HUMAN GENETICS

Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe

A. Raveane[1,2*†], S. Aneli[2,3,4*†], F. Montinaro[2,5*†], G. Athanasiadis[6], S. Barlera[7], G. Birolo[3,4], G. Boncoraglio[8,9], A. M. Di Blasio[10], C. Di Gaetano[3,4], L. Pagani[5,11], S. Parolo[12‡], P. Paschou[13], A. Piazza[3,14], G. Stamatoyannopoulos[15§], A. Angius[16], N. Brucato[17], F. Cucca[16], G. Hellenthal[18], A. Mulas[19], M. Peyret-Guzzon[20], M. Zoledziewska[16], A. Baali[21], C. Bycroft[20‖], M. Cherkaoui[21], J. Chiaroni[22,23], J. Di Cristofaro[22,23], C. Dina[24], J. M. Dugoujon[17], P. Galan[25], J. Giemza[24], T. Kivisild[5,26], S. Mazieres[22], M. Melhaoui[27], M. Metspalu[5], S. Myers[20], L. Pereira[28,29], F. X. Ricaut[17], F. Brisighelli[30], I. Cardinali[31], V. Grugni[1], H. Lancioni[31], V. L. Pascali[30], A. Torroni[1], O. Semino[1], G. Matullo[3,4¶], A. Achilli[1¶], A. Olivieri[1¶], C. Capelli[2*¶]

European populations display low genetic differentiation as the result of long-term blending of their ancient founding ancestries. However, it is unclear how the combination of ancient ancestries related to early foragers, Neolithic farmers, and Bronze Age nomadic pastoralists can explain the distribution of genetic variation across Europe. Populations in natural crossroads like the Italian peninsula are expected to recapitulate the continental diversity, but have been systematically understudied. Here, we characterize the ancestry profiles of Italian populations using a genome-wide dataset representative of modern and ancient samples from across Italy, Europe, and the rest of the world. Italian genomes capture several ancient signatures, including a non–steppe contribution derived ultimately from the Caucasus. Differences in ancestry composition, as the result of migration and admixture, have generated in Italy the largest degree of population structure detected so far in the continent, as well as shaping the amount of Neanderthal DNA in modern-day populations.

INTRODUCTION

Our understanding of the events that shaped European genetic variation has been redefined by the availability of ancient DNA (aDNA). In particular, it has emerged that, in addition to the contributions of early hunter-gatherer populations, major genetic components can be traced back to Neolithic and Bronze Age expansions (*1*). Extensive gene flow across the continent over the last few thousand years (*1*, *2*) has further contributed to the correlation between geography and genetic variation observed in modern Europe (*3*).

The arrival of farming in Europe led to admixture between incoming Anatolian farmers and autochthonous hunter-gatherers, a process that generated individuals genetically close to present-day Sardinians (*4*, *5*). During the Bronze Age, the dispersal of a population related to the pastoralist nomadic Yamnaya from the Pontic-Caspian

[1]Department of Biology and Biotechnology "L. Spallanzani", University of Pavia, Pavia, Italy. [2]Department of Zoology, University of Oxford, Oxford, UK. [3]Department of Medical Sciences, University of Turin, Turin, Italy. [4]IIGM (Italian Institute for Genomic Medicine), Turin, Italy. [5]Estonian Biocentre, Institute of Genomics, University of Tartu, Tartu, Estonia. [6]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [7]Department of Cardiovascular Research, Istituto di Ricovero e Cura a Carattere Scientifico–Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy. [8]Department of Cerebrovascular Diseases, IRCCS Istituto Neurologico Carlo Besta, Milan, Italy. [9]PhD Program in Neuroscience, University Milano-Bicocca, Monza, Italy. [10]Istituto Auxologico Italiano, IRCCS, Centro di Ricerche e Tecnologie Biomediche, Milano, Italy. [11]APE lab, Department of Biology, University of Padua, Padua, Italy. [12]Computational Biology Unit, Institute of Molecular Genetics, National Research Council, Pavia, Italy. [13]Department of Biological Sciences, Purdue University, West Lafayette, IN, USA. [14]Academy of Sciences, Turin, Italy. [15]Department of Medicine and Genome Sciences, University of Washington, Seattle, WA, USA. [16]Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), Monserrato, Cagliari, Italy. [17]Evolutionary Medicine Group, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse, Centre National de la Recherche Scientifique (CNRS), Université de Toulouse, Toulouse, France. [18]University College London Genetics Institute (UGI), University College London, London, UK. [19]Istituto di Ricerca Genetica e Biomedica (IRGB), CNR, Lanusei, Italy. [20]The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [21]Faculté des Sciences Semlalia de Marrakech (FSSM), Université Cadi Ayyad, Marrakech, Morocco. [22]Aix Marseille Univ, CNRS, EFS, ADES, Marseille, France. [23]Etablissement Français du Sang PACA Corse, Biologie des Groupes Sanguins, Marseille, France. [24]l'institut du thorax, INSERM, CNRS, University of Nantes, Nantes, France. [25]Equipe de Recherche en Epidémiologie Nutritionnelle (EREN), Centre de Recherche en Epidémiologie et Statistiques, Université Paris 13/Inserm U1153/Inra U1125/ Cnam, COMUE Sorbonne Paris Cité, F-93017 Bobigny, France. [26]Department of Human Genetics, KU Leuven, Herestraat 49, box 604, Leuven 3000, Belgium. [27]Faculté des Sciences, Université Mohammed Premier, Oujda, Morocco. [28]i3S–Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal. [29]IPATIMUP–Instituto de Patologia e Imunologia Molecular, Universidade do Porto, Porto, Portugal. [30]Section of Legal Medicine, Institute of Public Health, Catholic University of the Sacred Heart, Rome, Italy. [31]Department of Chemistry, Biology and Biotechnology, University of Perugia, Perugia, Italy.
*Corresponding author. Email: alessandro.raveane01@universitadipavia.it (A.R.); serena.aneli@gmail.com (S.A.); francesco.montinaro@gmail.com (F.M.); cristian.capelli@zoo.ox.ac.uk (C.C.)

†These authors contributed equally to this work
‡Present address: Fondazione the Microsoft Research, University of Trento Centre for Computational and Systems Biology (COSBI), Piazza Manifattura 1, 38068 Rovereto (TN), Italy.
§Deceased.
‖Present address: Genomics Plc, King Charles House, Park End Street, Oxford OX1 1JD, UK.
¶Co-senior authors.

steppe further markedly affected the genetic landscape of the continent (*1*, *6*). This migration, supported by archeological and genetic data, has also been linked to the spread of the Indo-European languages in Europe and the introduction of several technological innovations in peninsular Eurasia (*7*). Genetically, ancient steppe populations have been described as a combination of Eastern and Caucasus hunter-gatherer/Iran Neolithic (EHG and CHG/IN) ancestries (*4*). However, the analysis of aDNA from Southern East Europe has identified the existence of additional contributions ultimately from the Caucasus (*8*, *9*) and suggests a more complex ancient ancestry composition for Europeans (*4*).

The geographic location of Italy, enclosed between continental Europe and the Mediterranean Sea, makes the Italian people relevant for the investigation of continent-wide demographic events to complement and enrich the information provided by aDNA studies. To characterize the genetic variability of modern-day European populations and their relationship to early European foragers, Neolithic farmers, and Bronze Age nomadic pastoralists, we investigated the population structure of present-day Italians and other Europeans in terms of their ancestry composition as the result of migration and admixture, both ancient and historical. To do this, we assembled and analyzed a comprehensive genome-wide single nucleotide polymorphism (SNP) dataset composed of 1616 individuals from all 20 Italian administrative regions and more than 140 worldwide reference populations to give a total of 5192 modern-day samples (fig. S1, A and B, and data file S1), to which we added genomic data available for ancient individuals (data file S1).

RESULTS

Distinctive genetic structure in Italy

We initially investigated patterns of genetic differentiation in Italy and surrounding regions by exploring the information embedded in the SNP-based haplotypes of modern samples [full modern dataset (FMD) including 218,725 SNPs]. The phased genome-wide dataset was analyzed using the ChromoPainter (CP) and fineSTRUCTURE (fS) pipeline (see the Supplementary Materials) (*10*, *11*) to generate a tree of groups of individuals with similar "copying vectors" (clusters; Fig. 1A). The fraction of pairs of individuals placed in the same cluster across multiple runs was, on average, 0.95 for Italian clusters and 0.96 across the whole set of clusters (see Materials and Methods and the Supplementary Materials). Non-European clusters were pooled into larger groups in subsequent analyses (see Materials and Methods and the Supplementary Materials).

Italian clusters separated into three main groups: Sardinia, Northern (North/Central-North Italy), and Southern Italy (South/Central-South Italy and Sicily); the first two were close to populations originally from Western Europe, while the last was closer to Middle Eastern groups (Fig. 1, A and B; figs. S1D and S2, A to C; and data file S1). These observations were confirmed using a subset of the dataset genotyped for a larger number of SNPs [high-density dataset (HDD) including 591,217 SNPs; see Materials and Methods and the Supplementary Materials; fig. S1D and data file S1]. To highlight the geographic distribution of the identified clusters along the Italian peninsula, we reconstructed the cluster composition of the various administrative regions of Italy by using the best sampling origin information available for each individual in our dataset (Fig. 1C and data file S1). Recent migrants and admixed individuals, as identified on the basis of their copying vectors (fig. S3, data file S2), were removed in subsequent CP/fS analyses (see Materials and Methods and the Supplementary Materials).

A sharp north-south division in cluster distribution was detected, the separation between northern and southern areas being shifted north along the peninsula (Fig. 1B) (*12*). The reported structure dismissed the possibility that the Central Italian populations differentiated from the Northern and Southern Italian groups (Fig. 1A) (*13*). Individuals from Central Italy were, in fact, assigned mostly to the Southern Italian clusters, except for samples from Tuscany, which grouped instead with the Northern Italian clusters (Fig. 1, A and B) (*12*). Contrary to previous results, no outliers were detected among the Northern Italian clusters (*12*).

We evaluated the impact of geography in shaping the genomic variability of Italy by testing the correlation between geographic and genetic coordinates, applying a Procrustes analysis (Fig. 1B; Materials and Methods). A significant correlation was observed in our dataset, in agreement with reports for Italy and Europe (*3*, *13*) (correlation in Procrustes rotation, 0.77 and 0.78, $P < 0.05$ including or excluding Sardinia, respectively). Intracluster variation within Southern and Northern Italian clusters was comparable (data file S2). Sardinian clusters were characterized by a high proportion of genome copied by individuals from the same cluster (self-copying), in agreement with previous indications of drift in Sardinian groups (*12*, *13*). Southern Italian samples showed higher among-cluster F_{ST} values than the Northern Italian ones, but lower TVD (total variation distance) values (data file S2; Materials and Methods) (*1*, *2*). Sardinian clusters showed both high TVD and F_{ST} intercluster variation, combining the effects of drift and variation in ancestry composition.

We compared the degree of variation among genetic clusters in Italy with those in several European countries (*11*, *14*–*16*) and across the whole of Europe (Fig. 1D). Among-cluster variability in Italy was significantly higher than in any other country examined (F_{ST}; median Italy, 0.004; data file S2; range medians for listed countries, 0.0001 to 0.002) and showed differences comparable with estimates across European clusters (median European clusters, 0.004; Materials and Methods and the Supplementary Materials), as previously suggested using unilinear data (*17*, *18*). The analysis of the migration surfaces [estimated effective migration surfaces (EEMS)] (*19*) highlighted several barriers to gene flow within and around Italy, but also suggested the existence of migration corridors in the southern part of the Adriatic and Ionian seas and between Sardinia, Corsica, and continental Italy (Fig. 1E and fig. S4) (*9*).

Multiple ancient ancestries in Italian clusters

To further characterize the observed genetic structure, we investigated the ancestry composition of modern clusters. We tested different combinations of ancient putative sources using a "mixture fit" approach [non-negative least square (NNLS) algorithm] (*11*, *20*). We applied this approach to ancient samples using the "unlinked" mode implemented in CP, similar to other routinely performed analyses based on genotype data, such as qpAdm and ADMIXTURE. In addition, data from modern individuals (from the FMD dataset) were harnessed as donor populations (Materials and Methods and Supplementary Materials). Following Lazaridis *et al.* (*8*), we performed two separate CP/NNLS analyses, "ultimate" and "proximate," referring to the least and the most recent putative sources, respectively (Fig. 2 and fig. S5, A to E). In the ultimate analysis, all the Italian clusters were characterized by relatively high amounts of Anatolian Neolithic (AN) contributions, ranging from 56% (SItaly1) to 72% (NItaly4),

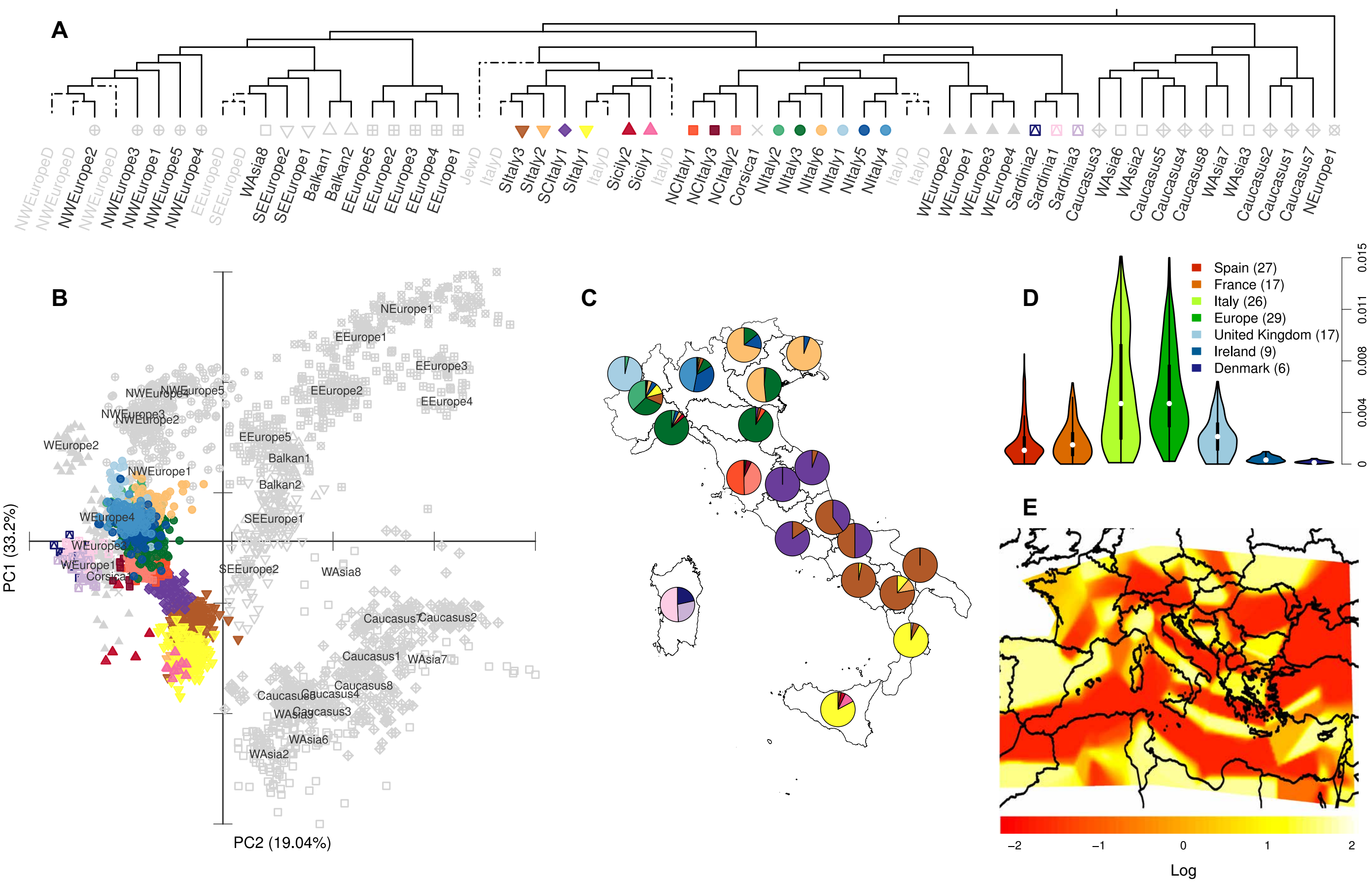

Fig. 1. Genetic structure of the Italian populations. (**A**) Simplified dendrogram of 3057 Eurasian samples clustered by the fS algorithm using the CP output (complete dendrogram in fig. S1C). Each leaf represents a cluster of individuals with similar copying vectors. Clusters with more than five individuals are labeled in black. Italian clusters are color coded. Gray labels ending in the form <<NAME>>_D refer to clusters containing less than five individuals or individuals of uncertain origin that have been removed in the following analyses. (**B**) Principal components analysis (PCA) based on the CP chunkcount matrix [colors as in (A)]. The centroid of the individuals belonging to non-Italian clusters is identified by the label for each cluster. The plot was rotated to the left by 90° to highlight the correspondence with the geography of the Italian samples. (**C**) Pie charts summarizing the relative proportions of inferred fS genetic clusters for all the 20 Italian administrative regions [colors as in (A)]. (**D**) Between-cluster F_{ST} estimates within European groups. Clusters were generated using only individuals belonging to the population analyzed (see Materials and Methods and the Supplementary Materials). The number of genetic clusters analyzed for each population is reported within brackets. For the comparisons across Europe, the cluster NEurope1 containing almost exclusively Finnish individuals was excluded (F_{ST} estimates for Italian and European clusters are in data file S2). F_{ST} distributions statistically lower than the Italian one are in colors other than green. (**E**) Estimated effective migration surfaces (EEMS) analysis in Southern Europe. Colors represent the $\log_{10}$ scale of the effective migration rate, from low (red) to high (yellow).

distributed along a north-south cline (Spearman ρ = 0.52, $P < 0.05$; Fig. 2, A to C, fig. S5A, and data file S3), with Sardinians showing values above 80%, as previously suggested (*1*, *21*). A closer affinity of Northern Italian than Southern Italian clusters to AN was also supported by D-statistics (fig. S6A). The remaining ancestry was mainly assigned to WHG (western hunter-gatherer), CHG, and EHG. In particular, the first two components were more present in populations from the South of Italy ($P < 0.05$, Student's *t* test), while the latter was higher in Northern Italian clusters ($P < 0.05$, Student's *t* test). These observations suggest the existence of different secondary source contributions to the two edges of the peninsula, with the north affected more by EHG-related populations and the south by CHG-related groups. IN ancestry was detected in Europe only in Southern Italy (Fig. 2 and fig. S5A).

North-south differences across Italy were also detected in the proximate analysis. When proximate sources were evaluated, significantly higher ABA (Anatolia Bronze Age) and SBA (Steppe Bronze Age) ancestries were detected in Southern and Northern Italy, respectively (Fig. 2, D to F, and fig. S5B; $P < 0.05$, Student's *t* test; $P < 0.05$, Wilcoxon rank sum test; Supplementary Materials), in line with the results based on the D-statistics (fig. S6, A and B) and mirroring the CHG and EHG patterns, respectively (Fig. 2, A to C). Contrary to previous reports (*4*), the occurrence of CHG as detected by our CP/NNLS analysis did not mirror the presence of SBA, with several populations testing positive for the latter but not for the former (Fig. 2 and fig. S5, A and B). When we compared this analysis and the one using a different CHG sample (SATP) (*5*), the two were highly correlated (Spearman ρ = 0.972, $P < 0.05$; fig. S5F). We therefore speculate that our approach might, in general, underestimate the presence of CHG across the continent; however, we note that even considering this scenario, the excess of Caucasus-related ancestry detected in the south of the European continent, and in Southern Italy in particular, is

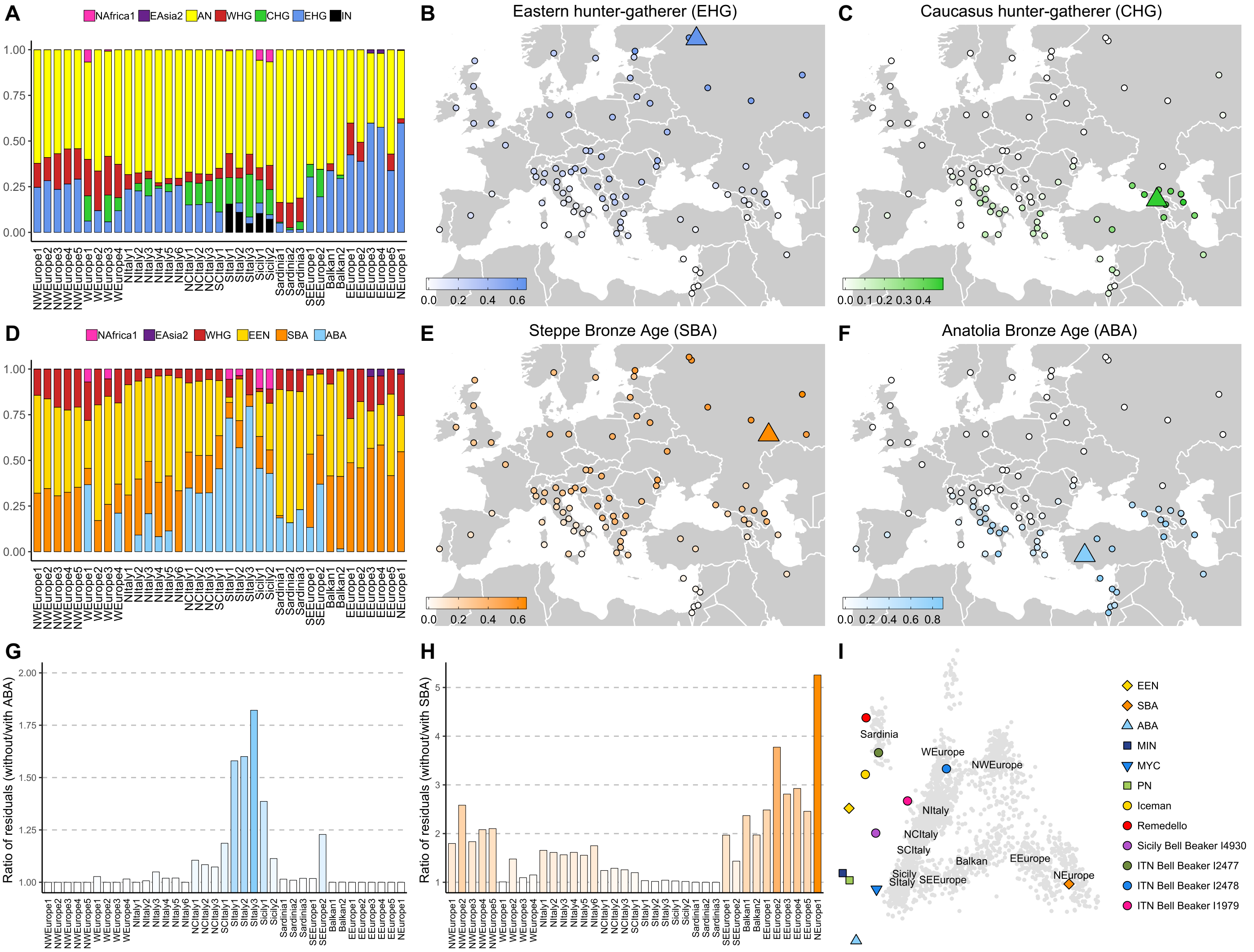

Fig. 2. Ancient ancestries in Western Eurasian modern-day clusters and Italian ancient samples. CP/NNLS analysis on all Italian and European clusters using as donors different sets of ancient samples and two modern clusters (NAfrica1, North Africa; EAsia2, East Asia) [full results in fig. S5 (A and B)]. (**A**) Ultimate sources: AN, Anatolian Neolithic (Bar8); WHG, western hunter-gatherer (Bichon); CHG, Caucasus hunter-gatherer (KK1); EHG, Eastern hunter-gatherer (I0061); IN, Iranian Neolithic (WC1). (**B**) EHG and (**C**) CHG ancestry contributions in Western Eurasia, as inferred in (A) and figs. S8A and S5A. (**D**) Same as in (A), using proximate sources: WHG, western hunter-gatherer (Bichon); EEN, European Early Neolithic (Stuttgart); SBA, Bronze Age from steppe (I0231); ABA, Bronze Age from Anatolia (I2683). (**E**) SBA and (**F**) ABA ancestry contributions, as inferred in (D) and fig. S5B. Triangles refer to the location of ancient samples used as sources (data file S1). (**G**) Ratio of the residuals in the NNLS analysis (see Materials and Methods and the Supplementary Materials) for all the Italian and European clusters when ABA was excluded and included in the set of proximate sources; (**H**) as in (**G**), but excluding/including SBA instead of ABA. (**I**) Ancient Italian and other selected ancient samples projected on the components inferred from modern European individuals. Labels are placed at the centroid of the individuals belonging to the indicated clusters.

notable and unexplained by currently proposed models for the peopling of the continent. The different impact of ABA and SBA across Italy was additionally supported by the reduced fit of the NNLS (sum of the squared residuals; Materials and Methods and Supplementary Materials) when the proximate analysis was run excluding one of these two sources. The residuals were almost twice as much for Southern Italians when ABA was not included as a source, while a substantial increase in the residual values was observed for Northern Italians when SBA was removed from the panel of proximate sources (Fig. 2, G and H). The different affinities of Southern and Northern Italians for ABA and SBA were also highlighted in the principal components analysis (PCA) and ADMIXTURE analysis on ancient and modern samples (Fig. 2I and fig. S7).

These results were confirmed by the qpAdm analysis, where all the analyzed Italian clusters could be modeled as a combination of ABA, SBA, and European Middle-Neolithic/Chalcolithic populations, their contributions mirroring the pattern observed in the CP/NNLS analysis (fig. S5G and data file S4). Sardinian clusters were consistently modeled as AN + WHG + CHG/IN across runs, with the inclusion of North Africa and SBA when a different number of sources were considered. The qpAdm analyses of Italian HDD clusters generated similar results (Materials and Methods, Supplementary Materials, and data file S4).

We noted that in the Balkan peninsula, signatures related to ABA were present but were less evident than in Southern Italy across modern-day populations, possibly being masked by historical contributions from Central Europe (Figs. 2 and 3 and fig. S5B) (*2*, *21*, *22*). Overall, SBA and ABA appeared to have different distribution patterns in Italy and to reflect those present in Europe as a whole: more common in North Italy and across the continent in the case of the former, more localized in the south of Europe and Italy in the case of the latter. Similar results were obtained when other Southern European ancient sources replaced ABA in the proximate analysis (see fig. S5, C to E, Materials and Methods, and the Supplementary Materials).

Modeling the ancestry composition of ancient Italian samples

To obtain temporal insights into the emergence of the differences between Northern and Southern Italy in relation to SBA and ABA ancestries, we performed the same qpAdm analysis on post-Neolithic/Bronze Age Italian individuals (data file S4). Iceman and Remedello, the oldest Italian samples included here [3400 to 2800 Before Common Era (BCE)], were composed of high proportions of AN (74 and 85%, respectively). The Bell Beaker samples of Northern Italy (2200 to 1930 BCE) were modeled as ABA and AN + SBA and WHG. Although ABA estimates in these samples were characterized by large standard errors (SE), the detection of steppe ancestry, at approximately 14%, was more robust. In contrast, Bell Beaker samples from Sicily (2500 to 1900 BCE) were modeled almost exclusively as ABA, with less than 5% SBA (data file S4). Despite the fact that the small number of SNPs and prehistoric individuals tested prevents the formulation of conclusive results, differences in the occurrence of AN ancestry, and possibly also of Bronze Age–related contributions, are suggested to be present between ancient samples from North and South Italy. Differences across ancient Italian samples were also supported by their projections on the PCA of modern-day data (Fig. 2I). Remedello and Iceman clustered with European Early Neolithic samples, together with one of the three Bell Beaker individuals from North Italy, as previously reported (*23*), and modern-day Sardinians. The other two Bronze Age North Italian samples clustered with modern North Italians, while the Bell Beaker sample from Sicily was projected in between European Early Neolithic, Bronze Age Southern European, and modern-day Southern Italian samples (Fig. 2I). These results suggest a differentiation in ancient ancestry composition between different areas of Italy, dating at least in part back to the Bronze Age.

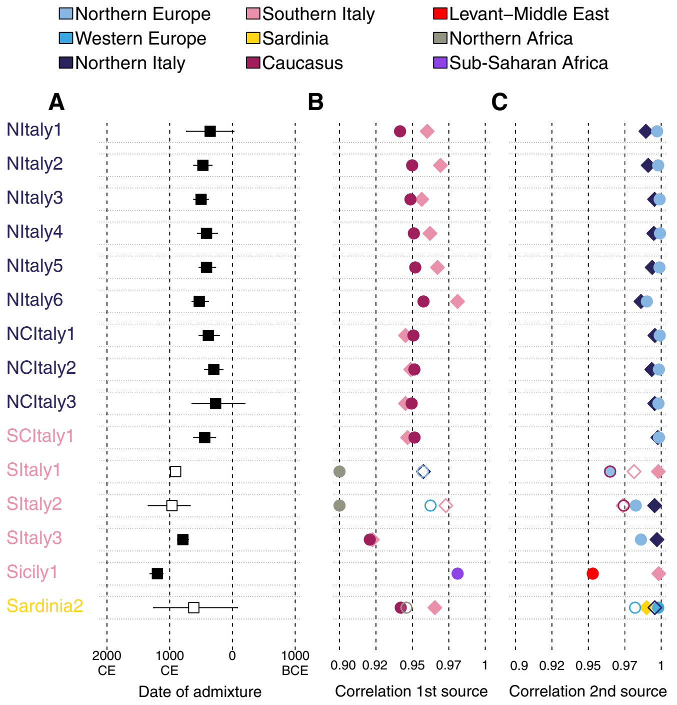

Fig. 3. Admixture events inferred by GT. (**A**) Dates of the events inferred in the GT noItaly analysis on all the Italian clusters (labels as in Fig. 1A and data file S1; full results in fig. S8 and data file S5; see Materials and Methods and the Supplementary Materials); lines encompassed the 95% confidence interval. GT events were distinguished in "one date" (black squares; 1D in data file S5) and "one date multiway" (white squares; 1 MW). (**B**) Correlation values between copying vectors of first source(s) identified by GT and the best proxy in the noItaly analysis (circles) or the best proxy among Italian clusters (diamonds). (**C**) Same as in (B), referring to second source(s) copying vectors. Empty symbols refer to additional first (B) and second (C) sources detected in multiway events. African best proxies in (B) for clusters SItaly1 and SItaly2 were plotted on the 0.90 boundary for visualization only, the correlation values being 0.78 and 0.87, respectively. The symbols referring to the best Italian proxies for the African sources identified for clusters SItaly1, SItaly2, Sicily1, and Sardinia3 in (B) are not included as the correlation values are lower than the African ones and below the threshold used in the figure. The colors of the symbols refer to the ancestry to which proxies were assigned (see Materials and Methods and the Supplementary Materials).

Historical admixture

To investigate the contribution of historical admixture events in shaping the modern distribution of ancient ancestries and the observed population structure, we characterized admixture events for Italian and European populations using GLOBETROTTER (GT) (Fig. 3, fig. S8, and data file S5)(*22*). We discuss here the results based on the nonlocal (*2*) (GT "noItaly") analysis of the FMD (data file S5) as it provides a wider coverage at the population level; concordant results were found when the HDD was similarly analyzed (data file S5).

We ran the analysis excluding the Italians as donors to reduce copying between highly similar groups (Fig. 3). The events detected in Italy occurred mostly between 1000 and 2000 years ago (ya) and extended to 2500 ya in the rest of Europe (Fig. 3A and fig. S8). The profile of inferred sources varied across Italy. Clusters from the Caucasus and North-West Europe were identified for many Italian clusters as best proxies for the admixing sources in agreement with previous studies (*21*), while Middle Eastern and African groups were detected for Southern Italy and Sardinia (Fig. 3, B and C). When Italian clusters were included among putative sources, they were as good as, or better proxies than, clusters from the Caucasus and the Middle East. On the other hand, North-West European and African clusters were mostly confirmed as better proxies than groups from any other area (Fig. 3, B and C), as observed when GT was run including all clusters as donors ("GTall" analysis; data file S5). Overall, these results supported a scenario in which gene flow mostly occurred between Italian and African/other Eurasian populations. SBA and ABA ancestries were detected in Italian and non-Italian best proxies (Figs. 2D and 3 and data file S5), which suggests that part of these ancestries arrived from outside Italy in historical times (*21*), but also that these components were already present in Italian groups at the time of these admixture events. The timing of the admixture events and the sources involved differ between Northern and Southern Italian clusters, pointing to different admixture histories for the two areas. Episodes of gene flow were also detected in Sardinia, combining signals from both the African continent and North-West Europe. MALDER results for the more recent episodes replicated the admixture pattern identified by GT (fig. S8 and data file S5).

The Neanderthal legacy across Italy and Europe

The variation in ancestry composition reported across Italy and Europe is expected to influence other aspects of the genetic profiles of European populations, including the presence of archaic genetic material (*4*). We investigated the degree of Neanderthal ancestry in Italian and other Eurasian populations (data file S1) by focusing on SNPs tagging Neanderthal introgressed regions (*24*). SNPs were pruned for linkage disequilibrium (LD), and a final set of 3969 SNPs was used to estimate the number of Neanderthal alleles in samples genotyped for the Infinium Omni2.5-8 Illumina Beadchip. Asian and Northern European populations had significantly more Neanderthal alleles than European and Southern European groups, as previously reported (*25*, *26*), with significant differences also highlighted within Italy (Fig. 4, A and B). Contributions from the African groups may have influenced these patterns, particularly in Southern European populations (*2*) (Figs. 2 and 3). However, differences within Europe and Italy were still present once individuals belonging to clusters with African contributions were removed (see fig. S9, A and B, Materials and Methods, and the Supplementary Materials). Ancient samples have been reported to differ in their amount of Neanderthal DNA partially because of the variation in proportion of the "Basal Eurasian" lineage, which harbors only a negligible fraction of Neanderthal ancestry (*4*). Consistent with this (*4*), we found the estimated amounts of Basal Eurasian and Neanderthal to be negatively correlated across modern-day European clusters and populations (Fig. 4C and fig. S9, E to H). Our estimation of Basal Eurasian ancestry might be affected by non-Eurasian contributions, which could therefore be partially responsible for the observed Neanderthal patterns. The correlation between the Basal Eurasian component and Neanderthal allele sharing was still present once populations with African and East Asian contributions (Fig. 3 and fig. S8) were removed (Fig. 2 and fig. S9D). These results suggest that the pattern of Neanderthal ancestry observed in Italy and across Europe has been shaped by different factors: variation in the amount of Basal Eurasian present in ancient sources, variation in the ancient ancestry composition of modern samples, and variation in the historical contribution from Africa and East Asia.

The variation in Neanderthal ancestry was also evident when SNP frequencies were evaluated. A total of 144 SNPs were identified among the Neanderthal-tag SNPs showing the largest differences in allelic frequency in genome-wide comparisons across Eurasian and African populations (Materials and Methods; Supplementary Materials, Neanderthal-tag SNPs within the top 1% of the genome-wide

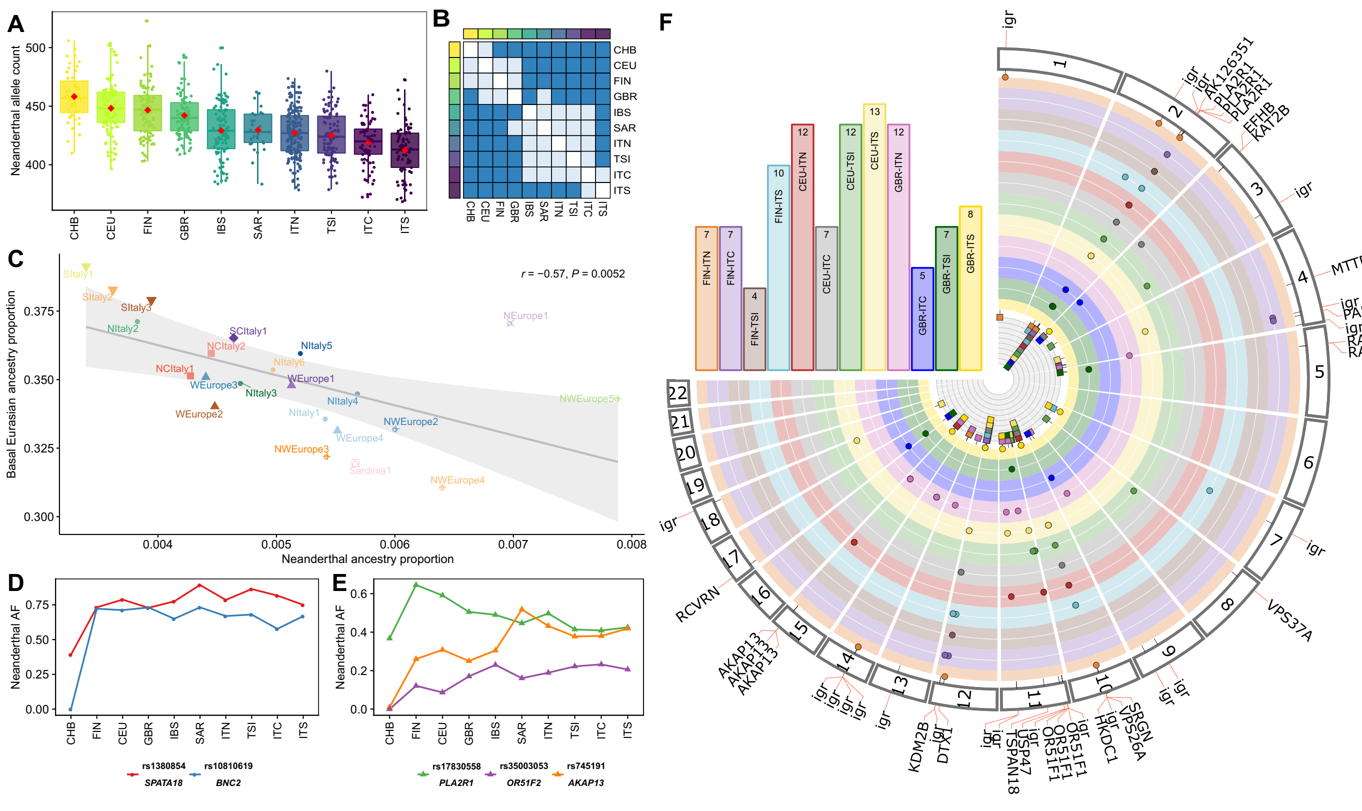

Fig. 4. Neanderthal ancestry distribution in Eurasian populations. (**A**) Neanderthal allele counts in individuals from Eurasian populations, sorted by median values on 3969 LD-pruned Neanderthal tag-SNPs. CEU, Utah Residents with Northern and Western European ancestry; GBR, British in England and Scotland; FIN, Finnish in Finland; IBS, Iberian Population in Spain; TSI, Tuscans from Italy; ITN, Italians from Northern Italy; ITC, Italians from Central Italy; ITS, Italians from Southern Italy; SAR, Italians from Sardinia; CHB, Han Chinese. (**B**) Matrix of significances based on Wilcoxon rank sum test between pairs of populations including (lower triangular matrix) and removing (upper) outliers (see Materials and Methods and Supplementary Materials; dark blue, adjusted $P < 0.05$; light blue, adjusted $P > 0.05$). Colored squares at the sides of the heatmap refer to the populations compared, as per Fig. 4A. (**C**) Correlation between Neanderthal ancestry proportions and the amount of Basal Eurasian ancestry in European clusters (see Materials and Methods and the Supplementary Materials). (**D** and **E**) Neanderthal allele frequency (AF) for selected SNPs within the indicated genes: (D) high-frequency alleles in Europe and (E) North-South Europe divergent alleles. (**F**) Comparisons between Northern European and Italian populations (excluding Sardinia). Bars refer to comparison for reported pairs of populations; the number of NTT SNPs is reported within bars. Each section of the circos represents a tested chromosome; points refer to NTT SNPs. Colors are the same as for bars; igr, intergenic region variant.

distributions of each of the 55 pairwise population comparisons—NTT SNPs; fig. S9I, data file S6). The top 1% of each distribution was significantly depleted in Neanderthal SNPs (Materials and Methods, Supplementary Materials, and data file S6), in agreement with a scenario of Neanderthal mildly deleterious variants being removed more efficiently in human populations (*27*).

The 50 genes containing NTT SNPs were enriched for phenotypes related to facial morphology, body size, metabolism, and muscular diseases (Materials and Methods, Supplementary Materials, and data file S6). A total of 34 NTT SNPs were found to have at least one known phenotypic association (data file S6) (*28*). Among these, we found Neanderthal alleles associated with increased gene expression in testis and in skin after sun exposure (SNPs within the *IP6K3* and *ITPR3* genes), susceptibility to cardiovascular and renal conditions (*AGTR1*), and Brittle cornea syndrome (*PRDM5*) (*24*). NTT SNPs between European and Asian/African populations included previously reported variants in *BNC2 (29)* and *SPATA18* genes (*30*, *31*) (Materials and Methods, Supplementary Materials, and Fig. 4D), while 80 NTT SNPs were involved in at least one comparison between Northern (CEU, GBR, and FIN) and Southern European populations (IBS and Italian groups). Among these SNPs, three mapped to the Neanderthal introgressed haplotype hosting the *PLA2R1* gene (*32*), the archaic allele at these positions reaching frequencies of at least 43% in Northern European and at most 35% in Southern European populations (Fig. 4, E and F). Ten SNPs showed an opposite frequency gradient: seven mapped to one Neanderthal-introgressed region spanning the *OR51F1*, *OR51F2*, and *OR52R1* genes (Fig. 4, E and F), and the other three identified regions hosting the *AKAP13* gene, within one of the high-frequency European Neanderthal introgressed haplotypes recently reported (Fig. 4, E and F) (*31*). The locus-specific differences explored here might be the result of selection-based evolutionary dynamics combined with the demographic events that shaped the more general genome-wide patterns present across Italy and Europe.

DISCUSSION

The pattern of variation reported across Italian groups appears geographically structured across three main regions in Italy: Southern, Northern, and Sardinia. Similarly to Europe, the genetic structure reflects isolation by distance following population movements since prehistoric times (*1*) and historical admixture from the fringes of the continent (*2*). The analysis of both modern and ancient data suggests that in Italian populations, ancestries related to CHG and EHG derive from at least two sources. One is the well-characterized steppe (SBA) signature associated with nomadic groups from the Pontic-Caspian steppes. This component reached Italy from mainland Europe at least as long ago as the Bronze Age, as suggested by its presence in Bell Beaker samples from North Italy (data file S4). The other contribution is ultimately associated with CHG ancestry, as previously suggested (*21*), and predominantly affected Southern Italy, where it represents a substantial component of the ancestry profile of local modern populations. Although the details of the origins of this signature are still uncharacterized, it may have been present as early as the Bronze Age in Southern Italy (data file S4). The very low presence of CHG signatures in Sardinia and in older Italian samples (Remedello and Iceman), but its occurrence in modern-day Southern Italians, might be explained by different scenarios not mutually exclusive: (i) population structure among early foraging groups across Italy, reflecting different affinities to CHG; (ii) the presence in Italy of different Neolithic contributions, characterized by a different proportion of CHG-related ancestry; (iii) the combination of a post-Neolithic, prehistoric CHG-enriched contribution with a previous AN-related Neolithic layer; and (iv) a substantial historical contribution from Southeastern Europe across the whole of Southern Italy.

No major structure has been highlighted so far in pre-Neolithic Italian samples (*6*). An arrival of the CHG-related component in Southern Italy from the Southern part of the Balkan Peninsula, including the Peloponnese, is compatible with the identification of genetic corridors linking the two regions (Fig. 1E) (*9*) and the presence of Southern European ancient signatures in Italy (Fig. 2). The temporal appearance of CHG signatures in Anatolia and Southern East Europe in the Late Neolithic/Bronze Age suggests its relevance for post-Neolithic contributions (*33*). Our results suggest contributions from ancestries additional to the three "canonical" ones considered so far in the literature (WHG, AN, and SBA). The differential distribution of these ancestries contributed to the differentiation observed between Northern and Southern Italian clusters. Additional analyses of aDNA samples from around this time in Italy are expected to clarify what ancient scenario might best support the structure related to ancient ancestry composition presented here.

Historical events possibly involving continental groups at the end of the Roman Empire and African contributions following the establishment of Arab kingdoms in Southern Europe around 1300 to 1200 ya (*2*, *13*, *21*, *22*, *34*) played a role in further shaping the ancestry profiles and population structure of Italians (Fig. 3). In particular, African contributions might have contributed to the increased diversity detected among clusters in Southern Italy and Sardinia (Fig. 3 and data file S2) (*13*).

Significantly, despite Sardinia being confirmed as the most closely related population to Early European Neolithic farmers (Fig. 2, D and I), there is no evidence for a simple genetic continuity between the two groups. Populations in Sardinia were not completely isolated and, like the rest of Italy, experienced historical episodes of gene flow (Figs. 2 and 3 and data file S4) that contributed to the further dispersal of ancient ancestries and the introduction of other components, including African ones.

It has been previously reported that variation in effective population size might explain differences in the amount of Neanderthal DNA detected in European and Asian populations (*24*, *26*). Additional Neanderthal introgression events in Asia and gene flow from populations with lower Neanderthal ancestry in Europe may provide further explanations for differences in Neanderthal occurrence across populations. The spatial heterogeneity of Neanderthal legacy within Europe reported here appears to be the result of ancient and historical events that brought together in different combinations groups harboring different amounts of Neanderthal genetic material. While these events have shaped the overall continental distribution of Neanderthal DNA, locus-specific differences in the occurrence of Neanderthal alleles are also expected to reflect selective pressures acting on these variants since their introgression in the populations (*27*).

The Bronze Age migration from the Pontic-Caspian steppe region has been linked to the arrival of the Indo-European languages in mainland Europe. Our identification of an additional substantial component in Italy possibly arriving in the Bronze Age raises the possibility of multiple Indo-European waves into the continent. Similarly, the persistence in Italy of non-Indo European languages into historical times (e.g., Etruscan) could be linked to reduced SBA penetration along the peninsula. These associations, while fascinating, will require dedicated and multidisciplinary approaches to be properly explored and validated. In particular, ancient samples spanning diachronic, geographic, and cultural transects in the Italian peninsula and nearby regions will need to be analyzed to complement the interpretative framework proposed here for the processes that have shaped Southern European variation.

MATERIALS AND METHODS

Analysis of modern samples

Dataset

Two hundred twenty-two samples are presented here. Of these, 167 Italians and 6 Albanians were specifically selected and sequenced for this project with two versions (1.2 and 1.3) of the Infinium Omni2.5-8 Illumina beadchip, while 49 Italians and Europeans were genotyped with the Human660W-Quad BeadChip in the frame of another research (data file S1) *(35)*. Two separate worldwide datasets were prepared. The FMD included 4852 samples (*2*, *12*, *22*, *36*–*51*) (1589 Italians) and 218,725 SNPs genotyped with Illumina arrays; the HDD contained 1651 samples (*12*, *36*, *38*, *40*, *44*, *47*, *50*, *51*) (524 Italians) and 591,217 SNPs genotyped with the Illumina Omni array (Supplementary Materials).

The merging, the removal of ambiguous C/G and A/T and triallelic markers, the exclusion of related individuals, and the discarding of SNPs in LD were performed using PLINK 1.9 (*52*). Only autosomal markers were considered.

Haplotype analysis (CP and fS)

We investigated patterns of genetic differentiation in Italy by exploring the information provided by SNP-based haplotypes. Phased haplotypes were generated using SHAPEIT 2 (*53*) and by applying the HapMap b37 genetic map.

CP was used to generate a matrix of recipient individuals "painted" as a combination of donor samples (copying vector). Three runs of CP were done for each dataset generating three different outputs: (i) a matrix of all the individuals painted as a combination of all the individuals, for cluster identification and GT analysis; (ii) a matrix of all Italians as a combination of all Italians, for F_{ST} analysis; and (iii) a matrix of all the samples as a combination of all the other samples but excluding Italians, for noItaly GT analysis. The median numbers of SNPs per painted fragment were 13.7 and 31.6 for the FMD and HDD, respectively.

Clusters were inferred using fS. After an initial search based on the "greedy" mode, the dendrogram was processed by visual inspection (*2*, *20*) according to the geographical origin of the samples. The robustness of the cluster was obtained by processing the Markov Chain Monte Carlo (MCMC) pairwise coincidence matrix (Supplementary Materials).

Cluster self-copy analysis

Recently admixed individuals were identified as those copying from members of the cluster to which they belong less than the amount of cluster self-copying for samples with all four grandparents from the same geographic region (Supplementary Materials).

F_{ST} and TVD within and between Italian clusters

To have a comprehensive overview of the genetic diversity in Italy, we estimated the pairwise F_{ST} within Italian clusters using smartpca implemented in EIGENSOFT (*54*). TVD estimates were obtained using the TVD function (*11*) on the CP chunklength matrix.

Pairwise genetic diversity among clusters was computed estimating pairwise F_{ST} and TVD metrics on Italian clusters belonging to the same or to different macroareas. In detail, the NItaly macroarea comprised clusters named as NItaly and NCItaly; the SItaly macroarea included SItaly, SCItaly, and Sicily named clusters; while the Sardinia macroarea included only the Sardinia clusters.

F_{ST} estimates among clusters

Pairwise F_{ST} estimates among newly generated Italian clusters and originally generated European clusters (Supplementary Materials) were inferred using the smartpca software implemented in the EIGENSOFT package (*54*). Comparisons between the F_{ST} distributions were performed using a Wilcoxon rank sum test in the R programming language environment.

Principal components analysis

PCA was performed on the CP chunkcount matrix (Supplementary Materials) and was generated using the *prcomp()* function on R software. Allele frequency PCA was performed using smartpca implemented in EIGENSOFT (*54*) after pruning the datasets for LD.

Procrustes analysis

To validate the correlation observed between the haplotype-based PCA (Fig. 1) and the cluster distribution (Fig. 1C), we performed a symmetric Procrustes analysis with 100,000 permutations. In detail, we used the values of the first two PCs of the PCA estimated on the CP chunkcount matrix generated using only Italian individuals for which the place of origin (administrative region) was available along with the geographic coordinates of the administrative center ("capoluogo di regione") of the region to which they were assigned to on the basis of available information (data file S1).

Characterization of the migration landscape (EEMS analysis)

We highlighted the spatial patterns of genetic differentiation by EEMS analysis (*19*). This was performed estimating the average pairwise distances between populations using the bed2diffs tool, and the resulting output was visualized using the Reems package (*19*).

ADMIXTURE analysis

ADMIXTURE 1.3.0 software (*55*) was used, performing 10 different runs using a random seed. The results were combined with CLUMPAK (*56*) using the largeKGreedy algorithm and random input orders with 10,000 repeats. *Distruct for many K's* implemented in CLUMPAK (*56*) was then used to identify the best alignment of CLUMPP results. Results were processed using the R statistical software.

The time and the sources of admixture events (GT and MALDER analyses)

Times of admixture events were investigated using the GLOBETROTTERv2 software. GT was utilized using two approaches: complete and nonlocal (referred to as noItaly; Supplementary Materials) in default modality (*2*, *11*). The difference between the two approaches was the inclusion or the exclusion respectively of all the Italian clusters as donors in the CP matrix used as the input file. To improve the precision of the admixture signals, the "null.ind: 1" parameter was set (*2*). Unclear signals were corrected using the default parameters, and a total of 100 bootstraps were performed. MALDER (*57*) uses allele frequencies to dissect the time of admixture signals. The best amplitude was identified and used to calculate a *Z* score (Supplementary Materials). A *Z* score equal to or lower than 2 identifies not significantly different amplitude curves (Supplementary Materials) (*58*).

Sources for both GT and MALDER were grouped in different ancestries as indicated in the legend of Fig. 3 and fig. S8. The expression [1950 − (*g* + 1)*29], where *g* is the number of generation, was used to convert the GT and MALDER results into years.

Analyses including ancient samples

Dataset

To explore the extent to which the European and Italian genetic variation has been shaped by ancient demographic events, we merged modern samples from FMD with 63 ancient samples selected from recent studies (data file S1) (*4*, *5*, *8*, *23*, *33*, *59–61*).

Principal component analysis (PCA)

To visualize the genetic affinities of ancient and modern samples, we performed two PCAs with the EIGENSOFT (*54*) smartpca software and the "*lsqproject*" and "*shrinkmode*" option, projecting the ancient samples onto components inferred from modern European, West Asian, and Caucasian individuals and then only on the modern European clusters. To evaluate the potential impact of DNA damage in calling variants from aDNA samples, we repeated the PCA with the 63 ancient samples and the modern European, Caucasian, and West Asian samples by removing transition polymorphisms and recorded significant correlations for the localization of ancient samples along PC1 and PC2 (Pearson $r > 0.99$, $P < 0.05$).

ADMIXTURE analysis

We explored the genetic relationships between modern and ancient samples by projecting the ancient samples on the previously inferred ancestral allele frequencies from 10 ADMIXTURE (*55*) runs on modern samples (see the "Analysis of modern samples" section and the Supplementary Materials). We used CLUMPP for merging the resulting matrices and *distruct* for the visualization in CLUMPAK program (*56*).

D-statistics

We tested for admixture using the D-statistics as implemented in the qpDstat tool in the software ADMIXTOOLS v4.2 (*62*). We performed the D-statistics analyses evaluating the relationship of the Italian cluster with AN, ABA, and SBA. In detail, we performed the D-statistics D(Ita1,Ita2,AN/ABA/SBA,Mbuti), where Ita1 and Ita2 are the different clusters composed mainly of Italian individuals as inferred by fS.

CP/NNLS analysis

We used an approach based on the CP software (*10*) and a slight adaptation of the NNLS function (*11*, *20*) to estimate the proportions of the genetic contributions from ancient population to our modern clusters. We ran CP using the unlinked mode (*59*) with the same Ne and θ parameters of the modern dataset, painting both modern and ancient individuals and using modern samples as donors (*59*, *60*). We analyzed the output of CP by solving an appropriately formulated NNLS problem, reconstructing the modern clusters in terms of the ancients. We applied this combined approach on different sets of ancient samples (ultimate and proximate sources).

Goodness of fit was measured evaluating the residuals of the NNLS analysis. In detail, we focused on the proximate sources and compared the sum of squared residuals when ABA or SBA was included/excluded as putative sources. Furthermore, for the ultimate and proximate analyses, we estimated the SEs by applying a weighted jackknife bootstrap (data file S3), estimating the mixture profile removing one chromosome at time and averaging the values taking into account the total number of markers analyzed for iteration (*58*).

qpAdm analysis

We used the ancestral reconstruction method qpAdm, which harnesses different relationships of populations related to a set of outgroups (e.g., f4[Target, O1, O2, O3]) (*1*) to model the ancestry composition of modern and ancient Italian samples as different combinations of ancient sources. In detail, for each tested cluster of the FMD and HDD, we evaluated all the possible combinations of N "left" sources with $N = \{2..5\}$ and one set of right/left outgroups (see the Supplementary Materials) (*8*).

For each of the tested combinations, we used qpWave to evaluate whether the set of chosen outgroups is able to (i) discriminate the combinations of sources and (ii) establish if the target may be explained by the sources. We used a *P* value threshold of 0.01. Last, we used qpAdm to infer the admixture proportions and reported it and the associated SEs in data file S4. In addition, we performed the same analysis with $N = \{2..4\}$ for Iceman, Remedello, and Bell Beaker individuals from Sicily and North Italy (data file S4).

Archaic contribution

Dataset

We assembled an additional HDD by retaining only samples genotyped on the Illumina Infinium Omni2.5-8 BeadChip from our larger modern dataset. In particular, we included seven populations from the 1000 Genomes Project: the five European populations (Northern European from Utah, CEU; England, GBR; Finland, FIN; Spain, IBS; Italy from Tuscany, TSI), one from Asia (Han Chinese, CHB), and one from Africa (Yoruba from Nigeria, YRI). We also retained 466 Italian samples, whose four grandparents were born in the same Italian region. The Italian samples were broadly clustered according to their geographical origin into Northern, Central, and Southern Italians, and Sardinians, while TSI samples from the 1000 Genome Project formed a separate cluster (data file S1).

From this dataset, we extracted 7164 Neanderthal SNPs tagging Neanderthal introgressed regions (*24*). To select which allele was inherited from Neanderthals, we chose the one from the Altai Neanderthal (*26*) genome when it was homozygous and the minor allele in YRI when it was heterozygous.

Number of Neanderthal alleles in present-day human populations

To provide a direct (relative) estimate of the Neanderthal DNA present in different individuals, we initially pruned Neanderthal tagging SNPs in LD and counted the number of Neanderthal alleles considering all the tag-SNP across all samples. Then, we compared the distribution of Neanderthal allele counts across populations with the two-sample Wilcoxon rank sum test. We repeated the same analyses after removing outlier individuals.

Basal Eurasian ancestry and Neanderthal contribution

To infer the proportion of Basal Eurasian present in European populations and investigate its impact in shaping variation in the Neanderthal legacy across populations (*4*, *5*), we used the f4 ratio implemented in the ADMIXTOOLS package (*62*) in the form f4(Target, Loschbour, Ust_Ishim, Kostenki14)/f4(Mbuti, Loschbour, Ust_Ishim, Kostenki14). We repeated this approach to infer the Neanderthal ancestry in the form f4(Mbuti, Chimp Target, Altai)/f4(Mbuti, Chimp, Dinka, Altai) (fig. S9, E to H). We then performed the same analyses by grouping the modern individuals according to the CP/fS inferred clusters (see the "Analysis of modern samples" section) and retained only clusters with at least 10 samples (Fig. 4C).

African ancestry and Neanderthal legacy

The impact of African contributions in shaping the amount of Neanderthal occurrence was evaluated by exploring how the removal of the clusters showing African gene flow as detected by the GT analysis (Fig. 3), and of individuals belonging to these clusters, affected the correlation between Basal Eurasian/Neanderthal estimates and the degree of population differentiation in the amount of Neanderthal alleles, respectively (Supplementary Materials and fig. S9, A to D).

Comparison of Neanderthal allele frequencies across modern populations

We explored significant differences in the frequencies of Neanderthal alleles across populations by computing the allele frequency differences for every SNP for each of the possible pairs of the 11 populations in our dataset, thus obtaining 55 distributions (see the Supplementary Materials). Then, the NTT SNPs, i.e., the Neanderthal-tag SNPs in the top 1% of each distribution, were selected (data file S6).

The biological implications of Neanderthal introgression

Given the list of genes overlapping the Neanderthal introgressed regions harboring the NTT SNPs and the list of genes directly harboring the NTT SNPs, we performed different enrichment tests with the online tool EnrichR (*63*). Particularly, we searched for significant enrichments compared to the human genome using the EnrichR collection of database, e.g., dbGaP, Panther 2016, HPO, and KEGG 2016 (data file S6). We then investigated known direct associations between the Neanderthal alleles of the NTT SNPs and phenotypes by looking in the GWAS and PheWAS catalogs (https://phewascatalog.org/phewas) and applying the PheGenI tool (https://www.ncbi.nlm.nih.gov/gap/phegeni) (data file S6). We used the circos representation as in Kanai *et al.* (*64*) to highlight different sets of NTT SNPs (Fig. 4F).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/9/eaaw3492/DC1

Fig. S1. Geographic location of populations included in FMD and HDD, and fineStructure dendrogram for all the 4,852 (FMD) and 1,641 (HDD) samples.

Fig. S2. Allele frequency PCA (genotype based) and individual-level ADMIXTURE of modern samples.

Fig. S3. "Cluster self-copy" analysis and PCA with admixed Italian individuals.

Fig. S4. Results of the EEMS analysis on Italy-only populations.

Fig. S5. CP/NNLS and qpAdm results for different sets of ancient sources for all modern clusters.

Fig. S6. D-statistics analyses.

Fig. S7. PCA and ADMIXTURE analyses of 63 ancient samples.

Fig. S8. GT and MALDER analyses for all the Eurasian and North African clusters.

Fig. S9. Neanderthal ancestry distribution in Eurasian population and its relationship with African admixture and Basal Eurasian ancestry.

Data file S1. Modern and ancient samples used in this study.

Data file S2. Cluster self-copy analysis.

Data file S3. Weighted jackknife bootstraps.

Data file S4. qpAdm results.

Data file S5. GT and MALDER results.

Data file S6. NTT SNPs (Neanderthal-Tag SNPs within the top 1% of the genome-wide distributions of each of the 55 pairwise population comparisons).

REFERENCES AND NOTES

1. W. Haak, I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, S. Nordenfelt, E. Harney, K. Stewardson, Q. Fu, A. Mittnik, E. Bánffy, C. Economou, M. Francken, S. Friederich, R. G. Pena, F. Hallgren, V. Khartanovich, A. Khokhlov, M. Kunst, P. Kuznetsov, H. Meller, O. Mochalov, V. Moiseyev, N. Nicklisch, S. L. Pichler, R. Risch, M. A. Rojo Guerra, C. Roth, A. Szécsényi-Nagy, J. Wahl, M. Meyer, J. Krause, D. Brown, D. Anthony, A. Cooper, K. W. Alt, D. Reich, Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
2. G. B. J. Busby, G. Hellenthal, F. Montinaro, S. Tofanelli, K. Bulayeva, I. Rudan, T. Zemunik, C. Hayward, D. Toncheva, S. Karachanak-Yankova, D. Nesheva, P. Anagnostou, F. Cali, F. Brisighelli, V. Romano, G. Lefranc, C. Buresi, J. Ben Chibani, A. Haj-Khelil, S. Denden, R. Ploski, P. Krajewski, T. Hervig, T. Moen, R. J. Herrera, J. F. Wilson, S. Myers, C. Capelli, The role of recent admixture in forming the contemporary West Eurasian genomic landscape. *Curr. Biol.* **25**, 2518–2526 (2015).
3. J. Novembre, T. Johnson, K. Bryc, Z. Kutalik, A. R. Boyko, A. Auton, A. Indap, K. S. King, S. Bergmann, M. R. Nelson, M. Stephens, C. D. Bustamante, Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
4. I. Lazaridis, D. Nadel, G. Rollefson, D. C. Merrett, N. Rohland, S. Mallick, D. Fernandes, M. Novak, B. Gamarra, K. Sirak, S. Connell, K. Stewardson, E. Harney, Q. Fu, G. Gonzalez-Fortes, E. R. Jones, S. A. Roodenberg, G. Lengyel, F. Bocquentin, B. Gasparian, J. M. Monge, M. Gregg, V. Eshed, A. S. Mizrahi, C. Meiklejohn, F. Gerritsen, L. Bejenaru, M. Blüher, A. Campbell, G. Cavalleri, D. Comas, P. Froguel, E. Gilbert, S. M. Kerr, P. Kovacs, J. Krause, D. McGettigan, M. Merrigan, D. A. Merriwether, S. O'Reilly, M. B. Richards, O. Semino, M. Shamoon-Pour, G. Stefanescu, M. Stumvoll, A. Tönjes, A. Torroni, J. F. Wilson, L. Yengo, N. A. Hovhannisyan, N. Patterson, R. Pinhasi, D. Reich, Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
5. Q. Fu, C. Posth, M. Hajdinjak, M. Petr, S. Mallick, D. Fernandes, A. Furtwängler, W. Haak, M. Meyer, A. Mittnik, B. Nickel, A. Peltzer, N. Rohland, V. Slon, S. Talamo, I. Lazaridis, M. Lipson, I. Mathieson, S. Schiffels, P. Skoglund, A. P. Derevianko, N. Drozdov, V. Slavinsky, A. Tsybankov, R. G. Cremonesi, F. Mallegni, B. Gély, E. Vacca, M. R. González Morales, L. G. Straus, C. Neugebauer-Maresch, M. Teschler-Nicola, S. Constantin, O. T. Moldovan, S. Benazzi, M. Peresani, D. Coppola, M. Lari, S. Ricci, A. Ronchitelli, F. Valentin, C. Thevenet, K. Wehrberger, D. Grigorescu, H. Rougier, I. Crevecoeur, D. Flas, P. Semal, M. A. Mannino, C. Cupillard, H. Bocherens, N. J. Conard, K. Harvati, V. Moiseyev, D. G. Drucker, J. Svoboda, M. P. Richards, D. Caramelli, R. Pinhasi, J. Kelso, N. Patterson, J. Krause, S. Pääbo, D. Reich, The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
6. E. R. Jones, G. Gonzalez-Fortes, S. Connell, V. Siska, A. Eriksson, R. Martiniano, R. L. McLaughlin, M. Gallego Llorente, L. M. Cassidy, C. Gamba, T. Meshveliani, O. Bar-Yosef, W. Müller, A. Belfer-Cohen, Z. Matskevich, N. Jakeli, T. F. G. Higham, M. Currat, D. Lordkipanidze, M. Hofreiter, A. Manica, R. Pinhasi, D. G. Bradley, Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
7. D. W. Anthony, *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World* (Princeton University Press, 2010).
8. I. Lazaridis, A. Mittnik, N. Patterson, S. Mallick, N. Rohland, S. Pfrengle, A. Furtwängler, A. Peltzer, C. Posth, A. Vasilakis, P. J. P. McGeorge, E. Konsolaki-Yannopoulou, G. Korres, H. Martlew, M. Michalodimitrakis, M. Özsait, N. Özsait, A. Papathanasiou, M. Richards, S. A. Roodenberg, Y. Tzedakis, R. Arnott, D. M. Fernandes, J. R. Hughey, D. M. Lotakis, P. A. Navas, Y. Maniatis, J. A. Stamatoyannopoulos, K. Stewardson, P. Stockhammer, R. Pinhasi, D. Reich, J. Krause, G. Stamatoyannopoulos, Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218 (2017).
9. P. Paschou, P. Drineas, E. Yannaki, A. Razou, K. Kanaki, F. Tsetsos, S. S. Padmanabhuni, M. Michalodimitrakis, M. C. Renda, S. Pavlovic, A. Anagnostopoulos, J. A. Stamatoyannopoulos, K. K. Kidd, G. Stamatoyannopoulos, Maritime route of colonization of Europe. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9211–9216 (2014).
10. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense haplotype data. *PLOS Genet.* **8**, e1002453 (2012).
11. S. Leslie, B. Winney, G. Hellenthal, D. Davison, A. Boumertit, T. Day, K. Hutnik, E. C. Royrvik, B. Cunliffe; Wellcome Trust Case Control Consortium 2; International Multiple Sclerosis Genetics Consortium, D. J. Lawson, D. Falush, C. Freeman, M. Pirinen, S. Myers, M. Robinson, P. Donnelly, W. Bodmer, The fine-scale genetic structure of the British population. *Nature* **519**, 309–314 (2015).
12. G. Fiorito, C. Di Gaetano, S. Guarrera, F. Rosa, M. W. Feldman, A. Piazza, G. Matullo, The Italian genome reflects the history of Europe and the Mediterranean basin. *Eur. J. Hum. Genet.* **24**, 1056–1062 (2016).
13. M. Sazzini, G. A. Gnecchi Ruscone, C. Giuliani, S. Sarno, A. Quagliariello, S. De Fanti, A. Boattini, D. Gentilini, G. Fiorito, M. Catanoso, L. Boiardi, S. Croci, P. Macchioni, V. Mantovani, A. M. Di Blasio, G. Matullo, C. Salvarani, C. Franceschi, D. Pettener, P. Garagnani, D. Luiselli, Complex interplay between neutral and adaptive evolution shaped differential genomic background and disease susceptibility along the Italian peninsula. *Sci. Rep.* **6**, 32513 (2016).
14. G. Athanasiadis, J. Y. Cheng, B. J. Vilhjálmsson, F. G. Jørgensen, T. D. Als, S. Le Hellard, T. Espeseth, P. F. Sullivan, C. M. Hultman, P. C. Kjærgaard, M. H. Schierup, T. Mailund, Nationwide Genomic Study in Denmark Reveals Remarkable Population Homogeneity. *Genetics* **204**, 711–722 (2016).
15. R. P. Byrne, R. Martiniano, L. M. Cassidy, M. Carrigan, G. Hellenthal, O. Hardiman, D. G. Bradley, R. L. McLaughlin, Insular Celtic population structure and genomic footprints of migration. *PLOS Genet.* **14**, e1007152 (2018).
16. C. Bycroft, C. Fernandez-Rozadilla, C. Ruiz-Ponte, I. Quintela, Á. Carracedo, P. Donnelly, S. Myers, Patterns of genetic differentiation and the footprints of historical migrations in the Iberian Peninsula. *Nat. Commun.* **10**, 551 (2019).
17. G. Destro Bisol, P. Anagnostou, C. Batini, C. Battaggia, S. Bertoncini, A. Boattini, L. Caciagli, M. C. Caló, C. Capelli, M. Capocasa, L. Castrí, G. Ciani, V. Coia, L. Corrias, F. Crivellaro, M. E. Ghiani, D. Luiselli, C. Mela, A. Melis, V. Montano, G. Paoli, E. Sanna, F. Rufo, M. Sazzini, L. Taglioli, S. Tofanelli, A. Useli, G. Vona, D. Pettener, Italian isolates today: Geographic and linguistic factors shaping human biodiversity. *J. Anthropol. Sci.* **86**, 179–188 (2008).
18. M. Capocasa, P. Anagnostou, V. Bachis, C. Battaggia, S. Bertoncini, G. Biondi, A. Boattini, I. Boschi, F. Brisighelli, C. M. Caló, M. Carta, V. Coia, L. Corrias, F. Crivellaro, S. De Fanti, V. Dominici, G. Ferri, P. Francalacci, Z. A. Franceschi, D. Luiselli, L. Morelli, G. Paoli, O. Rickards, R. Robledo, D. Sanna, E. Sanna, S. Sarno, L. Sineo, L. Taglioli, G. Tagarelli, S. Tofanelli, G. Vona, D. Pettener, G. Destro Bisol, Linguistic, geographic and genetic

isolation: A collaborative study of Italian populations. *J. Anthropol. Sci.* **92**, 201–231 (2014).

19. D. Petkova, J. Novembre, M. Stephens, Visualizing spatial population structure with estimated effective migration surfaces. *Nat. Genet.* **48**, 94–100 (2016).
20. F. Montinaro, G. B. J. Busby, V. L. Pascali, S. Myers, G. Hellenthal, C. Capelli, Unravelling the hidden ancestry of American admixed populations. *Nat. Commun.* **6**, 6596 (2015).
21. S. Sarno, A. Boattini, L. Pagani, M. Sazzini, S. De Fanti, A. Quagliariello, G. A. Gnecchi Ruscone, E. Guichard, G. Ciani, E. Bortolini, C. Barbieri, E. Cilli, R. Petrilli, I. Mikerezi, L. Sineo, M. Vilar, S. Wells, D. Luiselli, D. Pettener, Ancient and recent admixture layers in Sicily and Southern Italy trace multiple migration routes along the Mediterranean. *Sci. Rep.* **7**, 1984 (2017).
22. G. Hellenthal, G. B. J. Busby, G. Band, J. F. Wilson, C. Capelli, D. Falush, S. Myers, A genetic atlas of human admixture history. *Science* **343**, 747–751 (2014).
23. I. Olalde, S. Brace, M. E. Allentoft, I. Armit, K. Kristiansen, T. Booth, N. Rohland, S. Mallick, A. Szécsényi-Nagy, A. Mittnik, E. Altena, M. Lipson, I. Lazaridis, T. K. Harper, N. Patterson, N. Broomandkhoshbacht, Y. Diekmann, Z. Faltyskova, D. Fernandes, M. Ferry, E. Harney, P. de Knijff, M. Michel, J. Oppenheimer, K. Stewardson, A. Barclay, K. W. Alt, C. Liesau, P. Ríos, C. Blasco, J. V. Miguel, R. M. García, A. A. Fernández, E. Bánffy, M. Bernabò-Brea, D. Billoin, C. Bonsall, L. Bonsall, T. Allen, L. Büster, S. Carver, L. C. Navarro, O. E. Craig, G. T. Cook, B. Cunliffe, A. Denaire, K. E. Dinwiddy, N. Dodwell, M. Ernée, C. Evans, M. Kuchařík, J. F. Farré, C. Fowler, M. Gazenbeek, R. G. Pena, M. Haber-Uriarte, E. Haduch, G. Hey, N. Jowett, T. Knowles, K. Massy, S. Pfrengle, P. Lefranc, O. Lemercier, A. Lefebvre, C. H. Martínez, V. G. Olmo, A. B. Ramírez, J. L. Maurandi, T. Majó, J. I. McKinley, K. McSweeney, B. G. Mende, A. Mod, G. Kulcsár, V. Kiss, A. Czene, R. Patay, A. Endrődi, K. Köhler, T. Hajdu, T. Szeniczey, J. Dani, Z. Bernert, M. Hoole, O. Cheronet, D. Keating, P. Velemínský, M. Dobeš, F. Candilio, F. Brown, R. F. Fernández, A.-M. Herrero-Corral, S. Tusa, E. Carnieri, L. Lentini, A. Valenti, A. Zanini, C. Waddington, G. Delibes, E. Guerra-Doce, B. Neil, M. Brittain, M. Luke, R. Mortimer, J. Desideri, M. Besse, G. Brücken, M. Furmanek, A. Hałuszko, M. Mackiewicz, A. Rapiński, S. Leach, I. Soriano, K. T. Lillios, J. L. Cardoso, M. P. Pearson, P. Włodarczak, T. D. Price, P. Prieto, P.-J. Rey, R. Risch, M. A. Rojo Guerra, A. Schmitt, J. Serralongue, A. M. Silva, V. Smrčka, L. Vergnaud, J. Zilhão, D. Caramelli, T. Higham, M. G. Thomas, D. J. Kennett, H. Fokkens, V. Heyd, A. Sheridan, K.-G. Sjögren, P. W. Stockhammer, J. Krause, R. Pinhasi, W. Haak, I. Barnes, C. Lalueza-Fox, D. Reich, The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196 (2018).
24. C. N. Simonti, B. Vernot, L. Bastarache, E. Bottinger, D. S. Carrell, R. L. Chisholm, D. R. Crosslin, S. J. Hebbring, G. P. Jarvik, I. J. Kullo, R. Li, J. Pathak, M. D. Ritchie, D. M. Roden, S. S. Verma, G. Tromp, J. D. Prato, W. S. Bush, J. M. Akey, J. C. Denny, J. A. Capra, The phenotypic legacy of admixture between modern humans and Neandertals. *Science* **351**, 737–741 (2016).
25. F. Arcuri, C. Albore Livadie, G. Di Maio, E. Espósito, G. Napoli, S. Scala, E. Soriano, Influssi balcanici e genesi del Bronzo antico in Italia meridionale: la koinè Cetina e la facies di Palma Campania. *Riv. di Sci. Preist.* **LXVI**, 77–95 (2016).
26. K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
27. M. Dannemann, F. Racimo, Something old, something borrowed: Admixture and adaptation in human evolution. *Curr. Opin. Genet. Dev.* **53**, 1–8 (2018).
28. J. MacArthur, E. Bowler, M. Cerezo, L. Gil, P. Hall, E. Hastings, H. Junkins, A. McMahon, A. Milano, J. Morales, Z. M. Pendlington, D. Welter, T. Burdett, L. Hindorff, P. Flicek, F. Cunningham, H. Parkinson, The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
29. L. C. Jacobs, A. Wollstein, O. Lao, A. Hofman, C. C. Klaver, A. G. Uitterlinden, T. Nijsten, M. Kayser, F. Liu, Comprehensive candidate gene study highlights *UGT1A* and *BNC2* as new genes determining continuous skin color variation in Europeans. *Hum. Genet.* **132**, 147–158 (2013).
30. M. Dannemann, K. Prüfer, J. Kelso, Functional implications of Neandertal introgression in modern humans. *Genome Biol.* **18**, 61 (2017).
31. R. M. Gittelman, J. G. Schraiber, B. Vernot, C. Mikacenic, M. M. Wurfel, J. M. Akey, Archaic hominin admixture facilitated adaptation to out-of-africa environments. *Curr. Biol.* **26**, 3375–3382 (2016).
32. P. Sekula, Y. Li, H. C. Stanescu, M. Wuttke, A. B. Ekici, D. Bockenhauer, G. Walz, S. H. Powis, J. T. Kielstein, P. Brenchley; GCKD Investigators, K.-U. Eckardt, F. Kronenberg, R. Kleta, A. Köttgen, Genetic risk variants for membranous nephropathy: Extension of and association with other chronic kidney disease aetiologies. *Nephrol. Dial. Transplant.* **32**, 325–332 (2017).
33. I. Mathieson, S. Alpaslan-Roodenberg, C. Posth, A. Szécsényi-Nagy, N. Rohland, S. Mallick, I. Olalde, N. Broomandkhoshbacht, F. Candilio, O. Cheronet, D. Fernandes, M. Ferry, B. Gamarra, G. G. Fortes, W. Haak, E. Harney, E. Jones, D. Keating, B. Krause-Kyora, I. Kucukkalipci, M. Michel, A. Mittnik, K. Nägele, M. Novak, J. Oppenheimer, N. Patterson, S. Pfrengle, K. Sirak, K. Stewardson, S. Vai, S. Alexandrov, K. W. Alt, R. Andreescu, D. Antonović, A. Ash, N. Atanassova, K. Bacvarov, M. B. Gusztáv, H. Bocherens, M. Bolus, A. Boroneanţ, Y. Boyadzhiev, A. Budnik, J. Burmaz, S. Chohadzhiev, N. J. Conard, R. Cottiaux, M. Čuka, C. Cupillard, D. G. Drucker, N. Elenski, M. Francken, B. Galabova, G. Ganetsovski, B. Gély, T. Hajdu, V. Handzhyiska, K. Harvati, T. Higham, S. Iliev, I. Janković, I. Karavanić, D. J. Kennett, D. Komšo, A. Kozak, D. Labuda, M. Lari, C. Lazar, M. Leppek, K. Leshtakov, D. L. Vetro, D. Los, I. Lozanov, M. Malina, F. Martini, K. McSweeney, H. Meller, M. Menđušić, P. Mirea, V. Moiseyev, V. Petrova, T. D. Price, A. Simalcsik, L. Sineo, M. Šlaus, V. Slavchev, P. Stanev, A. Starović, T. Szeniczey, S. Talamo, M. Teschler-Nicola, C. Thevenet, I. Valchev, F. Valentin, S. Vasilyev, F. Veljanovska, S. Venelinova, E. Veselovskaya, B. Viola, C. Virag, J. Zaninović, S. Zäuner, P. W. Stockhammer, G. Catalano, R. Krauß, D. Caramelli, G. Zariņa, B. Gaydarska, M. Lillie, A. G. Nikitin, I. Potekhina, A. Papathanasiou, D. Borić, C. Bonsall, J. Krause, R. Pinhasi, D. Reich, The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018).
34. C. Capelli, V. Onofri, F. Brisighelli, I. Boschi, F. Scarnicci, M. Masullo, G. Ferri, S. Tofanelli, A. Tagliabracci, L. Gusmao, A. Amorim, F. Gatto, M. Kirin, D. Merlitti, M. Brion, A. B. Verea, V. Romano, F. Cali, V. Pascali, Moors and Saracens in Europe: Estimating the medieval North African male legacy in southern Europe. *Eur. J. Hum. Genet.* **17**, 848–852 (2009).
35. E. Tamm, J. Di Cristofaro, S. Mazières, E. Pennarun, A. Kushniarevich, A. Raveane, O. Semino, J. Chiaroni, L. Pereira, M. Metspalu, F. Montinaro, Genome-wide analysis of Corsican population reveals a close affinity with Northern and Central Italy. *BioRxiv*, 10.1101/722165.
36. B. Yunusbayev, M. Metspalu, E. Metspalu, A. Valeev, S. Litvinov, R. Valiev, V. Akhmetova, E. Balanovska, O. Balanovsky, S. Turdikulova, D. Dalimova, P. Nymadawa, A. Bahmanimehr, H. Sahakyan, K. Tambets, S. Fedorova, N. Barashkov, I. Khidiyatova, E. Mihailov, R. Khusainova, L. Damba, M. Derenko, B. Malyarchuk, L. Osipova, M. Voevoda, L. Yepiskoposyan, T. Kivisild, E. Khusnutdinova, R. Villems, The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLOS Genet.* **11**, e1005068 (2015).
37. J. Z. Li, D. M. Absher, H. Tang, A. M. Southwick, A. M. Casto, S. Ramachandran, H. M. Cann, G. S. Barsh, M. Feldman, L. L. Cavalli-Sforza, R. M. Myers, Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
38. D. M. Behar, M. Metspalu, Y. Baran, N. M. Kopelman, B. Yunusbayev, A. Gladstein, S. Tzur, H. Sahakyan, A. Bahmanimehr, L. Yepiskoposyan, K. Tambets, E. K. Khusnutdinova, A. Kushniarevich, O. Balanovsky, E. Balanovsky, L. Kovacevic, D. Marjanovic, E. Mihailov, A. Kouvatsi, C. Triantaphyllidis, R. J. King, O. Semino, A. Torroni, M. F. Hammer, E. Metspalu, K. Skorecki, S. Rosset, E. Halperin, R. Villems, N. A. Rosenberg, No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. *Hum. Biol.* **85**, 859–900 (2013).
39. M. Rasmussen, Y. Li, S. Lindgreen, J. S. Pedersen, A. Albrechtsen, I. Moltke, M. Metspalu, E. Metspalu, T. Kivisild, R. Gupta, M. Bertalan, K. Nielsen, M. T. P. Gilbert, Y. Wang, M. Raghavan, P. F. Campos, H. M. Kamp, A. S. Wilson, A. Gledhill, S. Tridico, M. Bunce, E. D. Lorenzen, J. Binladen, X. Guo, J. Zhao, X. Zhang, H. Zhang, Z. Li, M. Chen, L. Orlando, K. Kristiansen, M. Bak, N. Tommerup, C. Bendixen, T. L. Pierre, B. Grønnow, M. Meldgaard, C. Andreasen, S. A. Fedorova, L. P. Osipova, T. F. G. Higham, C. B. Ramsey, T. V. O. Hansen, F. C. Nielsen, M. H. Crawford, S. Brunak, T. Sicheritz-Pontén, R. Villems, R. Nielsen, A. Krogh, J. Wang, E. Willerslev, Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
40. M. Raghavan, P. Skoglund, K. E. Graf, M. Metspalu, A. Albrechtsen, I. Moltke, S. Rasmussen, T. W. Stafford Jr., L. Orlando, E. Metspalu, M. Karmin, K. Tambets, S. Rootsi, R. Mägi, P. F. Campos, E. Balanovska, O. Balanovsky, E. Khusnutdinova, S. Litvinov, L. P. Osipova, S. A. Fedorova, M. I. Voevoda, M. DeGiorgio, T. Sicheritz-Ponten, S. Brunak, S. Demeshchenko, T. Kivisild, R. Villems, R. Nielsen, M. Jakobsson, E. Willerslev, Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
41. D. M. Behar, B. Yunusbayev, M. Metspalu, E. Metspalu, S. Rosset, J. Parik, S. Rootsi, G. Chaubey, I. Kutuev, G. Yudkovsky, E. K. Khusnutdinova, O. Balanovsky, O. Semino, L. Pereira, D. Comas, D. Gurwitz, B. Bonne-Tamir, T. Parfitt, M. F. Hammer, K. Skorecki, R. Villems, The genome-wide structure of the Jewish people. *Nature* **466**, 238–242 (2010).
42. L. Kovacevic, K. Tambets, A.-M. Ilumäe, A. Kushniarevich, B. Yunusbayev, A. Solnik, T. Bego, D. Primorac, V. Skaro, A. Leskovac, Z. Jakovski, K. Drobnic, H.-V. Tolk, S. Kovacevic, P. Rudan, E. Metspalu, D. Marjanovic, Standing at the gateway to Europe - the genetic structure of Western balkan populations based on autosomal and haploid markers. *PLOS ONE* **9**, e105090 (2014).
43. A. Kushniarevich, O. Utevska, M. Chuhryaeva, A. Agdzhoyan, K. Dibirova, I. Uktveryte, M. Möls, L. Mulahasanovic, A. Pshenichnov, S. Frolova, A. Shanko, E. Metspalu, M. Reidla, K. Tambets, E. Tamm, S. Koshel, V. Zaporozhchenko, L. Atramentova, V. Kučinskas, O. Davydenko, O. Goncharova, I. Evseeva, M. Churnosov, E. Pocheshchova,

B. Yunusbayev, E. Khusnutdinova, D. Marjanović, P. Rudan, S. Rootsi, N. Yankovsky, P. Endicott, A. Kassian, A. Dybo; Genographic Consortium, C. Tyler-Smith, E. Balanovska, M. Metspalu, T. Kivisild, R. Villems, O. Balanovsky, Genetic heritage of the Balto-Slavic speaking populations: A synthesis of autosomal, mitochondrial and Y-chromosomal data. *PLOS ONE* **10**, e0135820 (2015).

44. 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
45. M. Metspalu, I. G. Romero, B. Yunusbayev, G. Chaubey, C. B. Mallick, G. Hudjashov, M. Nelis, R. Mägi, E. Metspalu, M. Remm, R. Pitchappan, L. Singh, K. Thangaraj, R. Villems, T. Kivisild, Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *Am. J. Hum. Genet.* **89**, 731–744 (2011).
46. S. Chaubey, C. Benson, H. Khan, M. Monaghan, A. M. Shah, I. Judson, O. Wendler, Six-year survival of a patient with pulmonary artery angiosarcoma. *Asian Cardiovasc. Thorac. Ann.* **20**, 728–730 (2012).
47. L. Pagani, S. Schiffels, D. Gurdasani, P. Danecek, A. Scally, Y. Chen, Y. Xue, M. Haber, R. Ekong, T. Oljira, E. Mekonnen, D. Luiselli, N. Bradman, E. Bekele, P. Zalloua, R. Durbin, T. Kivisild, C. Tyler-Smith, Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. *Am. J. Hum. Genet.* **96**, 986–991 (2015).
48. S. Parolo, A. Lisa, D. Gentilini, A. M. Di Blasio, S. Barlera, E. B. Nicolis, G. B. Boncoraglio, E. A. Parati, S. Bione, Characterization of the biological processes shaping the genetic structure of the Italian population. *BMC Genet.* **16**, 132 (2015).
49. U. Hodoğlugil, R. W. Mahley, Turkish population structure and genetic ancestry reveal relatedness among Eurasian populations. *Ann. Hum. Genet.* **76**, 128–141 (2012).
50. M. Haber, D. Gauguier, S. Youhanna, N. Patterson, P. Moorjani, L. R. Botigué, D. E. Platt, E. Matisoo-Smith, D. F. Soria-Hernanz, R. S. Wells, J. Bertranpetit, C. Tyler-Smith, D. Comas, P. A. Zalloua, Genome-wide diversity in the levant reveals recent structuring by culture. *PLOS Genet.* **9**, e1003316 (2013).
51. M. Haber, M. Mezzavilla, A. Bergström, J. Prado-Martinez, P. Hallast, R. Saif-Ali, M. Al-Habori, G. Dedoussis, E. Zeggini, J. Blue-Smith, R. S. Wells, Y. Xue, P. A. Zalloua, C. Tyler-Smith, Chad genetic diversity reveals an African history marked by multiple holocene Eurasian migrations. *Am. J. Hum. Genet.* **99**, 1316–1324 (2016).
52. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
53. O. Delaneau, J.-F. Zagury, J. Marchini, Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
54. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLOS Genet.* **2**, e190 (2006).
55. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
56. N. M. Kopelman, J. Mayzel, M. Jakobsson, N. A. Rosenberg, I. Mayrose, Clumpak: A program for identifying clustering modes and packaging population structure inferences across *K. Mol. Ecol. Resour.* **15**, 1179–1191 (2015).
57. J. K. Pickrell, N. Patterson, P.-R. Loh, M. Lipson, B. Berger, M. Stoneking, B. Pakendorf, D. Reich, Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2632–2637 (2014).
58. F. Montinaro, G. B. J. Busby, M. Gonzalez-Santos, O. Oosthuitzen, E. Oosthuitzen, P. Anagnostou, G. Destro-Bisol, V. L. Pascali, C. Capelli, Complex ancient genetic structure and cultural transitions in Southern African populations. *Genetics* **205**, 303–316 (2017).
59. Z. Hofmanová, S. Kreutzer, G. Hellenthal, C. Sell, Y. Diekmann, D. Díez-del-Molino, L. van Dorp, S. López, A. Kousathanas, V. Link, K. Kirsanow, L. M. Cassidy, R. Martiniano, M. Strobel, A. Scheu, K. Kotsakis, P. Halstead, S. Triantaphyllou, N. Kyparissi-Apostolika, D. Urem-Kotsou, C. Ziota, F. Adaktylou, S. Gopalan, D. M. Bobo, L. Winkelbach, J. Blöcher, M. Unterländer, C. Leuenberger, Ç. Çilingiroğlu, B. Horejs, F. Gerritsen, S. J. Shennan, D. G. Bradley, M. Currat, K. R. Veeramah, D. Wegmann, M. G. Thomas, C. Papageorgopoulou, J. Burger, Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6886–6891 (2016).
60. F. Broushaki, M. G. Thomas, V. Link, S. López, L. van Dorp, K. Kirsanow, Z. Hofmanová, Y. Diekmann, L. M. Cassidy, D. Díez-del-Molino, A. Kousathanas, C. Sell, H. K. Robson, R. Martiniano, J. Blöcher, A. Scheu, S. Kreutzer, R. Bollongino, D. Bobo, H. Davudi, O. Munoz, M. Currat, K. Abdi, F. Biglari, O. E. Craig, D. G. Bradley, S. Shennan, K. Veeramah, M. Mashkour, D. Wegmann, G. Hellenthal, J. Burger, Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 499–503 (2016).
61. I. Mathieson, I. Lazaridis, N. Rohland, S. Mallick, N. Patterson, S. A. Roodenberg, E. Harney, K. Stewardson, D. Fernandes, M. Novak, K. Sirak, C. Gamba, E. R. Jones, B. Llamas, S. Dryomov, J. Pickrell, J. L. Arsuaga, J. M. de Castro, E. Carbonell, F. Gerritsen, A. Khokhlov, P. Kuznetsov, M. Lozano, H. Meller, O. Mochalov, V. Moiseyev, M. A. Guerra, J. Roodenberg, J. M. Vergès, J. Krause, A. Cooper, K. W. Alt, D. Brown, D. Anthony, C. Lalueza-Fox, W. Haak, R. Pinhasi, D. Reich, Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
62. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
63. M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, M. G. McDermott, C. D. Monteiro, G. W. Gundersen, A. Ma'ayan, Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
64. M. Kanai, M. Akiyama, A. Takahashi, N. Matoba, Y. Momozawa, M. Ikeda, N. Iwata, S. Ikegawa, M. Hirata, K. Matsuda, M. Kubo, Y. Okada, Y. Kamatani, Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).

Acknowledgments: We would like to thank St Hugh's College and the Department of Zoology for facilitating the visits of A.R. and S.A. to the University of Oxford, and the PhD programs of the University of Pavia and the University of Turin for supporting these visits; the High Performance Computing Facility of Oxford University and CINECA for computational resources, programming assistance, and advice given during this project; the SU.VI.MAX for the access to the F_{ST} estimates of their unpublished work; T. Capra for sharing the list of Neanderthal introgressed regions in humans; R. Daniels and M. Gonzales Santos for computing advice during the early stages of this project; L. Saag for the discussions for aDNA suggestions; L. Alessandri and M. Capardoni for comments on the archeological context of the Bronze Age in Italy and surrounding regions; S. Guarrera for technical support; P. Mitchell for the useful comments and suggestions; and the National Alpini Association (Associazione Nazionale Alpini) for help in collecting Italian DNA samples at the 86th national assembly in Piacenza in 2013, in particular B. Plucani, G. Basile, C. Ferrari, and the municipality of Piacenza. Last, we would like to acknowledge all the people who donated their DNA and made this work possible. **Funding:** The Leverhulme Trust (F.M. and C.C.); the Italian Ministry of Education, University and Research (MIUR): "Progetti Futuro in Ricerca 2012" (RBFR126B8I) (A.O. and A.A); the "Dipartimenti di Eccellenza (2018–2022)" [Department of Medical Sciences of Turin (G.M.); Department of Biology and Biotechnology of Pavia (A.A., A.O., O.S., and A.T.)]; the Fondazione Cariplo (project 2018-2045; A.T., A.A., and A.O.); the Italian Institute for genomic Medicine (IIGM) and Compagnia di San Paolo Torino, Italy (G.M.); the European Community, Sixth Framework Program (PROCARDIS: LSHM-CT-2007-037273) (S.B.); the Italian Ministry of Health (Besta CEDIR project: RC 2007/LR6, RC 2008/LR6; RC 2009/LR8; RC 2010/LR8; GR-2011-02347041) (G.B.B.); "Progetti di Ricerca finanziati dall'Università degli Studi di Torino (ex 60%) (2015)" (C.D.G. and G.M.); and ANR-14-CE10-0001 and Région Pays de la Loire (J.G.). **Author contributions:** A.O., A.A., G.M., F.M., and C.C. conceived the idea for the study. A.R., S.A., F.M., and C.C. performed, devised, or supervised the analyses. A.A., A.O., F.B., and V.L.P. provided reagents for the genotyping of novel samples. All the authors contributed to this study, providing data, computational facilities, or other resources. A.R., S.A., F.M., and C.C. wrote the manuscript with inputs from co-authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Data genotyped as part of this project and presented here for the first time (129 Italian samples and 6 samples from Albania, genotyped on the Infinium Omni2.5-8 Illumina beadchip) can be downloaded at the webpage https://capelligroup.wordpress.com/data/. The sampling procedure and project aims related to these samples were reviewed and approved by the Ethics Committee for Clinical Experimentation of the University of Pavia, board minutes of 11 April 2013 and of 5 October 2010, respectively. Data for 49 samples from Italy and Europe were generated at the Institute of Genomics in Tartu and will be released in the accession link provided in *(35)*. Requests for accessing previously published data should be directed to the corresponding authors of the publications where they were originally presented. F_{ST} estimates among clusters in France can be requested from C.D. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 12 December 2018
Resubmitted 10 April 2019
Accepted 6 August 2019
Published 4 September 2019
10.1126/sciadv.aaw3492

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

Science Advances

Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe

A. Raveane, S. Aneli, F. Montinaro, G. Athanasiadis, S. Barlera, G. Birolo, G. Boncoraglio, A. M. Di Blasio, C. Di Gaetano, L. Pagani, S. Parolo, P. Paschou, A. Piazza, G. Stamatoyannopoulos, A. Angius, N. Brucato, F. Cucca, G. Hellenthal, A. Mulas, M. Peyret-Guzzon, M. Zoledziewska, A. Baali, C. Bycroft, M. Cherkaoui, J. Chiaroni, J. Di Cristofaro, C. Dina, J. M. Dugoujon, P. Galan, J. Giemza, T. Kivisild, S. Mazieres, M. Melhaoui, M. Metspalu, S. Myers, L. Pereira, F. X. Ricaut, F. Brisighelli, I. Cardinali, V. Grugni, H. Lancioni, V. L. Pascali, A. Torroni, O. Semino, G. Matullo, A. Achilli, A. Olivieri and C. Capelli

Sci Adv **5** (9), eaaw3492.
DOI: 10.1126/sciadv.aaw3492

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science Advances* is a registered trademark of AAAS.