (2011) Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature*, **479**, 525–528.
2. Mellars, P. (2005) The impossible coincidence. A single-species model for the origins of modern human behavior in Europe. *Evol. Anthropol. Issues News Rev.*, **14**, 12–27.
3. Moroni, A., Ronchitelli, A., Simona, A., Aureli, D., Bailey, S.E., Boscato, P., Boschin, F., Capecchi, G., Crezzini, J., Douka, K., *et al.* (2018) Grotta del Cavallo (Apulia – Southern Italy). The Uluzzian in the mirror. *J. Anthropol. Sci.*, **96**, 125–160.
4. Villa, P., Pollarolo, L., Conforti, J., Marra, F., Biagioni, C., Degano, I., Lucejko, J.J., Tozzi, C., Pennacchioni, M., Zanchetta, G., *et al.* (2018) From Neandertals to modern humans: New data on the Uluzzian. *PLOS ONE*, **13**, e0196786.
5. Martini, F., Ronchitelli, A. and Sarti, L. (2017) Il Paleolitico e il Mesolitico della Puglia. *Preistoria e Protostoria della Puglia*, Istituto italiano di preistoria e protostoria, pp. 25–38.
6. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., *et al.* (2016) The genetic history of Ice Age Europe. *Nature*, **534**, 200–205.
7. Pala, M., Olivieri, A., Achilli, A., Accetturo, M., Metspalu, E., Reidla, M., Tamm, E., Karmin, M., Reisberg, T., Kashani, B.H., *et al.* (2012) Mitochondrial DNA Signals of Late Glacial Recolonization of Europe from Near Eastern Refugia. *Am. J. Hum. Genet.*, **90**, 915–924.
8. Bortolini, E., Pagani, L., Oxilia, G., Posth, C., Fontana, F., Badino, F., Saupe, T., Montinaro, F., Margaritora, D., Romandini, M., *et al.* (2021) Early Alpine occupation backdates westward human migration in Late Glacial Europe. *Curr. Biol.*, **31**, 2484-2493.e7.
9. Villalba-Mouco, V., van de Loosdrecht, M.S., Posth, C., Mora, R., Martínez-Moreno, J., Rojo-Guerra, M., Salazar-García, D.C., Royo-Guillén, J.I., Kunst, M., Rougier, H., *et al.* (2019) Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula. *Curr. Biol.*, **29**, 1169-1177.e7.
10. Cristiani, E., Farbstein, R. and Miracle, P. (2014) Ornamental traditions in the Eastern Adriatic: The Upper Palaeolithic and Mesolithic personal adornments from Vela Spila (Croatia). *J. Anthropol. Archaeol.*, **36**, 21–31.
11. Betti, L., Beyer, R.M., Jones, E.R., Eriksson, A., Tassi, F., Siska, V., Leonardi, M., Maisano Delser, P., Bentley, L.K., Nigst, P.R., *et al.* (2020) Climate shaped how Neolithic farmers and European hunter-gatherers interacted after a major slowdown from 6,100 BCE to 4,500 BCE. *Nat. Hum. Behav.*, **4**, 1004–1010.
12. Lo Vetro, D. and Martini, F. (2016) Mesolithic in Central–Southern Italy: Overview of lithic productions. *Quat. Int.*, **423**, 279–302.
13. Natali, E. and Forgia, V. (2018) The beginning of the Neolithic in Southern Italy and Sicily. *Quat. Int.*, **470**, 253–269.
14. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., *et al.* (2016) Genomic insights into the origin of farming in the ancient Near East. *Nature*, **536**, 419–424.
15. Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., *et al.* (2018) The genomic history of southeastern Europe. *Nature*, **555**, 197–203.
16. Clemente, F., Unterländer, M., Dolgova, O., Amorim, C.E.G., Coroado-Santos, F., Neuenschwander, S., Ganiatsou, E., Cruz Dávalos, D.I., Anchieri, L., Michaud, F., *et al.* (2021) The genomic history of the Aegean palatial civilizations. *Cell*, **184**, 2565-2586.e21.
17. Fernandes, D.M., Mittnik, A., Olalde, I., Lazaridis, I., Cheronet, O., Rohland, N., Mallick, S., Bernardos, R., Broomandkhoshbacht, N., Carlsson, J., *et al.* (2020) The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. *Nat.*

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*Ecol. Evol.*, **4**, 334–345.
18. Lazaridis, I., Mittnik, A., Patterson, N., Mallick, S., Rohland, N., Pfrengle, S., Furtwängler, A., Peltzer, A., Posth, C., Vasilakis, A., *et al.* (2017) Genetic origins of the Minoans and Mycenaeans. *Nature*, **548**, 214–218.
19. Raveane, A., Aneli, S., Montinaro, F., Athanasiadis, G., Barlera, S., Birolo, G., Boncoraglio, G., Blasio, A.M.D., Gaetano, C.D., Pagani, L., *et al.* (2019) Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe. *Sci. Adv.*
20. Broodbank, C. (2013) The Making of the Middle Sea: A History of the Mediterranean from the Beginning to the Emergence of the Classical World. *The Making of the Middle Sea: A History of the Mediterranean from the Beginning to the Emergence of the Classical World*; Oxford University Press, (2013) .
21. Teržan, B. and Karavanić, S. (2013) The Western Balkans in the Bronze Age. *Oxf. Handb. Eur. Bronze Age*, 837–863.
22. Cazzella, A. (2020) 2500-2000 BC□: connectivity phenomena between the Balkans, Greece, Southern Italy, Eastern Sicily, the Aeolian Islands and Malta. *Riv. Sci. Preistoriche LXX*, **S1**, 181–197.
23. Bettelli, M. and Vagnetti, L. (2019) Southern Italy. *A Companion to the Archaeology of Early Greece and the Mediterranean*, John Wiley & Sons, Ltd, pp. 1261–1286.
24. Vagnetti, L. (2012) Western Mediterranean. *Oxf. Handb. Bronze Age Aegean*, 890–905.
25. Jung, R., Guglielmino, R., Iacono, F. and Mommsen, H. (2021) Neutron Activation Analysis of Aegean and Aegeanising Ceramics from Roca Vecchia and the Circulation of Pottery in Southern Italy. *Punta Zambrone*, OAW, pp. 459–490.
26. Martin, T.R. (2013) Ancient Greece: From Prehistoric to Hellenistic Times: Second Edition. *Ancient Greece: From Prehistoric to Hellenistic Times: Second Edition*; Yale University Press, (2013) .
27. Agostino, B. (2009) Pithecusa e Cuma all'alba della colonizzazione, in Cuma. *Atti del 48° Convegno di Studi sulla Magna Grecia, Taranto*, pp. 3–28.
28. Agostino, B. and D'Acunto, M. (2009) La città e le mura: nuovi dati dall'area nord della città antica, in Cuma. *Atti del 48° Convegno di Studi sulla Magna Grecia, Taranto*, pp. 481–522.
29. Antonaccio, C.M. (2007) Colonization Greece on the Move, 900-480. In Shapiro, H. A. (ed.), *The Cambridge Companion to Archaic Greece*, Cambridge Companions to the Ancient World, Cambridge University Press, Cambridge, pp. 201–224.
30. Mele, A. (2014) Greci in Campania. *Greci in Campania*; Scienze e Lettere, (2014) .
31. Braccesi, L. and Nocita, M. (2016) HESPERÌA, 33. I Fondatori delle Colonie. Tra Sicilia e Magna Grecia. *HESPERÌA, 33. I Fondatori delle Colonie. Tra Sicilia e Magna Grecia.*; L'ERMA di BRETSCHNEIDER, (2016) .
32. Guzzo, P.G. (2011) Fondazioni greche: l'Italia meridionale e la Sicilia (VIII e VII sec. a.C.). *Fondazioni greche: l'Italia meridionale e la Sicilia (VIII e VII sec. a.C.)*; Carocci, (2011) .
33. Sarno, S., Boattini, A., Pagani, L., Sazzini, M., De Fanti, S., Quagliariello, A., Gnecchi Ruscone, G.A., Guichard, E., Ciani, G., Bortolini, E., *et al.* (2017) Ancient and recent admixture layers in Sicily and Southern Italy trace multiple migration routes along the Mediterranean. *Sci. Rep.*, **7**, 1984.
34. Tofanelli, S., Brisighelli, F., Anagnostou, P., Busby, G.B.J., Ferri, G., Thomas, M.G., Taglioli, L., Rudan, I., Zemunik, T., Hayward, C., *et al.* (2016) The Greeks in the West: genetic signatures of the Hellenic colonisation in southern Italy and Sicily. *Eur. J. Hum. Genet.*, **24**, 429–436.
35. Aneli, S., Saupe, T., Montinaro, F., Solnik, A., Molinaro, L., Scaggion, C., Carrara, N., Raveane, A., Kivisild, T., Metspalu, M., *et al.* (2022) The genetic origin of Daunians and the Pan-Mediterranean southern Italian Iron Age context. *Mol. Biol. Evol.*, msac014.
36. Sazzini, M., Gnecchi Ruscone, G.A., Giuliani, C., Sarno, S., Quagliariello, A., De Fanti, S., Boattini, A., Gentilini, D., Fiorito, G., Catanoso, M., *et al.* (2016) Complex interplay between neutral and adaptive evolution shaped differential genomic background and

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

disease susceptibility along the Italian peninsula. *Sci. Rep.*, **6**, 32513.
37. Birolo, G., Aneli, S., Gaetano, C.D., Cugliari, G., Russo, A., Allione, A., Casalone, E., Giorgio, E., Paraboschi, E.M., Ardissino, D., *et al.* (2021) Functional and clinical implications of genetic structure in 1686 Italian exomes. *Hum. Mutat.*, **42**, 272–289.
38. Sazzini, M., Abondio, P., Sarno, S., Gnecchi-Ruscone, G.A., Ragno, M., Giuliani, C., De Fanti, S., Ojeda-Granados, C., Boattini, A., Marquis, J., *et al.* (2020) Genomic history of the Italian population recapitulates key evolutionary dynamics of both Continental and Southern Europeans. *BMC Biol.*, **18**, 51.
39. Stamatoyannopoulos, G., Bose, A., Teodosiadis, A., Tsetsos, F., Plantinga, A., Psatha, N., Zogas, N., Yannaki, E., Zalloua, P., Kidd, K.K., *et al.* (2017) Genetics of the peloponnesean populations and the theory of extinction of the medieval peloponnesean Greeks. *Eur. J. Hum. Genet.*, **25**, 637–645.
40. Henn, B.M., Botigué, L.R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J.K., Fadhlaoui-Zid, K., Zalloua, P.A., Moreno-Estrada, A., Bertranpetit, J., *et al.* (2012) Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *PLOS Genet.*, **8**, e1002397.
41. Moorjani, P., Patterson, N., Hirschhorn, J.N., Keinan, A., Hao, L., Atzmon, G., Burns, E., Ostrer, H., Price, A.L. and Reich, D. (2011) The History of African Gene Flow into Southern Europeans, Levantines, and Jews. *PLOS Genet.*, **7**, e1001373.
42. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., *et al.* (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, **513**, 409–413.
43. Kovacevic, L., Tambets, K., Ilumäe, A.-M., Kushniarevich, A., Yunusbayev, B., Solnik, A., Bego, T., Primorac, D., Skaro, V., Leskovac, A., *et al.* (2014) Standing at the Gateway to Europe - The Genetic Structure of Western Balkan Populations Based on Autosomal and Haploid Markers. *PLOS ONE*, **9**, e105090.
44. Deguchi, H., Shukla, M., Hayat, M., Torkamani, A., Elias, D.J. and Griffin, J.H. (2020) Novel exomic rare variants associated with venous thrombosis. *Br. J. Haematol.*, **190**, 783–786.
45. Warrington, N.M., Beaumont, R.N., Horikoshi, M., Day, F.R., Helgeland, Ø., Laurin, C., Bacelis, J., Peng, S., Hao, K., Feenstra, B., *et al.* (2019) Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat. Genet.*, **51**, 804–814.
46. Cheng, S., Wen, Y., Liu, L., Cheng, B., Liang, C., Ye, J., Chu, X., Yao, Y., Jia, Y., Kafle, O.P., *et al.* (2021) Traumatic events during childhood and its risks to substance use in adulthood: an observational and genome-wide by environment interaction study in UK Biobank. *Transl. Psychiatry*, **11**, 431.
47. Emilsson, V., Ilkov, M., Lamb, J.R., Finkel, N., Gudmundsson, E.F., Pitts, R., Hoover, H., Gudmundsdottir, V., Horman, S.R., Aspelund, T., *et al.* (2018) Co-regulatory networks of human serum proteins link genetics to disease. *Science*, **361**, 769–773.
48. Chen, M.-H., Raffield, L.M., Mousas, A., Sakaue, S., Huffman, J.E., Moscati, A., Trivedi, B., Jiang, T., Akbari, P., Vuckovic, D., *et al.* (2020) Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell*, **182**, 1198-1213.e14.
49. Nan, H., Xu, M., Zhang, J., Zhang, M., Kraft, P., Qureshi, A.A., Chen, C., Guo, Q., Hu, F.B., Rimm, E.B., *et al.* (2011) Genome-wide association study identifies nidogen 1 ( NID1 ) as a susceptibility locus to cutaneous nevi and melanoma risk. *Hum. Mol. Genet.*, **20**, 2673–2679.
50. Voisinne, G., García-Blesa, A., Chaoui, K., Fiore, F., Bergot, E., Girard, L., Malissen, M., Burlet-Schiltz, O., Gonzalez de Peredo, A., Malissen, B., *et al.* (2016) Co-recruitment analysis of the CBL and CBLB signalosomes in primary T cells identifies CD5 as a key regulator of TCR-induced ubiquitylation. *Mol. Syst. Biol.*, **12**, 876.
51. Jorgenson, E., Thai, K.K., Hoffmann, T.J., Sakoda, L.C., Kvale, M.N., Banda, Y., Schaefer, C., Risch, N., Mertens, J., Weisner, C., *et al.* (2017) Genetic contributors to variation in alcohol consumption vary by race/ethnicity in a large multi-ethnic genome-wide association study. *Mol. Psychiatry*, **22**, 1359–1367.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

52. Endo, C., Johnson, T.A., Morino, R., Nakazono, K., Kamitsuji, S., Akita, M., Kawajiri, M., Yamasaki, T., Kami, A., Hoshi, Y., *et al.* (2018) Genome-wide association study in Japanese females identifies fifteen novel skin-related trait associations. *Sci. Rep.*, **8**, 8974.
53. Aneli, S., Caldon, M., Saupe, T., Montinaro, F. and Pagani, L. (2021) Through 40,000 years of human presence in Southern Europe: the Italian case study. *Hum. Genet.*, **140**, 1417–1431.
54. Antonio, M.L., Gao, Z., Moots, H.M., Lucci, M., Candilio, F., Sawyer, S., Oberreiter, V., Calderon, D., Devitofranceschi, K., Aikens, R.C., *et al.* (2019) Ancient Rome: A genetic crossroads of Europe and the Mediterranean. *Science*.
55. Saupe, T., Montinaro, F., Scaggion, C., Carrara, N., Kivisild, T., D'Atanasio, E., Hui, R., Solnik, A., Lebrasseur, O., Larson, G., *et al.* (2021) Ancient genomes reveal structural shifts after the arrival of Steppe-related ancestry in the Italian Peninsula. *Curr. Biol.*, **31**, 2576-2591.e12.
56. Grugni, V., Raveane, A., Mattioli, F., Battaglia, V., Sala, C., Toniolo, D., Ferretti, L., Gardella, R., Achilli, A., Olivieri, A., *et al.* (2018) Reconstructing the genetic history of Italians: new insights from a male (Y-chromosome) perspective. *Ann. Hum. Biol.*, **45**, 44–56.
57. Busby, G.B.J., Hellenthal, G., Montinaro, F., Tofanelli, S., Bulayeva, K., Rudan, I., Zemunik, T., Hayward, C., Toncheva, D., Karachanak-Yankova, S., *et al.* (2015) The Role of Recent Admixture in Forming the Contemporary West Eurasian Genomic Landscape. *Curr. Biol.*, **25**, 2518–2526.
58. McKee, S. (2000) Uncommon Dominion: Venetian Crete and the Myth of Ethnic Purity. *Uncommon Dominion: Venetian Crete and the Myth of Ethnic Purity*; University of Pennsylvania Press, (2000) .
59. Cline, E.H. (2010) The Oxford Handbook of the Bronze Age Aegean. *The Oxford Handbook of the Bronze Age Aegean*; OUP USA, (2010) .
60. Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
61. Behar, D.M., Metspalu, M., Baran, Y., Kopelman, N.M., Yunusbayev, B., Gladstein, A., Tzur, S., Sahakyan, H., Bahmanimehr, A., Yepiskoposyan, L., *et al.* (2013) No Evidence from Genome-Wide Data of a Khazar Origin for the Ashkenazi Jews. *Hum. Biol.*, **85**, 859–900.
62. Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., *et al.* (2010) The genome-wide structure of the Jewish people. *Nature*, **466**, 238–242.
63. Kushniarevich, A., Utevska, O., Chuhryaeva, M., Agdzhoyan, A., Dibirova, K., Uktveryte, I., Möls, M., Mulahasanovic, L., Pshenichnov, A., Frolova, S., *et al.* (2015) Genetic Heritage of the Balto-Slavic Speaking Populations: A Synthesis of Autosomal, Mitochondrial and Y-Chromosomal Data. *PLOS ONE*, **10**, e0135820.
64. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., *et al.* (2008) Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. *Science*, **319**, 1100–1104.
65. Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S.Q., Thomas, M.G., Luiselli, D., *et al.* (2012) Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex Influences on the Ethiopian Gene Pool. *Am. J. Hum. Genet.*, **91**, 83–96.
66. Sallustio, F., Cox, S.N., Serino, G., Curci, C., Pesce, F., De Palma, G., Papagianni, A., Kirmizis, D., Falchi, M. and Schena, F.P. (2015) Genome-wide scan identifies a copy number variable region at 3p21.1 that influences the TLR9 expression levels in IgA nephropathy patients. *Eur. J. Hum. Genet.*, **23**, 940–948.
67. Yunusbayev, B., Metspalu, M., Metspalu, E., Valeev, A., Litvinov, S., Valiev, R., Akhmetova, V., Balanovska, E., Balanovsky, O., Turdikulova, S., *et al.* (2015) The

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genetic Legacy of the Expansion of Turkic-Speaking Nomads across Eurasia. *PLOS Genet.*, **11**, e1005068.

68. Yunusbayev, B., Metspalu, M., Järve, M., Kutuev, I., Rootsi, S., Metspalu, E., Behar, D.M., Varendi, K., Sahakyan, H., Khusainova, R., *et al.* (2012) The Caucasus as an Asymmetric Semipermeable Barrier to Ancient Human Migrations. *Mol. Biol. Evol.*, **29**, 359–365.
69. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. and Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**, s13742-015-0047–8.
70. Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., *et al.* (2015) Population genomics of Bronze Age Eurasia. *Nature*, **522**, 167–172.
71. Broushaki, F., Thomas, M.G., Link, V., López, S., van Dorp, L., Kirsanow, K., Hofmanová, Z., Diekmann, Y., Cassidy, L.M., Díez-del-Molino, D., *et al.* (2016) Early Neolithic genomes from the eastern Fertile Crescent. *Science*, **353**, 499–503.
72. Hofmanová, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Díez-del-Molino, D., Dorp, L. van, López, S., Kousathanas, A., Link, V., *et al.* (2016) Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl. Acad. Sci.*, **113**, 6886–6891.
73. Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., *et al.* (2015) Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.*, **6**, 8912.
74. Keller, A., Graefen, A., Ball, M., Matzas, M., Boisguerin, V., Maixner, F., Leidinger, P., Backes, C., Khairat, R., Forster, M., *et al.* (2012) New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.*, **3**, 698.
75. Kılınç, G.M., Omrak, A., Özer, F., Günther, T., Büyükkarakaya, A.M., Bıçakçı, E., Baird, D., Dönertaş, H.M., Ghalichi, A., Yaka, R., *et al.* (2016) The Demographic Development of the First Farmers in Anatolia. *Curr. Biol.*, **26**, 2659–2666.
76. Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., *et al.* (2017) Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature*, **551**, 368–372.
77. Martiniano, R., Caffell, A., Holst, M., Hunter-Mann, K., Montgomery, J., Müldner, G., McLaughlin, R.L., Teasdale, M.D., van Rheenen, W., Veldink, J.H., *et al.* (2016) Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nat. Commun.*, **7**, 10326.
78. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., *et al.* (2015) Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, **528**, 499–503.
79. Mittnik, A., Wang, C.-C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R., Khartanovich, V., Moiseyev, V., Tõrv, M., *et al.* (2018) The genetic prehistory of the Baltic Sea region. *Nat. Commun.*, **9**, 442.
80. Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M., *et al.* (2019) The formation of human populations in South and Central Asia. *Science*, **365**, eaat7487.
81. Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., *et al.* (2018) The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature*, **555**, 190–196.
82. Wang, C.-C., Reinhold, S., Kalmykov, A., Wissgott, A., Brandt, G., Jeong, C., Cheronet, O., Ferry, M., Harney, E., Keating, D., *et al.* (2019) Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nat. Commun.*, **10**, 590.
83. Alexander, D.H., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.*, **19**, 1655–1664.
84. Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. and Mayrose, I. (2015)

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.*, **15**, 1179–1191.

85. Patterson, N., Price, A.L. and Reich, D. (2006) Population Structure and Eigenanalysis. *PLOS Genet.*, **2**, e190.
86. Scrucca, L., Fop, M., Murphy, T.B. and Raftery, A.E. (2016) mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J.*, **8**, 289–317.
87. Petr, M., Vernot, B. and Kelso, J. (2019) admixr—R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics*, **35**, 3194–3195.
88. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T. and Reich, D. (2012) Ancient Admixture in Human History. *Genetics*, **192**, 1065–1093.
89. Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., *et al.* (2017) A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*, **358**, 655–658.
90. Tamm, E., Di Cristofaro, J., Mazières, S., Pennarun, E., Kushniarevich, A., Raveane, A., Semino, O., Chiaroni, J., Pereira, L., Metspalu, M., *et al.* (2019) Genome-wide analysis of Corsican population reveals a close affinity with Northern and Central Italy. *Sci. Rep.*, **9**, 13581.
91. Malaspinas, A.-S., Westaway, M.C., Muller, C., Sousa, V.C., Lao, O., Alves, I., Bergström, A., Athanasiadis, G., Cheng, J.Y., Crawford, J.E., *et al.* (2016) A genomic history of Aboriginal Australia. *Nature*, **538**, 207–214.
92. Rentzsch, P., Schubach, M., Shendure, J. and Kircher, M. (2021) CADD-Splice—improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med.*, **13**, 31.
93. Denny, J.C., Bastarache, L., Ritchie, M.D., Carroll, R.J., Zink, R., Mosley, J.D., Field, J.R., Pulley, J.M., Ramirez, A.H., Bowton, E., *et al.* (2013) Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.*, **31**, 1102–1110.
94. MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., *et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.*, **45**, D896–D901.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Figure Legends

**Figure 1. Admixture and genetic relatedness in modern South-Eastern Europeans.**
A) *ADMIXTURE* plot for K = 9 (lowest CV error) and 34 populations. Populations are ordered according to their geographical longitude. B) Barplots represent the proportion of individuals assigned to a specific cluster (pie in the inset) by the mclust clustering approach (see Fig. S4). Bars are placed on the population geographic origin.

**Figure 2. Modelling of ancestry proportion in South-Eastern European populations.**
A). Supported *qpAdm* proportions for the main ancestral populations of present-day Europeans. B) as in A) using a Steppe EMBA ancestry instead of WHG and EHG ancestral groups as indicated in the legend. Colours are as in Fig. S5.

**Figure 3. Graph modelling of present-day Apulians and Greeks.**
Z scores of both trees are between -3 and 3. Ancestries highlighted in shades of red are ancestries specifically shared with Bronze Age Greeks. Letters from A to F refer to unsampled ancestral populations.

**Figure 4. Karyotype plot of the top 10 windows from PBS analysis.** In red and blue are indicated the chromosome locations for the top 10 PBS in North and South Italy, respectively. Where present, the set of annotated genes is presented.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.26.482072; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Table Legends

**Table 1. *f4-outgroup* statistics to compare the relationships between Neolithic and Greek Bronze Age samples with Northern Italians and Sardinians.**

| Outgroup | Ancient Sources | OIG[1] | North Italy | f4 | SD | Z |
|---|---|---|---|---|---|---|
| Mbuti | Anatolia_N | Abruzzo | Lombardy | 0.000591 | 0.000196 | 3.02 |
| | | Calabria | | 0.001111 | 0.000172 | 6.442 |
| | | EastSicily | | 0.001381 | 0.000177 | 7.789 |
| | | Apulia | | 0.000915 | 0.000153 | 5.995 |
| | | Sardinia | | -0.001415 | 0.000177 | -7.98 |
| | | Sicily | | 0.001393 | 0.000349 | 3.992 |
| | | Tuscany | | 0.000436 | 0.000138 | 3.172 |
| | | WestSicily | | 0.001421 | 0.000171 | 8.321 |
| | Minoan | Abruzzo | | 0.000839 | 0.000235 | 3.573 |
| | | Calabria | | 0.001392 | 0.000212 | 6.576 |
| | | EastSicily | | 0.001676 | 0.000218 | 7.683 |
| | | Apulia | | 0.001016 | 0.000183 | 5.562 |
| | | Sardinia | | -0.000835 | 0.000215 | -3.884 |
| | | Sicily | | 0.001343 | 0.000433 | 3.099 |
| | | Tuscany | | 0.000582 | 0.00017 | 3.417 |
| | | WestSicily | | 0.001574 | 0.000206 | 7.628 |
| | Mycenaean | Abruzzo | | 0.00077 | 0.000372 | 2.068 |
| | | Calabria | | 0.000832 | 0.000335 | 2.486 |
| | | EastSicily | | 0.001749 | 0.000348 | 5.025 |
| | | Apulia | | 0.001128 | 0.000299 | 3.772 |
| | | Sardinia | | -0.000666 | 0.000339 | -1.968 |
| | | Sicily | | 0.002023 | 0.000672 | 3.008 |
| | | Tuscany | | 0.000668 | 0.00028 | 2.387 |
| | | WestSicily | | 0.00142 | 0.000344 | 4.13 |

[1]Other Italian Group

(which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A

South Italy

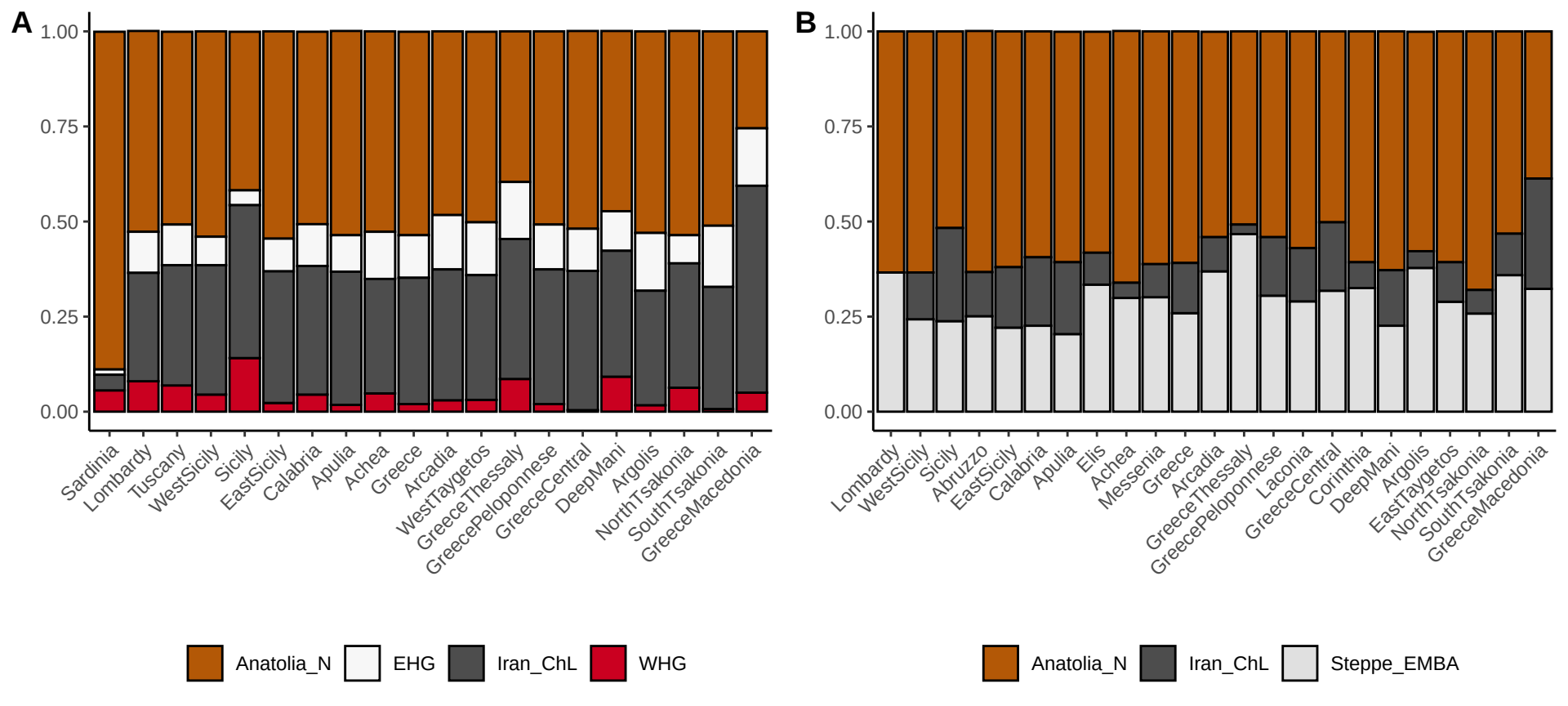
Pelopponese

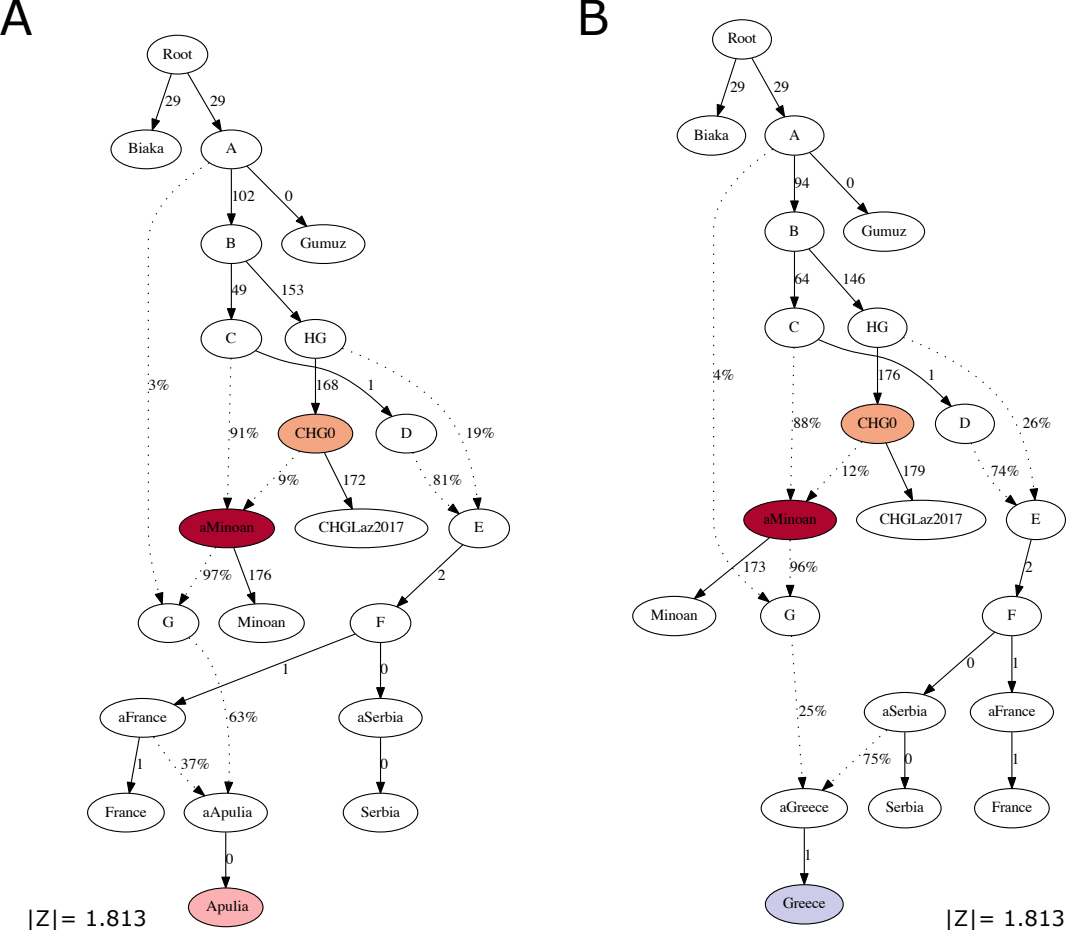
Sardinia Lombardy Tuscany EastSicily WestSicily Abruzzo Calabria Apulia Germany Slovenia Croatia Bosnia Montenegro Slovakia Serbia Kosovo Macedonia Greece Macedonia Elis Achea Messenia Arcadia WestTaygetos GreeceThessaly GreecePeloponnese Laconia Corinthia DeepMani Argolis EastTaygetos NorthTsakonia SouthTsakonia Caucasus_North Caucasus_South

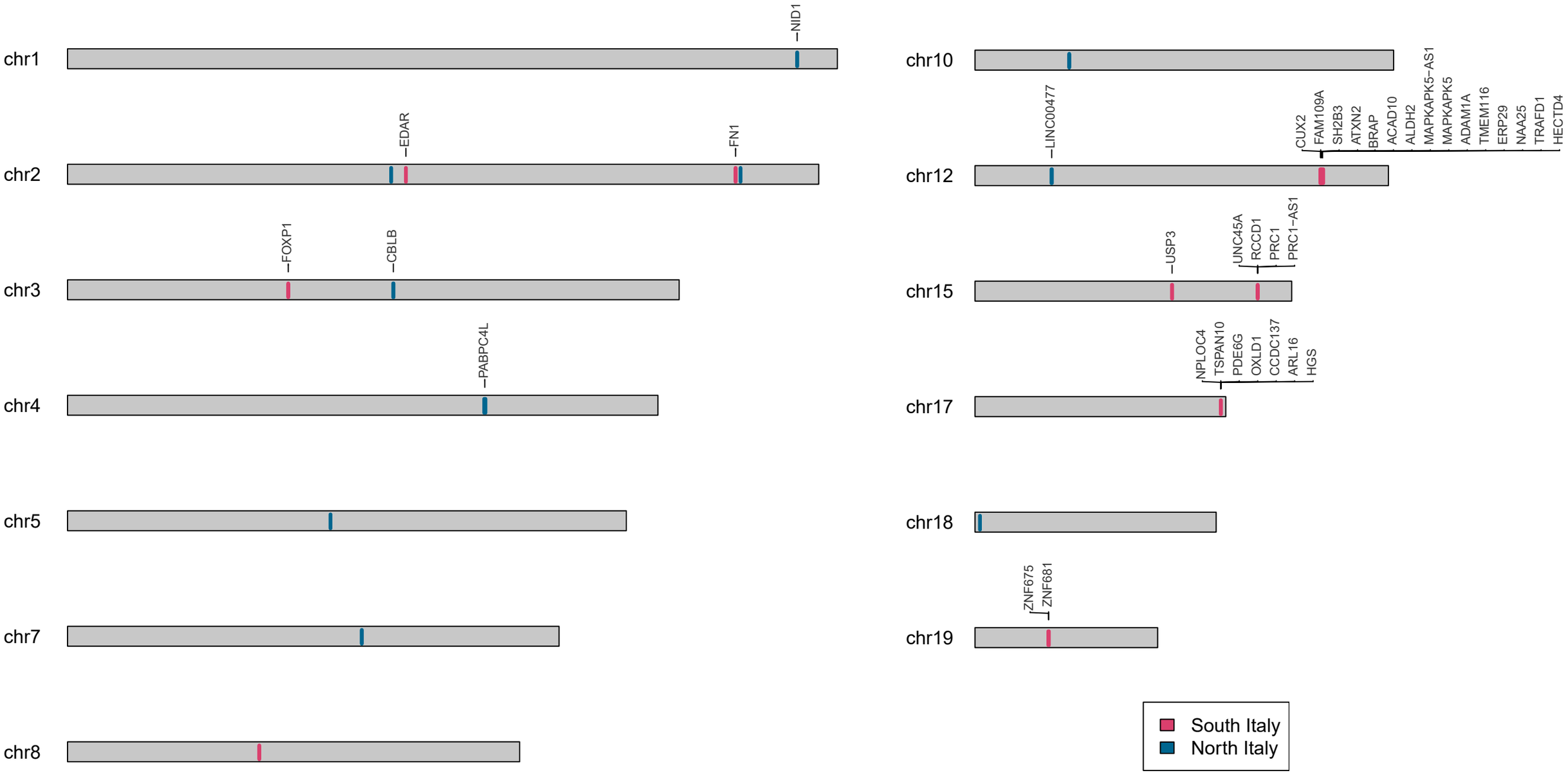
1.00
0.75
0.50
0.25
0.00

B

cl.1 cl.2 cl.3 cl.4 cl.5 cl.6 cl.7 cl.8 cl.9 cl.10 cl.11 cl.12


A
1.00
0.75
0.50
0.25
0.00
Sardinia
Lombardy
Tuscany
WestSicily
Sicily
EastSicily
Calabria
Apulia
Achea
Greece
Arcadia
WestTaygetos
GreeceThessaly
GreecePeloponnese
GreeceCentral
DeepMani
Argolis
NorthTsakonia
SouthTsakonia
GreeceMacedonia
Anatolia_N
EHG
Iran_ChL
WHG
B
1.00
0.75
0.50
0.25
0.00
Lombardy
WestSicily
Sicily
Abruzzo
EastSicily
Calabria
Apulia
Elis
Achea
Messenia
Greece
Arcadia
GreeceThessaly
GreecePeloponnese
Laconia
GreeceCentral
Corinthia
DeepMani
Argolis
EastTaygetos
NorthTsakonia
SouthTsakonia
GreeceMacedonia
Anatolia_N
Iran_ChL
Steppe_EMBA


A
Root
29
29
Biaka
A
102
0
B
Gumuz
49
153
C
HG
3%
168
1
91%
CHG0
D
19%
9%
172
81%
aMinoan
CHGLaz2017
E
97%
176
2
G
Minoan
F
1
0
aFrance
63%
aSerbia
1
37%
0
France
aApulia
Serbia
0
Apulia
|Z|= 1.813
B
Root
29
29
Biaka
A
94
0
B
Gumuz
64
146
C
HG
4%
176
1
88%
CHG0
D
26%
12%
179
74%
aMinoan
CHGLaz2017
E
173
96%
2
Minoan
G
F
0
1
25%
aSerbia
aFrance
75%
0
1
aGreece
Serbia
France
1
Greece
|Z|= 1.813


chr1
NID1
chr2
EDAR
FN1
chr3
FOXP1
CBLB
chr4
PABPC4L
chr5
chr7
chr8
chr10
chr12
LINC00477
CUX2
FAM109A
SH2B3
ATXN2
BRAP
ACAD10
ALDH2
MAPKAPK5-AS1
MAPKAPK5
ADAM1A
TMEM116
ERP29
NAA25
TRAFD1
HECTD4
chr15
USP3
UNC45A
RCCD1
PRC1
PRC1-AS1
chr17
NPLOC4
TSPAN10
PDE6G
OXLD1
CCDC137
ARL16
HGS
chr18
chr19
ZNF675
ZNF681
South Italy
North Italy