

# The Genetic Origin of Daunians and the Pan-Mediterranean Southern Italian Iron Age Context

Serena Aneli \*,<sup>†,1</sup> Tina Saupe ,<sup>†,2</sup> Francesco Montinaro,<sup>2,3</sup> Anu Solnik,<sup>4</sup> Ludovica Molinaro,<sup>2</sup> Cinzia Scaggion,<sup>5</sup> Nicola Carrara,<sup>6</sup> Alessandro Raveane ,<sup>7</sup> Toomas Kivilild,<sup>2,8</sup> Mait Metspalu,<sup>2</sup> Christiana L. Scheib ,<sup>‡,2,9</sup> and Luca Pagani<sup>‡,1,2</sup>

<sup>1</sup>Department of Biology, University of Padua, Padova, Italy

<sup>2</sup>Estonian Biocentre, Institute of Genomics, University of Tartu, Tartu, Estonia

<sup>3</sup>Department of Biology-Genetics, University of Bari, Bari, Italy

<sup>4</sup>Core Facility, Institute of Genomics, University of Tartu, Tartu, Estonia

<sup>5</sup>Department of Geosciences, University of Padua, Padova, Italy

<sup>6</sup>Anthropology Museum, University of Padova, Padova, Italy

<sup>7</sup>Laboratory of Hematology-Oncology, European Institute of Oncology IRCCS, Milan, Italy

<sup>8</sup>Department of Human Genetics, KU Leuven, Leuven, Belgium

<sup>9</sup>St John's College, Cambridge, United Kingdom

\*Corresponding author: E-mail: serena.aneli@gmail.com.

†These authors contributed equally to this work.

‡Senior authors.

Associate editor: Maria C. Ávila-Arcos

## Abstract

The geographical location and shape of Apulia, a narrow land stretching out in the sea at the South of Italy, made this region a Mediterranean crossroads connecting Western Europe and the Balkans. Such movements culminated at the beginning of the Iron Age with the Iapygian civilization which consisted of three cultures: Peucetians, Messapians, and Daunians. Among them, the Daunians left a peculiar cultural heritage, with one-of-a-kind stelae and pottery, but, despite the extensive archaeological literature, their origin has been lost to time. In order to shed light on this and to provide a genetic picture of Iron Age Southern Italy, we collected and sequenced human remains from three archaeological sites geographically located in Northern Apulia (the area historically inhabited by Daunians) and radiocarbon dated between 1157 and 275 calBCE. We find that Iron Age Apulian samples are still distant from the genetic variability of modern-day Apulians, they show a degree of genetic heterogeneity comparable with the cosmopolitan Republican and Imperial Roman civilization, even though a few kilometers and centuries separate them, and they are well inserted into the Iron Age Pan-Mediterranean genetic landscape. Our study provides for the first time a window on the genetic make-up of pre-Roman Apulia, whose increasing connectivity within the Mediterranean landscape, would have contributed to laying the foundation for modern genetic variability. In this light, the genetic profile of Daunians may be compatible with an at least partial autochthonous origin, with plausible contributions from the Balkan peninsula.

Key words: ancient DNA, Dauni, Iron Age, Southern Italy.

## Introduction

The Mediterranean's Iron Age populations, between 1100 and 600 BCE, lived in a time of previously unprecedented connectivity (Hodos 2020). Although the technological advances in seafaring had allowed great opportunities for long-distance mobility just a few millennia earlier, it was during the Iron Age that the cosmopolitan role of Mare Nostrum arose, by promoting the spreading of cultures, goods, languages, technical advances as well as heterogeneous ancestral genetic components coming from far and wide (Abulafia 2011; Hodos 2020).

Notable examples of such distant connections are the Greek and Phoenician settlements across the Central and Western Mediterranean shores beginning from the 9th and 8th centuries (Fernandes et al. 2020; Marcus et al. 2020).

The Iron Age is conventionally marked as beginning 950 BCE in the Italian peninsula and its Tyrrhenian islands (Nijboer 2006; Hodos 2020; Aneli et al. 2021). These areas too joined the cosmopolitan wave sweeping across Southern Europe, by hosting numerous trading posts along its shores. A patchwork of communities appeared in this period within the Italian

© The Author(s) 2022. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

borders, each characterized by unique and well-defined cultures and identities, which were later encapsulated and blurred by the Roman colonization. The shift from Republican to Imperial Rome, with the consequent inclusion of non-Roman civilizations from and beyond the Italian peninsula, has already been shown to be connected with a major eastward genetic shift in Central Italian samples (Antonio et al. 2019). Genetic material from Imperial Romans from Central Italy is indeed the first to colocalize with contemporary Italians in a Principal Component space, pointing to this time period (after 27 BCE) as a crucial one in shaping contemporary Italian genetic makeup. Whether such a shift can also be observed in the rest of the Italian peninsula and how the Republic to Imperial transition may have impacted the genetic landscape of local, pre-Roman populations remains an open question.

Despite the numerous written records and archaeological findings, questions about Iron Age populations, their origins, and mutual relationships remain unanswered. Among the many groups occupying Italy in the Iron Age, the Daunians, a lapygian population from northern Apulia, were first mentioned in the 7–6th century BCE (Lombardo 2014; Norman 2016). Similarly to their neighboring populations, Peucetians and Messapians (living in central and southern Apulia, respectively), the name of the Daunians comes from ancient Greek documents and, given the absence of written Daunian records, the scant information we have on their social, political, and religious life are wholly reliant upon the material record, such as their one-of-a-kind stelae (Norman 2016). For instance, we know that they were mainly farmers, animal breeders, horsemen, and maritime traders with an established trade network extending across the sea with Illyrian tribes (Tagliente 1986; Kirigin et al. 2010; Norman 2016). A fascinating aspect of this population, as opposed to their neighbors in Apulia, was their tenacious resistance to external influences. For instance, they did not acquire either social or cultural Hellenic elements and no Greek alphabet inscriptions have been found in their settlements. Indeed, they retained a strong cultural identity and political autonomy until the Roman arrival in the late 4th–early 3rd century BCE (Norman 2016).

Despite the extensive archaeological literature, their origin has been lost to time and, as early as in the Hellenistic period, various legends already existed connecting them to either Illyria (an ancient region broadly identifiable with the Balkan peninsula), Arkadia (present-day Peloponnesus), or Crete (Lombardo 2014; Norman 2016).

To shed light on this population and provide a glance of the Iron Age genetic landscape of the Southern Italian peninsula, we collected human remains from three Iron Age necropoleis geographically located within the area historically inhabited by Daunians: *Ordona* (ancient Herdonia), one of the largest settlements estimated at 600 hectares (Norman 2016), *Salapia*, and *San Giovanni Rotondo* (fig. 1A). Here, we provide for the first time a genetic investigation of Iron Age Apulian individuals associated with the Daunian culture, offering insight into the genetic landscape of pre-Roman Southern Italy.

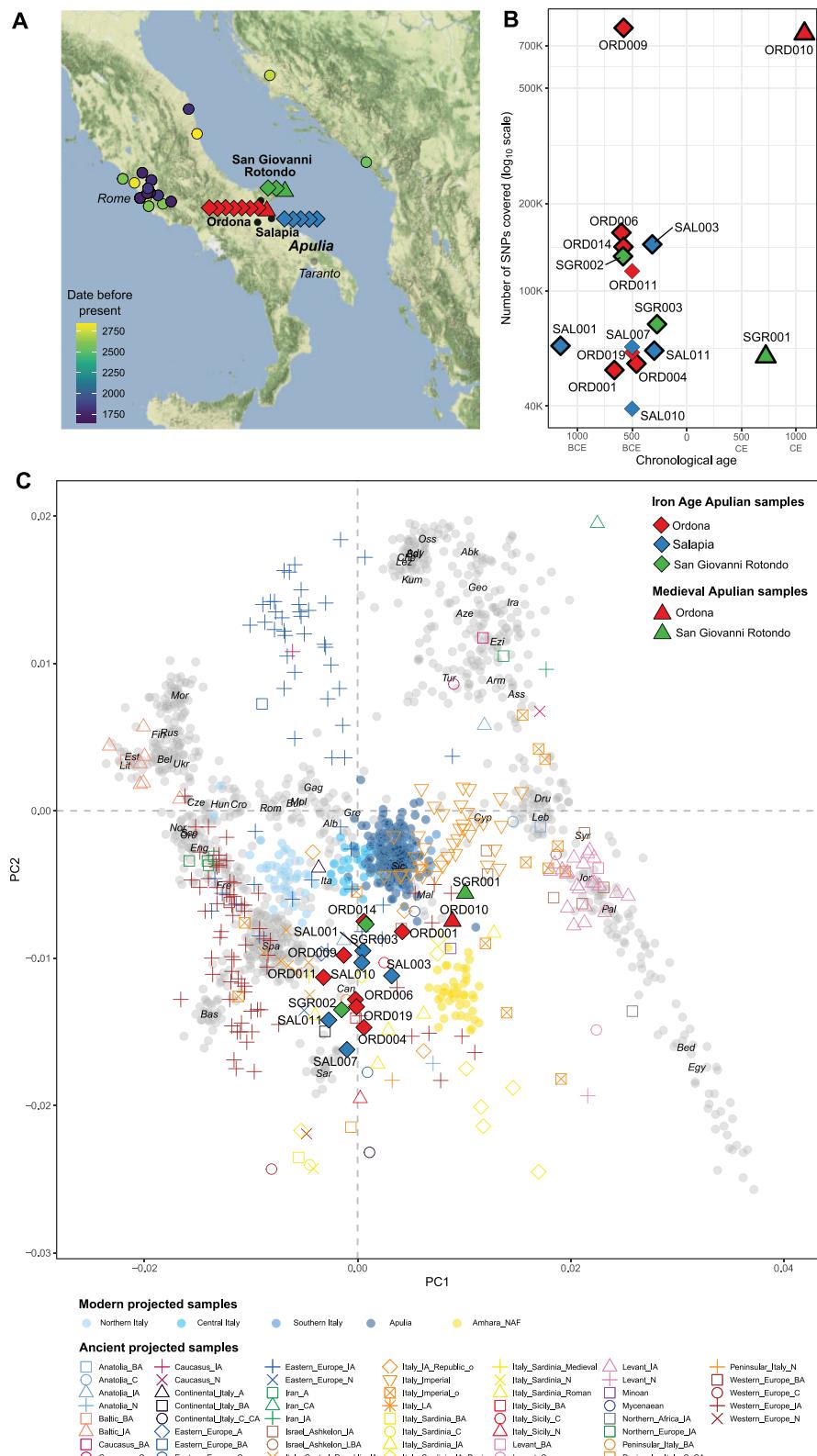
## Results

We extracted DNA from 34 human skeletal remains (petrous bone = 23 and teeth = 11) from three necropoleis (Ordona = 19; Salapia = 12; San Giovanni Rotondo = 3; hereafter ORD, SAL, and SGR, respectively) at the Ancient DNA Laboratory of the Institute of Genomics, University of Tartu in Estonia ([supplementary data 1A, Supplementary Material online](#)), processing between 250 and 1,200 mg of bone material for each sample. All three necropoleis are geographically located less than 50 km from each other in modern Apulia, Southeastern Italy. Ordona and Salapia are archaeologically dated to the Daunian period (6th–3rd century BCE) except one individual (ORD010) archaeologically dated to the Mediaeval time period (which in Europe dates approximately from the 5th to the late 15th century CE) (fig. 1B). Based on the museum record, the samples from the San Giovanni Rotondo necropole have been archaeologically inferred to be from the Iron Age period. After excluding 13 samples due to low DNA concentration (<0.5 ng/μl) and screening 21 libraries at a low depth ( $\pm 20$ M reads per library), we additionally sequenced to higher depth 16 libraries with endogenous DNA between 1.81% and 38.82% and mitochondrial DNA-based (mtDNA) contamination estimated at less than 2.89% ([supplementary data 1A, Supplementary Material online](#)).

The sequencing runs were merged resulting in 16 individuals for genome-wide analysis: eight from Ordona, five from Salapia, and three from San Giovanni Rotondo. The final data set includes individuals with an average genomic coverage between 0.031 and 0.995× and a number of single-nucleotide polymorphisms (SNPs) overlapping with Human Origins 1240k between  $\sim$ 40,000 and  $\sim$ 810,000 ([supplementary data 1A, Supplementary Material online](#)).

Out of those 16 individuals, we selected ten individuals based on their proximity within the principal component analysis (PCA) space (fig. 1C) for radiocarbon dating and estimated their age between 1157 and 275 calBCE with a median date of 521 calBCE ([supplementary data 1B, Supplementary Material online](#)). The radiocarbon dates confirm the archaeological dates of individuals from Salapia and Ordona as well as the Iron Age affiliation of the two San Giovanni Rotondo samples (SGR002 and SGR003). Two additional samples (ORD010 and SGR001) with a shift toward the Near East in the PCA (fig. 1C) were radiocarbon dated to 995–1156 calCE (95.4%) and 670–774 calCE (95.4%), respectively ([supplementary data 1B, Supplementary Material online](#)). Based on their dates, the samples were used as an external control for further analyses focusing on Iron Age Apulia (IAA).

We determined the mitochondrial DNA (mtDNA) haplotype of each individual and the Y chromosome (Ychr) haplogroup for nine male individuals (see Materials and Methods and [supplementary data 1A, C, and D, Supplementary Material online](#)). We found that the mtDNA haplotypes mostly belong to the mtDNA lineages H1, H5, K1, and U5, haplotypes found in previous studies of individuals from this time period in the Italian Peninsula (Antonio et al. 2019). Besides the Ychr haplogroup R1b, which is the most frequent



Downloaded from https://academic.oup.com/mbe/article/39/2/msa014/6509524 by University of Arizona user on 16 August 2022

**FIG. 1.** Geographical location, dating, and genetic structure. (A) Geographical location of the newly generated Apulian ancient individuals (green, red, and blue diamonds and triangles) and other published Iron Age samples from Italy and surrounding areas (color scale mirroring their “years before present” dates). (See also [table 1](#) and [supplementary data 1 and 2, Supplementary Material online](#).) (B) Dating and number of SNPs from 1240K set covered. Samples with thick edges have been radiocarbon dated (mean presented), whereas for the others, we represented their archaeological dating ([table 1](#) and [supplementary data 1B, Supplementary Material online](#)). (C) PCA of newly generated individuals with reference ancient individuals projected onto the genetic variation inferred from modern-day Western Eurasian populations ([supplementary data 2 and 3, Supplementary Material online](#)). For the ancient reference populations which have not been attributed to the Iron Age, we plotted their centroids.

haplogroup during the Bronze Age in the Italian Peninsula and on the islands Sardinia and Sicily (Allentoft et al. 2015; Haak et al. 2015; Antonio et al. 2019; Fernandes et al. 2020; Saupe et al. 2021), we found the Ychr lineages I1-M253, I2d-M223, and J2b-M241. The haplogroup I2d-M223 was one of the main Y chromosome lineages in Western Europe until the Late Neolithic whereas J2b-M241 first appears in the Bronze Age (Allentoft et al. 2015; Mathieson et al. 2015; Schuenemann et al. 2017; Fernandes et al. 2020; Marcus et al. 2020). We found one Early Mediaeval individual (SGR001) belonging to haplogroup I1-M253, which is common in Northern Europe and previously also detected in a 6th Century Langobard burial from North Italy (Amorim et al. 2018).

We used READ (Monroy Kuhn et al. 2018) to identify pairs with first or second degree of genetic relatedness from autosomal data (see Materials and Methods and [supplementary data 1E–H, Supplementary Material online](#), for details, [supplementary fig. 11, Supplementary Material online](#)). We found one first-degree relationship between the two female individuals ORD001 (age: 9–11 years, 779–544 calBCE; 95.4%) and ORD009 (age: 40–45 years, 749–406 calBCE; 95.4%) from Ordona sharing the identical mtDNA haplogroup H5c and indicating a mother–daughter or full sibling relationship ([supplementary data 1 and fig. 1A and B, Supplementary Material online](#)).

### The Making of Modern Italians

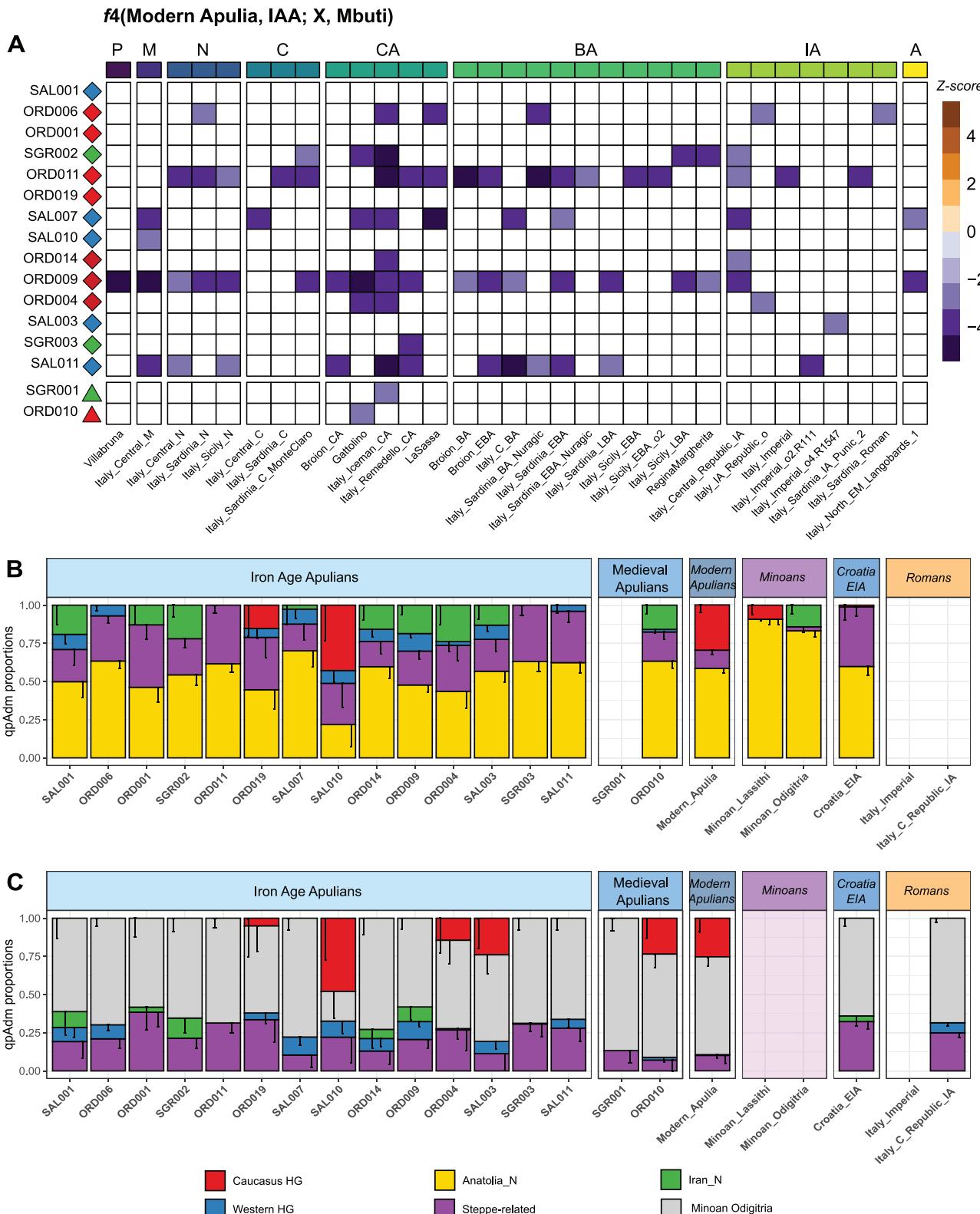
To explore the genetic make-up of the IAA population, we performed a PCA projecting the ancient individuals onto the genetic variation of modern Eurasian samples ([fig. 1C](#) and [supplementary data 2 and 3, Supplementary Material online](#)). Our samples are largely scattered between modern peninsular Italians and Sardinians, and, in contrast to what was generally described (Schiffels et al. 2016; Saag et al. 2019) for other European Iron Age populations (e.g., Northern\_Europe\_IA, Western\_Europe\_IA, and Levant\_IA in [fig. 1C](#)), they are still clearly distant from the genetic variability of modern-day inhabitants of Apulia. The downward shift of Iron Age Apulians from the present-day ones is further confirmed by the significantly negative  $f_4(\text{Modern Apulians}, \text{IAA}; X, \text{Mbuti})$ , where  $X$  is a Neolithic/Chalcolithic/Copper Age population ([fig. 2A](#) and [supplementary data 4, Supplementary Material online](#)). Within the heterogeneity reported by the PCA, which does not mirror the archaeological sites, the two Mediaeval individuals are shifted toward modern Middle Eastern and Caucasus populations (ORD010 and SGR001), whereas the others are stretched along the PC2. This pattern partially mirrors the chronological date with the most recent being more similar to present-day Southern Europeans and is further strengthened when considering the PC3 distribution ([supplementary fig. 2, Supplementary Material online](#)). Three samples located at the bottom of the PCA (ORD004, ORD019, SAL007) and one (SAL010) falling in the middle did not include modern Apulians among the top 25 results of an  $f_3$  outgroup analysis ([supplementary fig. 3, Supplementary Material online](#)). All of them showed an affinity to Copper and Bronze Age Italians (Saupe et al. 2021) as well as the

Aegean and the Mediterranean worlds (including Minoans, Greece, Croatians, and Gibraltar). A similar distribution is mirrored in the multidimensional scaling (MDS) built from the  $f_3$  outgroup measures, where the oldest IAA individual (SAL001: 1260–1048 calBCE; 95.4%) lies farthest from the modern samples, whereas the Mediaeval ones (ORD010 and SGR001) are the closest ([supplementary fig. 4, Supplementary Material online](#)). An additional MDS plot is reported in [supplementary figure 5, Supplementary Material online](#): There, we computed the median distances from the centroid of the IAA samples (excluding the two Mediaeval samples), the Roman Imperials and the Republicans, obtaining 0.0093, 0.0140, and 0.0101, respectively. Notably, the distance distributions were not statistically different, thus suggesting that the extent of genetic heterogeneity of IAA individuals is comparable to that of the cosmopolitan Republican and Imperial Roman samples, the latter having experienced a remarkable degree of admixture with foreign Mediterranean groups, pushed by strong geopolitical contingencies (Antonio et al. 2019). We also investigated whether the coverage was somehow responsible for the sample scattering by computing the Pearson correlation between the coverage itself and the distance from the IAA centroid. We found out that the two measures are not correlated ( $R = -0.17, P = 0.57$ ), thus showing that most of the sample variation was not due to technical reasons ([supplementary fig. 6, Supplementary Material online](#)).

The peculiar positioning of the IAA individuals casts doubt on when the major population shift resulting in modern Italian genetic composition took place. The shift toward the modern Italian genetic variability can be seen with the Republican-Imperial Roman samples (Antonio et al. 2019), the latter being more “similar” to modern Italians ([fig. 1C](#)). Whether Apulian individuals dating back to the Republican or Imperial phase would also show a repositioning toward modern genetic variability remains an open question, although the later Mediaeval samples of this study point to population changes having taken place after the Iron Age in the area.

### The Pan-Mediterranean Genetic Landscape of IAA

The geographic location of Apulia, a narrow peninsula stretching out in the sea at the South of Italy, has made this region an important Mediterranean crossroads connecting Western Europe, the Balkans, the Aegean, and Levant worlds. This is reflected in the PCA where IAA individuals are closely related to other Iron Age populations from the Mediterranean and surrounding areas (e.g., Montenegro, Bulgaria, and Sardinia) ([fig. 1C](#) and [supplementary fig. 4, Supplementary Material online](#)). Nomadic or cosmopolitan groups scatter like IAA: three Punic individuals from Sardinia (Italy\_Sardinia\_IA\_Punic; Marcus et al. 2020), three Moldova Scythians already reported to be genetically similar to Southern Europeans (Krzewińska et al. 2018), Spanish individuals from the Hellenistic and the Romans periods (Olalde et al. 2019), and an individual from the 12th century Iron Age Ashkelon (Feldman, Master, et al. 2019) which clusters with ORD001.



**Fig. 2.** Genetic relationship of Iron Age and Middle Age Apulians and their ancestral composition. (A) Heatmap representing the Z score values of  $f_4$ (Modern Apulia, IAA; X, Mbuti), where X is an ancient Italian population. We also added the two Middle Age samples for comparison. Tests with Z scores between  $-3$  and  $3$  or with less than  $5,000$  SNPs were not included (P, Palaeolithic; M, Mesolithic; N, Neolithic; C, Chalcolithic; CA, Copper Age; BA, Bronze Age; IA, Iron Age; A, Antiquity). (B) qpAdm proportions of ancient Apulian samples and other reference populations (Modern Apulians, Minoans, Croatia\_EIA, and the Roman individuals from the Republican and the Imperial period) using “base” sources: Caucasus and Western hunter-gatherer component (HG), Anatolia Neolithic (Anatolia\_N), Iranian Neolithic (Iran\_N), Steppe-related ancestry. (C) qpAdm proportions using also the Minoans as source (Materials and Methods). In (B) and (C), we plotted the model with the highest number of sources and P value (complete results may be seen in supplementary fig. 7, Supplementary Material online).

In order to shed light onto the genetic composition of the IAA individuals, we modeled them as a combination of the main ancestries documented across Western Europe at that time: Western Hunter-Gatherers (WHG), Anatolian Neolithic (AN), Steppe-related and, interchangeably, Caucasus Hunter-Gatherers (CHG), or Iranian Neolithic (IN) using the qpWave/qpAdm framework (fig. 2B and supplementary fig. 7A, Supplementary Material online, and Materials and Methods). Broadly, the contributions of such ancestries to the genetic variability of ancient European populations vary according to their geographical positions: in particular, northernmost locations received higher proportions of WHG, Steppe-related ancestry and, consequently, CHG ancestries, whereas Southern European groups carried variable Iranian Neolithic or CHG traces (Lazaridis et al. 2014). In view of this, we observed that although the IAA individuals could generally be modeled as a two-way admixture between AN and Steppe ( $0.63 \pm 0.08$  and  $0.37 \pm 0.08$ , respectively), the alternative model AN + CHG/IN could also fit for a subset of them, particularly in case of the samples ORD004, ORD010, and SAL010 with higher or comparable  $P$  values (supplementary fig. 7A, Supplementary Material online, first row with two sources). When three or four sources were tested, the presence of WHG ancestry in the majority of our individuals emerges and together with AN, Steppe, and CHG/IN, forms a supported model for IAA samples (fig. 2B and supplementary fig. 7A and data 5, Supplementary Material online). Notably, for the individuals stretching downward in the PCA (ORD004, ORD019, SGR002, and the Mediaeval ORD010) a three-way admixture involving AN, Steppe, and CHG/IN is generally preferable. To better understand the putative contribution of more recent populations, we modeled our samples with base sources (WHG, AN, Steppe-related, and CHG/IN) and, alternatively, Minoans (Odigtria and Lasithi), Amhara\_NAF, and Roman Republicans (fig. 2C and supplementary fig. 7B–D, Supplementary Material online, and Materials and Methods). Amhara\_NAF can be used as a proxy for the non-African component in modern Ethiopian individuals that was tentatively linked to the Sea People, a Bronze Age nomadic seafaring population (Feldman, Master, et al. 2019; Molinaro et al. 2019). Together with Minoans and Roman Republicans, this component can be broadly modeled as a Pan-Mediterranean population (constituted by AN and IN/CHG components) with the addition of WHG and Steppe-related ancestry in Roman Republicans. When modeled also with Minoans and Amhara\_NAF, which roughly proxies the same ancestral signature, the majority of the samples required an additional CHG/IN contribution (two-way admixtures in supplementary fig. 7B and C, Supplementary Material online) as well as Steppe-related and WHG. We further observed that, as previously seen, the WHG contribution is less clear in those samples stretching downward in the PCA. Although the CHG/IN additional contribution may simply proxy the presence of Steppe-related ancestry in IAA, the absence of which in Minoans has already been reported (Lazaridis et al. 2017), the same cannot be said about Roman Republicans (two-way admixtures in supplementary fig. 7D, Supplementary Material online), which

harbored a considerable amount of Steppe component (Antonio et al. 2019). However, this signature is not confirmed with  $f_4$  analyses (supplementary fig. 8B and C, Supplementary Material online), where just Mycenaean groups report less CHG ancestry than our samples.

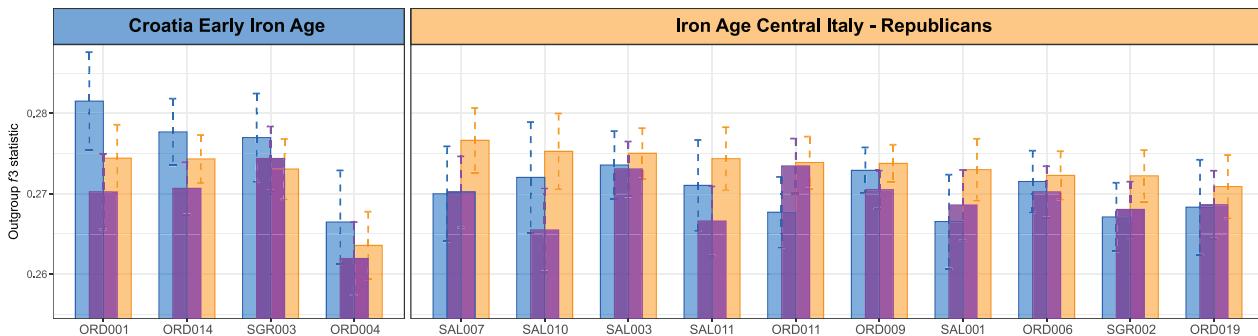
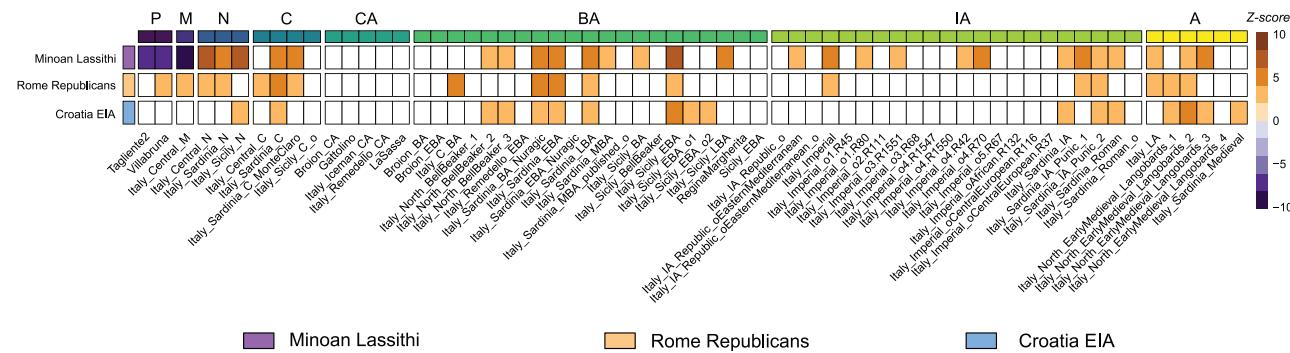
Even taking into consideration some uncertainty, especially for the samples with lower coverage, the broader picture emerging from qpAdm analyses is the pervasive presence of coeval Italian ancestries (Rome Republicans), as well as previous Bronze Age sources (Minoans and Sea People) spread far and wide across the Mediterranean Sea. In such a melting pot scenario, the genetic heterogeneity of IAA individuals, who lived in close temporal and geographical proximity, stand out (fig. 1C and supplementary figs. 4 and 9, Supplementary Material online). An example of the cosmopolitan nature of Iron Age Mediterranean is the first-degree relationship between ORD009 and ORD001, whose positions in the PCA strikingly differ, with the individual ORD001 being stretched toward Middle Eastern and Caucasus modern populations, potentially as a consequence, in case of a mother–daughter relationship, of the foreign origin of the father (fig. 1C and supplementary fig. 4, Supplementary Material online).  $F_4$  analyses in the form  $f_4(\text{ORD009}, \text{ORD001}; X, \text{Mbuti})$  report a slightly significant ( $Z$  score between  $-2$  and  $-3$ ) excess of Greece\_N, Portugal\_LN\_C, Lebanon\_Roman, and Italy\_Sicily\_EBA in ORD001, which may explain its eastward shift (supplementary fig. 10A, Supplementary Material online). Moreover, ORD001, together with SGR002, but not ORD009 harbored more CHG when compared with Lebanon\_Hellenistic and Lebanon\_IA3 samples, respectively (supplementary fig. 10C, Supplementary Material online).

We also investigated whether the PCA scattering was due to varying African or Levantine contributions with  $f_4(\text{Rome Republican}, \text{IAA}, \text{Levant_N/YRI}, \text{Mbuti})$  and tried the same on Mediaeval ancient Apulians (ORD010 and SGR001). However, none of the tested ancient Apulians shows a significant excess of YRI ancestry when compared with the contemporary Roman Republicans, even though ORD014, SAL007, and SAL011 show negative  $f_4$  values with a  $Z$  score between  $-2$  and  $-3$  (supplementary fig. 10B, Supplementary Material online).

### The Origin of Daunians

The apparent genetic heterogeneity of IAA individuals connected to the Daunian population and the cosmopolitan landscape of Iron Age Mediterranean populations hinder a full reconstruction of the demographic processes leading to the Daunians. Nevertheless, a few milestones can be spotted.

When we performed  $f_3$  analyses to investigate the nearest possible source for each IAA individual using Minoans, Iron Age Croatians, and the local Roman Republicans (fig. 3A), we found that none of the IAA individuals shows higher affinities with Minoans. Three of them, which clustered close to modern Italians in the PCA (ORD001, ORD014, and SGR003, fig. 1C), show higher affinity with the Iron Age Croatian sample (ORD004 followed this pattern too, but with lower  $f_3$  values). However, the remaining majority are closest to the Roman Republicans, which can be interpreted as

**A f3(IAA, putative Daunian sources; Mbuti)****B f4(putative Daunians sources, IAA\_P; X, Mbuti)**

**Fig. 3.** Genetic affinities of Iron Age Apulian samples with the putative populations of origin: Minoans (Minoan\_Lassithi), Illyrians (here proxied by the Croatia\_EIA individual), and the Roman Republicans (here proxying the autochthonous Iron Age Italian ancestry). (A) Outgroup  $f_3$  statistics of IAA samples compared with the putative Daunian sources. Samples have been sorted and grouped according to the source whose  $f_3$  values were higher. (B) Heatmap showing the Z score values of  $f_4$ (putative sources, IAA\_P; X, Mbuti), where X is an ancient Italian population and IAA\_P is the entire set of Iron Age Apulian samples taken together. Tests with Z scores between  $-3$  and  $3$  or with less than 5,000 SNPs were not included (P, Palaeolithic; M, Mesolithic; N, Neolithic; C, Chalcolithic; CA, Copper Age; BA, Bronze Age; IA, Iron Age; A, Antiquity).

representative of local Iron Age peninsular Italy ancestry, as also indicated by our MDS results.

Moreover, the WHG contribution, which was a necessary component to explain IAA and Roman Republicans according to qpAdm output, is absent from Minoans and Iron Age Croatians, thus making it a putative signature of an at least partial autochthonous origin (fig. 2B and C and supplementary fig. 7, Supplementary Material online). These results are confirmed for Minoans, but not for Croatia\_EIA, by the  $f_4$  analyses (fig. 3B and supplementary fig. 8A and data 6, Supplementary Material online). Indeed, the significantly negative  $f_4$ (Minoans, IAA; WHG, Mbuti) (fig. 3B and supplementary fig. 8A, Supplementary Material online), as well as the  $f_4$ (Greece\_Minoan\_Lassithi/Greece\_BA\_Mycenaean/ Greece\_BA\_Mycenaean\_Pylos, IAA; X, Mbuti), reported a significant excess of WHG ancestry (X being Tagliente2, Loschbour, Italy\_Central\_M, Switzerland\_Bichon, and so on) in all IAA individuals, with the exception of ORD004, SAL001, and SGR003 (also the Mediaeval samples ORD010 and SGR001 deviate from this pattern, supplementary data 6, Supplementary Material online). An excess of Bronze Age Steppe-related component, which at that time was already present along the Italian peninsula (Saupe et al. 2021), was clearly present in SGR002, ORD006, and ORD009 by the  $f_4$ (IAA, Greece\_Minoan\_Lassithi; X, Mbuti) (supplementary data 6, Supplementary Material online).

The excess of WHG ancestry, tentatively suggesting a local origin, is somewhat blurred by the genetic similarity of the two most probable sources—Illyrians (Croatia\_EIA) and an autochthonous one (Roman Republicans), which together make part of the same Mediterranean continuum. Indeed, although no significantly negative  $f_4$ (Croatia\_EIA, IAA; X, Mbuti) values have been found (with the exception of X being Italy\_C\_BA, Italy\_Iceman\_CA, Portugal\_MBA, and Anatolia\_MLBA in SAL007, which retains the highest proportion of AN in qpAdm in fig. 2B and supplementary fig. 7, Supplementary Material online), the same pattern is obtained when IAA individuals are compared with Roman Republicans ( $f_4$ (RomanRepublicans, IAA; X, Mbuti), supplementary data 6, Supplementary Material online). Imperial individuals from a few centuries later ( $f_4$ (Rome\_Imperial, IAA; X, Mbuti)) show quite the opposite pattern: no positive results have been produced with the exception of ORD004 (X = Lebanon\_Roman), SAL001 (X = Italy\_Sardinia\_Roman\_o), SAL010 (X = Russia\_LateMaikop), and ORD010. Conversely, the negative  $f_4$  values point toward WHG, Neolithic, and Bronze Age Steppe-related ancestries (supplementary data 6, Supplementary Material online).

Another signal coming from qpAdm analyses is the apparent excess of CHG ancestry in IAA; however, the predominant contribution of CHG to the Steppe-related ancestry that, by the Iron Age, had already spread to the Mediterranean area

makes it hard to properly detect a CHG signature independent from the Steppe wave, possibly brought by pan-Mediterranean influxes. When directly investigated with an f4 framework, IAA shows generally more CHG than Mycenaean, less CHG than contemporary Croatian\_EIA and, in some cases (ORD019, SGR002, and the Mediaeval SGR001 with Z scores higher than 2) more CHG than older Croatian samples (\_N = Neolithic and \_MN = Middle Neolithic) ([supplementary fig. 8, Supplementary Material online](#)).

## Discussion

The new genomic sequences of Daunian samples reveal that Iron Age (pre-Roman) Southern Italy (Apulia) can be placed within a Pan-Mediterranean genetic continuum that stretches from Crete (Minoans; [Lazaridis et al. 2017](#)) and the Levant (Sea People; [Feldman, Master, et al. 2019](#); [Molinaro et al. 2019](#)) to the Republican Rome and the Iberian Peninsula ([Antonio et al. 2019](#)), mainly composed by AN and IN/CHG genetic features with the addition of WHG and Steppe-related influences in Continental Italy. Pre-Roman Italian populations, being part of this broader landscape, are not directly superimposable with contemporary Italians which instead seem to be influenced by the homogenizing effect of Imperial Roman and late Antiquity events.

Within the described Pan-Mediterranean landscape, the IAA/Daunians show a sizeable heterogeneity, comparable to the one of the broader and cosmopolitan Republican and even more so Imperial Roman civilization, and the highest genetic affinity to Republican Romans and Iron Age Croatians, whereas Minoans and other Iron Age Greek samples show absent or reduced WHG contribution when compared with IAA. This makes a Cretan or Arkadian origin less likely, even though some tales have connected them with the Greek hero Diomedes and many ancient historians have claimed such origins for their neighbor Messapians and Peucetians.

The Daunians maintained strong commercial and political relations with the Illyrian people, controlling together the area spanning from the Dalmatia to the Gargano peninsula ([Kirigin et al. 2010](#)) and had many cultural affinities with them ([Nava and Descoedres 1990](#)). The material culture, involving peculiar anthropomorphic statue stelae, has provided some information on Daunian culture and may also help in unraveling their mysterious origin. In particular, the forearm decorations on a female stela have been interpreted as tattoos and, whereas tattooing practices were considered barbarian among the Greeks ([Jones 1987](#)), they were customary in populations from Tracia and Illyria and, more generally, among the women of status from the Balkans ([Norman 2011, 2016](#)).

It is not clear whether these connections indicate a movement of people or a sharing of cultural ideas and a conclusive answer to the origin of the Daunians remains elusive. From a parsimony perspective, the genetic results point to an autochthonous origin (e.g., a genetic continuity of Daunians with the population that inhabited the area prior to the examined historical period), here mainly marked by the presence of WHG signature, although we cannot exclude

additional influences from Croatia (ancient Illyria), as described by available historical sources and by the material remains ([De Juliis 1988](#); [Norman 2016](#)).

## Materials and Methods

### Samples Information and Ethical Statement

The ancient human remains analyzed in this work belong to the collections of the Museum of Anthropology, University of Padua and have been found and cataloged by different archaeological campaigns carried out during the 20th century. Given the samples were kept at the Museum of Anthropology of the University of Padova, we informed the territorially competent Soprintendenza (Soprintendenza Archeologia, Belle Arti e Paesaggio per l'Area metropolitana di Venezia e per le Province di Belluno, Padova e Treviso) about our sampling strategy and scientific research project, which was approved and our communication assigned the following protocol numbers: prot. 27903 (September 10, 2020) and prot. 29131 (November 23, 2020).

### Ordona

The necropolis of Herdonia (today's Ordona, within the Apulian Foggia province in Italy) was studied by different archaeological campaigns interested in the Daunians, Roman as well as the Mediaeval settlements in 1978 and 1981 ([Corrain 1986](#)). The human remains collected on such occasions were later extensively cataloged and studied and new paleopathological evidence was also brought back to light ([Scaggion and Carrara 2016](#)). Inhabited starting from the Neolithic period, Herdonia became an important Daunian center from the 6th century BCE.

For this study, in collaboration with the University of Padova and Museum of Anthropology of Padova, samples of human remains (petrous bone = 19) were taken. The samples ORD001, ORD004, ORD006, ORD009, ORD010, and ORD014 were dated at <sup>14</sup>CHRONO Centre for Climate, the Environment, and Chronology in Belfast, United Kingdom ([supplementary data 1B, Supplementary Material online](#)).

### Salapia

The necropolis of Salapia, an ancient town located 10 km from contemporary Cerignola, within the Apulian Foggia province in Italy. The osteological samples were brought by Prof. Santo Tiné to the Museum of Anthropology of the University of Padova with no further information on the archaeological context, and osteological studies were carried out by Cleto Corrain and colleagues in 1971 ([Corrain et al. 1972](#)). From this site, samples of 12 human remains (petrous bone = 5, teeth = 7) were taken. The samples SAL001, SAL003, and SAL011 were dated at <sup>14</sup>CHRONO Centre for Climate, the Environment, and Chronology in Belfast, United Kingdom ([supplementary data 1B, Supplementary Material online](#)).

### San Giovanni Rotondo

The San Giovanni Rotondo samples come from the osteoarchaeological collection of the Museum of Anthropology of

the University of Padova and are not associated to any further record with the exception of a broad “Iron Age” archaeological label and may be part of the samples brought to the Museum by Prof. Santo Tiné in the 1960s. From this site, samples of human remains (petrous bone = 3) were taken. The samples SGR001, SGR002, and SGR003 were dated at <sup>14</sup>CHRONO Centre for Climate, the Environment, and Chronology in Belfast, United Kingdom ([supplementary data 1B, Supplementary Material online](#)).

### DNA Extraction

In all ancient DNA work, we strive to reduce the impact of destructive sampling and retain as much information for future researchers as possible. We sample individuals with multiple elements present, ideally an antimere to the sample taken (when possible); we take detailed photographs of material prior to sampling; we sample the minimum amount that is necessary and we use protocols that will allow for different types of biomolecular or chemical analysis (either simultaneously or in the future).

All of the laboratory work was performed in dedicated ancient DNA laboratories at the Estonian Biocentre, Institute of Genomics, University of Tartu, Tartu, Estonia. The library quantification and sequencing were performed at the Core Facility of the Institute of Genomics, Tartu, Estonia. The main steps of the laboratory work are detailed below.

In total 34 samples from human remains were extracted for DNA analysis ([supplementary data 1, Supplementary Material online](#)). The first layer of pars petrous was removed with a sterilized drill bite to avoid exogenous contamination. A 10-mm core of the inner ear was sampled from the pars petrous. The drill bits and core drill were sterilized in between samples with 6% (w/v) bleach followed by distilled water and then ethanol rinse. Root portions of teeth were removed with a sterile drill wheel.

The root and the petrous portions were soaked in 6% (w/v) bleach for 5 min. Samples were rinsed three times with 18.2 MΩcm H<sub>2</sub>O and soaked in 70% (v/v) Ethanol for 2 min. The tubes were shaken during the procedure to dislodge particles. The samples were transferred to a clean paper towel on a rack inside a class IIB hood with the UV light on and allowed to dry for 2–3 h.

Afterward, the samples were weighed to calculate the accurate volume of EDTA (20× EDTA [μl] of sample mass [mg]) and Proteinase K (0.5× Proteinase K [μl] of sample mass [mg]). EDTA and Proteinase K were added into PCR-clean 5 ml or 15 ml conical tubes (Eppendorf) along with the samples inside the IIB hood and the tubes were incubated 72 h on a slow shaker at room temperature.

The DNA extracts (of root portions and pars petrous portions) were concentrated to 250 μl using the Vivaspin Turbo 15 (Sartorius) and purified in large volume columns (High Pure Viral Nucleic Acid Large Volume Kit, Roche) using 2.5 ml of PB buffer, 1 ml of PE buffer, and 100 μl of EB buffer (MinElute PCR Purification Kit, QIAGEN). For the elution of the endogenous DNA, the silica columns were transferred to a collection tube to dry and followed in 1.5 ml DNA lo-bind tubes (Eppendorf) to elute. The samples were incubated with

100 μl EB buffer at 37 °C for 10 min and centrifuged at 13,000 rpm for 2 min. After centrifugation, the silica columns were removed and the samples were stored at –20 °C. Only one extraction was performed per extraction for screening and 30 μl used for libraries.

### Library Preparation

Sequencing libraries were built using NEBNext DNA Library Prep Master Mix Set for 454 (E6070, New England Biolabs) and Illumina-specific adaptors ([Meyer and Kircher 2010](#)) following established protocols ([Meyer and Kircher 2010; Orlando et al. 2013; Malaspinas et al. 2014](#)). The end repair module was implemented using 18.75 μl of water, 7.5 μl of buffer, and 3.75 μl of enzyme mix, incubating at 20°C for 30 min. The samples were purified using 500 μl PB and 650 μl of PE buffer and eluted in 30 μl EB buffer (MinElute PCR Purification Kit, QIAGEN). The adaptor ligation module was implemented using 10 μl of buffer, 5 μl of T4 ligase, and 5 μl of adaptor mix ([Meyer and Kircher 2010](#)), incubating at 20°C for 15 min. The samples were purified as in the previous step and eluted in 30 μl of EB buffer (MinElute PCR Purification Kit, QIAGEN). The adaptor fill-in module was implemented using 13 μl of water, 5 μl of buffer, and 2 μl of Bst DNA polymerase, incubating at 37°C for 30 min and at 80°C for 20 min. Libraries were amplified using the following PCR set up: 50 μl DNA library, 1× PCR buffer, 2.5 mM MgCl<sub>2</sub>, 1 mg/ml BSA, 0.2 μM inPE1.0, 0.2 mM dNTP each, 0.1U/μl HGS Taq Diamond, and 0.2 μM indexing primer. Cycling conditions were: 5' at 94 °C, followed by 18 cycles of 30 s each at 94, 60, and 68 °C, with a final extension of 7 min at 72 °C. The samples were purified and eluted in 35 μl of EB buffer (MinElute PCR Purification Kit, QIAGEN). Three verification steps were implemented to make sure library preparation was successful and to measure the concentration of dsDNA/sequencing libraries—fluorometric quantitation (Qubit, Thermo Fisher Scientific), parallel capillary electrophoresis (Fragment Analyser, Agilent Technologies), and qPCR.

### DNA Sequencing

DNA was sequenced using the Illumina NextSeq500/550 High-Output single-end 75 cycle kit. As a norm, 15 samples were sequenced together on one flow cell; additional data was generated for 16 samples to increase coverage ([supplementary data 1, Supplementary Material online](#)).

### Mapping

Before mapping, the sequences of the adapters, indexes, and poly-G tails occurring due to the specifics of the NextSeq 500 technology were cut from the ends of DNA sequences using cutadapt-1.11 ([Martin 2011](#)). Sequences shorter than 30 bp were also removed with the same program to avoid random mapping of sequences from other species. The sequences were aligned to the reference sequence GRCh37 (hs37d5) using Burrows-Wheeler Aligner (BWA 0.7.12) ([Li and Durbin 2009](#)) and the command mem with reseeding disabled.

After alignment, the sequences were converted to BAM format and only sequences that mapped to the human genome were kept with samtools 1.3 ([Li et al. 2009](#)). Afterward,

the data from different flow cell lanes were merged and duplicates were removed using picard 2.12 (<https://github.com/broadinstitute/picard/tree/2.12.1>, last accessed January 24, 2022). Indels were realigned using GATK 3.5 (McKenna et al. 2010) and reads with a mapping quality less than 10 were filtered out using samtools 1.3 (Li et al. 2009).

### aDNA Authentication

As a result of degradation over time, aDNA can be distinguished from modern DNA by certain characteristics: short fragments and a high frequency of C=>T substitutions at the 5' ends of sequences due to cytosine deamination. The program mapDamage2.0 (Jónsson et al. 2013) was used to estimate the frequency of 5' C=>T transitions. Rates of contamination were estimated on mitochondrial DNA by calculating the percentage of nonconsensus bases at haplogroup-defining positions as detailed in (Jones et al. 2017). Each sample was mapped against the RSRS downloaded from phylotree.org and checked against haplogroup-defining sites for the sample-specific haplogroup (supplementary data 1A, Supplementary Material online).

Samtools 1.9 (Li et al. 2009) option stats was used to determine the number of final reads, average read length, and average coverage. The final average endogenous DNA content of all individuals (proportion of reads mapping to the human genome) was 10.39% (0.09–38.82%) (supplementary data 1A, Supplementary Material online).

### Calculating Genetic Sex Estimation

Genetic sex was calculated using the methods described in (Skoglund et al. 2013), estimating the fraction of reads mapping to Y chromosome out of all reads mapping to either X or Y chromosome. Additionally, sex was determined using a method described in (Fu et al. 2016), calculating the X and Y ratio by the division of the coverage by the autosomal coverage. Here, the sex was calculated for samples with a coverage >0.01× and only reads with a mapping quality >10 were counted for the autosomal, X, and Y chromosomes (supplementary data 1A, Supplementary Material online).

### Determining mtDNA Haplogroups

Mitochondrial DNA haplogroups were determined using Haplotype2 on the command line. For the determination, the reads were realigned to the reference sequence RSRS and the parameter -rsrs were given to estimate the haplogroups using Haplotype2 (van Oven and Kayser 2009; Weissensteiner et al. 2016) (supplementary data 1A, Supplementary Material online). Subsequently, the identical results between the individuals were checked visually by aligning mapped reads to the reference sequence using samtools 0.1.19 (Li et al. 2009) command tview and confirming the haplogroup assignment in PhyloTree. Additionally, private mutations were noted for further kinship analysis. The polymorphisms were estimated using the online platform of Haplotype2. Here, the variant calling files (vcf) were uploaded to the online platform and the known polymorphisms in the RSRS were converted to rCRS (supplementary data 1D, Supplementary Material online).

### Y Chromosome Variant Calling and Haplotyping

A total of 161,140 binary Y chromosome SNPs that have been detected as polymorphic in previous high coverage whole Y chromosome sequencing studies (Hallast et al. 2015; Karmin et al. 2015; Poznik et al. 2016) were called in nine male individuals with more than 0.01× autosomal coverage using ANGSD-0.916 (Hallast et al. 2015; Karmin et al. 2015; Poznik et al. 2016) –doHaploCall option. A subset of 826 sites had at least one of the nine individuals with a derived allele (supplementary data 1C, Supplementary Material online). Basal haplogroup affiliations of each sample were determined by assessing the proportion of derived allele calls in a set of primary haplogroup defining internal branches, as defined in (Karmin et al. 2015), using 1,677 informative sites.

### Kinship Analysis

All newly generated individuals with an average human coverage more than 0.03× were selected to assess kinship relationships up to the third degree (Saupe et al. 2021; Wang et al. 2021; Žegarac et al. 2021). We divided the individuals in different groups regarding their geographical area and resulting relationships. First, we were called all individuals using ANGSD-0.916 (Korneliussen et al. 2014) command –doHaploCall to sample a random base for the positions that are present at MAF > 0.1 in the 1000 Genomes GBR population (1000 Genomes Project Consortium et al. 2015) giving a total of 4,045,514 SNPs for autosomal kinship analysis. The ANGSD output files were converted to .tped format and used as an input for kinship analyses with READ (Monroy Kuhn et al. 2018), known to be reliable when working with sequencing depth in the order of 0.1× and routinely used in aDNA studies. First, we tested all individuals together and found one first degree relationship between ORD001 and ORD009 and two second degree relationships between ORD004, SAL007, and SAL011. We tested those relationships using all individuals from Ordona including SAL007 and SAL011 and all individuals from Salapia including ORD004. We compared the estimated mtDNA haplogroups, radiocarbon dates, genetic sexes, average human coverages, and the age of death from the individuals to confirm or dismiss relationships (supplementary data 1 and fig. 1, Supplementary Material online). Based on the outputs, we dismissed the possible second degree relationships.

### Data Set Preparation and Preprocessing for Autosomal Analysis

We assembled a genome-wide data set of ancient and modern samples by converting the following data sets, where needed, in PLINK format using convertf from the EIGENSOFT 8.0.0 (Patterson et al. 2006) and merging them together with PLINK 1.9 (Chang et al. 2015): 1) the ancient individuals and the Mbuti Congolese individuals (HGDP; Patterson et al. 2012) from the “1240K” data set from Dr David Reich laboratory (<https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>, version 44.3 release, last accessed January 21, 2021), 2) the modern samples from “1240K+HO” of the same release, 3) the Italian Chalcolithic and Bronze Age

samples produced by a recent study (Saupe et al. 2021), 4) the genotypes of an Italian hunter-gatherer from the Late Epigravettian site of Riparo Tagliente dated around 17kya (Bortolini et al. 2021), 5) additional genome-wide data of 129 modern-day Italian samples covering the entire country (Raveane et al. 2019), 6) genome-wide data of 217 present-day Apulian individuals (Sallustio et al. 2015), of which 41 individuals (unrelated and filtered for PCA outliers) were used for qpAdm and  $f_3$  analyses, and 7) 48 haploid genomes representing the Eurasian component of modern Ethiopians (here called Amhara\_NAF for “Non African Fraction”; Molinaro et al. 2019). We excluded the ancient samples showing issues in their group assignment (i.e., “Ignore” in the “1240K” data set) or in their DNA quality (e.g., those not containing the string “PASS,” with the exception of the *Iceman* individual, or those with less than 5,000 SNPs) and one of a pair of related samples. We further refined the list of ancient samples by retaining only those coming from geographical regions relevant for this study (supplementary data 2, Supplementary Material online) (Keller et al. 2012; Gamba et al. 2014; Lazaridis et al. 2014, 2016, 2017; Allentoft et al. 2015; Günther et al. 2015; Jones et al. 2015, 2017; Mathieson et al. 2015, 2018; Olalde et al. 2015, 2018, 2019; Broushaki et al. 2016; Cassidy et al. 2016; Fu et al. 2016; Hofmanová et al. 2016; Kılınç et al. 2016; Martiniano et al. 2016, 2017; Omrak et al. 2016; Schiffels et al. 2016; González-Fortes et al. 2017, 2019; Haber et al. 2017, 2019, 2020; Lipson et al. 2017; Saag et al. 2017, 2019; Schuenemann et al. 2017; Unterländer et al. 2017; van den Brink et al. 2017; Amorim et al. 2018; de Barros Damgaard, Marchi, et al. 2018; de Barros Damgaard, Martiniano, et al. 2018; Fregel et al. 2018; Harney et al. 2018; Krzewińska et al. 2018; Mitnik et al. 2018; Valdiosera et al. 2018; Veeramah et al. 2018; Zalloua et al. 2018; Antonio et al. 2019; Brace et al. 2019; Feldman, Fernández-Domínguez, et al. 2019; Feldman, Master, et al. 2019; Järve et al. 2019; Narasimhan et al. 2019; Sikora et al. 2019; Villalba-Mouco et al. 2019; Wang et al. 2019; Agranat-Tamir et al. 2020; Brunel et al. 2020; Fernandes et al. 2020; Furtwängler et al. 2020; Gokhman et al. 2020; Marcus et al. 2020; Margaryan et al. 2020; Rivollat et al. 2020; Skourtanioti et al. 2020). The same geographical filter was applied on the modern samples from “1240K+HO” and only those flagged as “PASS (genotyping)” have been selected in order to minimize variation due to different genotyping techniques (supplementary data 3, Supplementary Material online) (Patterson et al. 2012; Lazaridis et al. 2014, 2016; Biagini et al. 2019; Jeong et al. 2019). In case of duplicated samples, we selected the item showing the highest number of SNPs among the 1240K SNP set. Finally, we added the newly generated ancient Apulian samples to the merge.

We used two different SNPs set for the following analyses. The autosomal “1240K” SNP set (1,150,639 SNPs) was used for the analyses involving just ancient samples, such as the  $f$ -statistics and qpAdm analyses (see below). Conversely, the autosomal “HO” SNP set was used for those analyses relying also on modern populations (e.g., PCA and ADMIXTURE). In the latter case, we further refine the “HO” SNP set by removing monomorphic SNPs and those with more than 5% missing

count in modern populations with PLINK, thus obtaining 503,062 SNPs.

### Principal Component Analysis

We performed PCA using the program smartpca implemented in EIGENSOFT software 8.0.0 (Patterson et al. 2006) with the parameters “*lsqproject*” and “*shrinkmode*.” In particular, we projected the ancient samples’ genetic variation onto the principal components inferred from the modern samples of the “1240K+HO” data set. Modern Italian samples coming from Raveane et al. (2019), the modern Apulians from Sallustio et al. (2015), and the Amhara\_NAF were also projected.

### Admixture

We used the same data set of modern and ancient individuals presented in the PCA to explore the genetic components of each individual. For the analysis, we exploited the model-based algorithm implemented in ADMIXTURE (Alexander et al. 2009) projecting ancient individuals (-P flag) into the genetic structure calculated on the modern data set, due to missing data in the ancient individuals. We performed unsupervised Admixture for  $K \in \{2, 10\}$  for modern individuals and used the “per-cluster” inferred allele frequencies to project the ancient individuals. We visualized the Q output using R 3.6 (R Core Team 2021).

### $f_4$ Statistics

We performed a series of  $f_4$  statistics using the program *qpDstat* (option *f4mode*) implemented in the software ADMIXTOOLS 7.0.2 (Patterson et al. 2012) in different forms.  $f_4$  statistics are computed for a set of four different populations (W, X, Y, Z) and the sign of the resultant Z score indicates the direction of the gene flow: if it is positive, the gene flow occurred either between W and Y or X and Z, whereas if it is negative the gene flow occurred either between W and Z or X and Y). We grouped together ancient published samples from the same cultural/archaeological assemblages (labels used for  $f_3$  and  $f_4$  analyses can be found in the column “Analysis.Label” of supplementary data 2, Supplementary Material online). Conversely, ancient Apulian samples were both analyzed as individual samples and grouped together according to their archaeological age (Iron Age samples were grouped together under the acronym “IAA,” Iron Age Apulians). About 10 individuals from the Mbuti population from Congo coming from the “1240K” data set were used as an outgroup (population Z).

### Outgroup $f_3$ Statistics

To investigate genetic affinities between ancient Apulian samples with other ancient and modern human populations, we performed a series of Outgroup  $f_3$  analyses in the form  $f_3(X, Y; \text{outgroup} = \text{Mbuti})$  using the *qp3Pop* function implemented in ADMIXTOOLS 7.0.2 (Patterson et al. 2012). As above-described, we used the “Analysis.Label” groups for the ancient samples, whereas we used the “Group.Label” already present in the “1240K+HO” annotation file for modern samples.

**Table 1.** Archaeological Information, Dating, Genome Coverage, Genetic Sex, mtDNA, and Y Chromosome Haplogroups of the Individuals of This Study Selected for Genome-Wide Analysis.

| Individual | Site                 | Date                           | Genome Coverage | Gen. Sex | mtDNA HG      | Y Chromosome HG | No. SNPs |
|------------|----------------------|--------------------------------|-----------------|----------|---------------|-----------------|----------|
| ORD001     | Ordona               | 2516 ± 25 BP; 779–544 calBCE   | 0.0396          | XX       | H5c           | —               | 53,277   |
| ORD004     | Ordona               | 2374 ± 21 BP; 536–384 calBCE   | 0.0429          | XX       | U8b1b1        | R1b-M269        | 56,070   |
| ORD006     | Ordona               | 2476 ± 29 BP; 770–425 calBCE   | 0.12            | XY       | H + 16291T    | —               | 158,562  |
| ORD009     | Ordona               | 2433 ± 27 BP; 749–406 calBCE   | 0.995           | XX       | H5c           | —               | 805,966  |
| ORD010     | Ordona               | 984 ± 25 BP; 995–1156 calCE    | 0.91            | XX       | X2i           | —               | 769,336  |
| ORD011     | Ordona               | Not dated                      | 0.0889          | XY       | H1e           | R1b-P312        | 116,800  |
| ORD014     | Ordona               | 2438 ± 25 BP; 749–408 calBCE   | 0.108           | XY       | ISa2 + 16086C | J2b2-L283       | 141,736  |
| ORD019     | Ordona               | Not dated                      | 0.0484          | XY       | T2e           | I2d-Z2093/Y3670 | 61,292   |
| SAL001     | Salapia              | 2946 ± 30 BP; 1260–1048 calBCE | 0.049           | XY       | H1 + 16189!   | J2b-M241        | 64,574   |
| SAL003     | Salapia              | 2287 ± 24 BP; 401–354 calBCE   | 0.113           | XX       | K1a2a         | —               | 144,490  |
| SAL007     | Salapia              | Not dated                      | 0.0529          | XX       | K1a + 195T!   | —               | 64,034   |
| SAL010     | Salapia              | Not dated                      | 0.031           | XY       | U5a1          | J2b-M241        | 39,174   |
| SAL011     | Salapia              | 2241 ± 23 BP; 389–206 calBCE   | 0.0496          | XY       | U5b1          | I2d-M223        | 62,155   |
| SGR001     | San Giovanni Rotondo | 1285 ± 23 BP; 670–774 calCE    | 0.044           | XY       | U3a           | I1-M253         | 59,213   |
| SGR002     | San Giovanni Rotondo | 2451 ± 22 BP; 750–415 calBCE   | 0.102           | XY       | U5b1d1        | R1b-M269        | 131,437  |
| SGR003     | San Giovanni Rotondo | 2205 ± 26 BP; 368–176 calBCE   | 0.0572          | XX       | H1 + 16311T!  | —               | 76,687   |

NOTE.—Dates in BP are raw radiocarbon dates, calibrated dates are the 95.4% probability and were calibrated using IntCal20 (Reimer et al. 2020). See also [supplementary data 1](#), [Supplementary Material](#) online. Gen., genetic; HG, haplogroup.

We also computed the distance matrices from the Outgroup  $f_3$  results comparing ancient Apulian samples (both Iron and Middle Age samples) with other published ancient samples by subtracting the  $f_3$  values from 1 and we performed a classical (metric) MDS with default options using the function *cmdscale* in the stats package of R (version 4.0.4; R Core Team 2021).

### qpAdm

We exploited the qpWave/qpAdm framework within the software ADMIXTOOLS 7.0.2 to investigate the genetic composition of ancient Apulian samples and other ancient samples of interest with respect to human ancestries that contributed to coeval European population in the Mediterranean area. In particular, we tested a number of left sources from 2 to 5 in combination with a list of reference sources (right). We used the options “allsnps = YES” and “inbreed = NO” (the last option was used because some populations were composed of just one haploid individual). We tested four different sets of left sources. The “base” set included Western hunter-gatherers (*Loschbour\_published.DG*), Anatolian Neolithic (26 samples flagged as “Anatolia\_N” in the Analysis.Label column of [supplementary data 2](#), [Supplementary Material](#) online), Iranian Neolithic (samples *Iran\_GanjDareh\_N*), Caucasus hunter-gatherers (*KK1.SG* and *SATP.SG*), and Yamnaya (10 *Russia\_Samara\_EBA\_Yamnaya* samples). To this set of base sources, we separately added as left sources: Minoans (both *Greece\_Minoan\_Lassithi* and *Greece\_Minoan\_Odigritria*, but taken individually), Amhara\_NAF, and Roman Republican samples (*Italy\_Central\_Public\_1A*). In each run, we consistently used the same set of right sources (if present in the left set, they were removed from the right): *Russia\_Ust\_Ishim\_HG\_published.DG* (which was always kept as the first of the list), *Russia\_Kostenki14*, *Russia\_AfontovaGora3*, Eastern hunter-gatherers (*I0061 = Russia\_HG\_Karelia*, *I0124 = Russia\_HG\_Samara*,

*I0211 = Russia\_HG\_Karelia*, and *UzOO77 = Russia\_EHG*), *Spain\_ElMiron*, *Belgium\_UP\_GoyetQ116\_1\_published\_all*, *Levant\_N* (*Israel\_PPNB*), *Russia\_MA1\_HG.SG*, *Israel\_Natufian\_published*, *Czech\_Vestonice16*, and *Caucasus\_hunter-gatherers*.

### Ethical Statement

All ancient samples newly sequenced for this study are kept at the Museum of Anthropology of the University of Padova, hence we informed the territorially competent Soprintendenza (Soprintendenza Archeologia, Belle Arti e Paesaggio per l’Area metropolitana di Venezia e per le Province di Belluno, Padova e Treviso) about our sampling strategy and scientific research project, which was approved and our communication assigned the following protocol numbers: prot. 27903 (September 10, 2020) and prot. 29131 (November 23, 2020). We sampled individuals with multiple elements present, ideally an antimere to the sample taken (when possible) and took detailed photographs of material prior to performing the micro destructive sampling. Furthermore, we sampled the minimum amount that is necessary and we used protocols that allow for different types of biomolecular or chemical analysis (either simultaneously or in the future) to be carried out on the same bone material.

### Supplementary Material

[Supplementary data](#) are available at *Molecular Biology and Evolution* online.

### Acknowledgments

The study was supported by the Estonian Research Council grant PUT (PRG243) (A.S., T.K., M.M., C.L.S); the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.16-0030) (C.L.S., M.M., T.S., F.M., L.M.); the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012) (M.M.); and UniPd PRID 2019 (S.A., L.P.). The authors would like to thank Dr Francesco Bertolini

for facilitating the research of A.R. in the last stage of the manuscript preparation.

## Author Contributions

Conceptualization: L.P., S.A., and C.L.S. Formal analysis: S.A. and T.S. Funding acquisition: M.M. and L.P. Investigation: T.S. (aDNA data generation) and A.S. (quality control). Project administration: L.P. and C.L.S. Resources: C.L.S., N.C., and L.M. Supervision: L.P. and C.L.S. Visualization: S.A. and T.S. Writing—original draft: S.A., T.S., and L.P. Writing—review and editing: all authors.

## Data Availability

The DNA sequences generated during this study are available at the European Nucleotide Archive (ENA) at the accession number PRJEB47831. The data are also available in PLINK format through the data depository of the EBC-IG ([http://www.ebc.ee/free\\_data/AneliSaupe\\_2021/](http://www.ebc.ee/free_data/AneliSaupe_2021/)).

## References

- 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. 2015. A global reference for human genetic variation. *Nature* 526(7571):68–74.
- Abulafia D. 2011. The great sea: a human history of the Mediterranean. London: Allen Lane.
- Agranat-Tamir L, Waldman S, Martin MAS, Gokhman D, Mishol N, Eshel T, Cheronet O, Rohland N, Mallick S, Adamski N, et al. 2020. The genomic history of the Bronze Age Southern Levant. *Cell* 181(5):1146–1157.e11.
- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–1664.
- Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522(7555):167–172.
- Amorim CEG, Vai S, Posth C, Modi A, Koncz I, Hakenbeck S, La Rocca MC, Mende B, Bobo D, Pohl W, et al. 2018. Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nat Commun.* 9(1):3547.
- Aneli S, Caldon M, Saupe T, Montinaro F, Pagani L. 2021. Through 40,000 years of human presence in Southern Europe: the Italian case study. *Hum Genet.* 140(10):1417–1431.
- Antonio ML, Gao Z, Moots HM, Lucci M, Candilio F, Sawyer S, Oberreiter V, Calderon D, Devitofranceschi K, Aikens RC, et al. 2019. Ancient Rome: a genetic crossroads of Europe and the Mediterranean. *Science* 366(6466):708–714.
- Biagini SA, Solé-Morata N, Matisoo-Smith E, Zalloua P, Comas D, Calafell F. 2019. People from Ibiza: an unexpected isolate in the Western Mediterranean. *Eur J Hum Genet.* 27(6):941–951.
- 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(11):2484–2493.e7.
- Brace S, Diekmann Y, Booth TJ, van Dorp L, Faltskova Z, Rohland N, Mallick S, Olalde I, Ferry M, Michel M, et al. 2019. Author correction: ancient genomes indicate population replacement in Early Neolithic Britain. *Nat Ecol Evol.* 3(6):986–987.
- Broushaki F, Thomas MG, Link V, López S, van Dorp L, Kirsanow K, Hofmanová Z, Diekmann Y, Cassidy LM, Díez-Del-Molino D, et al. 2016. Early Neolithic genomes from the eastern Fertile Crescent. *Science* 353(6298):499–503.
- Brunel S, Bennett EA, Cardin L, Garraud D, Barrand Emam H, Beylier A, Boulestin B, Chenal F, Ciesielski E, Convertini F, et al. 2020. Ancient genomes from present-day France unveil 7,000 years of its demographic history. *Proc Natl Acad Sci U S A.* 117(23):12791–12798.
- Cassidy LM, Martiniano R, Murphy EM, Teasdale MD, Mallory J, Hartwell B, Bradley DG. 2016. Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proc Natl Acad Sci U S A.* 113(2):368–373.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4:7.
- Corrain C. 1986. Resti scheletrici umani da Ordona (Foggia), Secc. VII–IV aC. In: Iker R, editor. *Ordona VII/2. Les Tombes Dauniennes. Vol. 24. Etudes de Philologie, d'Archéologie et d'Histoire Ancienne Publiées par l'Institut Historique Belge de Rome.* Bruxelles, Rome: Academia Belgica. p. 787–807.
- Corrain C, Capitanio M, Ersperer G. 1972. I resti scheletrici della necropoli di Salapia (Cerignola), secoli IX–III aC. Padova: Società Cooperativa Tipografica.
- de Barros Damgaard B, Marchi N, Rasmussen S, Peyrot M, Renaud G, Korneliussen T, Moreno-Mayar JV, Pedersen MW, Goldberg A, Usmanova E, et al. 2018. 137 ancient human genomes from across the Eurasian steppes. *Nature* 557(7705):369–374.
- de Barros Damgaard P, Martiniano R, Kamm J, Víctor Moreno-Mayar J, Kroonen G, Peyrot M, Barjamovic G, Rasmussen S, Zacho C, et al. 2018. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* 360(6396):eaar7711.
- De Juliis EM. 1988. *Gli Iapigi: storia e civiltà della Puglia preromana.* Milano: Longanesi & C.
- Feldman M, Fernández-Domínguez E, Reynolds L, Baird D, Pearson J, Herskovitz I, May H, Goring-Morris N, Benz M, Gresky J, et al. 2019. Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia. *Nat Commun.* 10:1218.
- Feldman M, Master DM, Bianco RA, Burri M, Stockhammer PW, Mittnik A, Aja AJ, Jeong C, Krause J. 2019. Ancient DNA sheds light on the genetic origins of early Iron Age Philistines. *Sci Adv.* 5(7):eaax0061.
- Fernandes DM, Mittnik A, Olalde I, Lazaridis I, Cheronet O, Rohland N, Mallick S, Bernardo R, Brodmankhoshbacht N, Carlson J, et al. 2020. The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. *Nat Ecol Evol.* 4(3):334–345.
- Fregel R, Méndez FL, Bokbot Y, Martín-Socas D, Camalich-Massieu MD, Santana J, Morales J, Ávila-Arcos MC, Underhill PA, Shapiro B, et al. 2018. Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proc Natl Acad Sci U S A.* 115(26):6774–6779.
- 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(7606):200–205.
- Furtwängler A, Rohrlach AB, Lamnidis TC, Papac L, Neumann GU, Siebke I, Reiter E, Steuri N, Hald J, Denaire A, et al. 2020. Ancient genomes reveal social and genetic structure of Late Neolithic Switzerland. *Nat Commun.* 11(1):1915.
- Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, Domboróczki L, Kovári I, Pap I, Anders A, et al. 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nat Commun.* 5(1):5257.
- Gokhman D, Nissim-Rafinia M, Agranat-Tamir L, Housman G, García-Pérez R, Lizano E, Cheronet O, Mallick S, Nieves-Colón MA, Li H, et al. 2020. Differential DNA methylation of vocal and facial anatomy genes in modern humans. *Nat Commun.* 11(1):1189.
- González-Fortes G, Jones ER, Lightfoot E, Bonsall C, Lazar C, Grandal-d'Anglade A, Garralda MD, Drak L, Siska V, Simalcsik A, et al. 2017. Paleogenomic evidence for multi-generational mixing between Neolithic farmers and mesolithic hunter-gatherers in the lower Danube Basin. *Curr Biol.* 27(12):1801–1810.e10.
- González-Fortes G, Tassi F, Trucchi E, Henneberger K, Pajmans JLA, Díez-del-Molino D, Schroeder H, Susca RR, Barroso-Ruiz C, Bermudez FJ, et al. 2019. A western route of prehistoric human migration from Africa into the Iberian Peninsula. *Proc Biol Sci.* 286(1895):20182288.
- Günther T, Valdiosera C, Malmström H, Ureña I, Rodriguez-Varela R, Sverrisdóttir ÓO, Daskalaki EA, Skoglund P, Naidoo T, Svensson EM,

- et al. 2015. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc Natl Acad Sci U S A.* 112(38):11917–11922.
- Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K, et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522(7555):207–211.
- Haber M, Doumet-Serhal C, Scheib C, Xue Y, Danecek P, Mezzavilla M, Youhanna S, Martiniano R, Prado-Martinez J, Szpak M, et al. 2017. Continuity and admixture in the last five millennia of Levantine history from ancient Canaanite and present-day Lebanese genome sequences. *Am J Hum Genet.* 101(2):274–282.
- Haber M, Doumet-Serhal C, Scheib CL, Xue Y, Mikulski R, Martiniano R, Fischer-Genz B, Schutkowski H, Kivisild T, Tyler-Smith C. 2019. A transient pulse of genetic admixture from the crusaders in the near east identified from ancient genome sequences. *Am J Hum Genet.* 104(5):977–984.
- Haber M, Nassar J, Almarri MA, Saupe T, Saag L, Griffith SJ, Doumet-Serhal C, Chanteau J, Saghieh-Beydoun M, Xue Y, et al. 2020. A genetic history of the Near East from an aDNA time course sampling eight points in the past 4,000 years. *Am J Hum Genet.* 107(1):149–157.
- Hallast P, Batini C, Zadik D, Maisano Delser P, Wetton JH, Arroyo-Pardo E, Cavalleri GL, de Knijff P, Destro Bisol G, Dupuy BM, et al. 2015. The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. *Mol Biol Evol.* 32(3):661–673.
- Harney É, May H, Shalem D, Rohland N, Mallick S, Lazaridis I, Sarig R, Stewardson K, Nordenfelt S, Patterson N, et al. 2018. Publisher correction: ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. *Nat Commun.* 9(1):3913.
- Hodos T. 2020. The archaeology of the Mediterranean Iron Age: a globalising world c.1100–600 BCE. Cambridge: Cambridge University Press.
- Hofmanová Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Díez-Del-Molino D, van Dorp L, López S, Kousathanas A, Link V, et al. 2016. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc Natl Acad Sci U S A.* 113(25):6886–6891.
- Järve M, Saag L, Scheib CL, Pathak AK, Montinaro F, Pagani L, Flores R, Guellil M, Saag L, Tambets K, et al. 2019. Shifts in the genetic landscape of the Western Eurasian Steppe associated with the beginning and end of the Scythian dominance. *Curr Biol.* 29(14):2430–2441.e10.
- Jeong C, Balanovsky O, Lukianova E, Kahbatkyz N, Flegontov P, Zaporozhchenko V, Immel A, Wang C-C, Ixan O, Khussainova E, et al. 2019. The genetic history of admixture across inner Eurasia. *Nat Ecol Evol.* 3(6):966–976.
- Jones CP. 1987. Stigma: tattooing and branding in Graeco-Roman antiquity. *J Rom Stud.* 77:139–155.
- Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R, McLaughlin RL, Gallego Llorente M, Cassidy LM, Gamba C, et al. 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun.* 6:8912.
- Jones ER, Zarina G, Moiseyev V, Lightfoot E, Nigst PR, Manica A, Pinhasi R, Bradley DG. 2017. The Neolithic transition in the Baltic was not driven by admixture with early European farmers. *Curr Biol.* 27(4):576–582.
- Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29(13):1682–1684.
- Karmin M, Saag L, Vicente M, Wilson Sayres MA, Järve M, Talas UG, Rootsī S, Ilumäe A-M, Mägi R, Mitt M, et al. 2015. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* 25(4):459–466.
- 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.
- Kılıç GM, Omrak A, Özer F, Günther T, Büyükkarakaya AM, Biçakçı E, Baird D, Dönertas HM, Ghalichi A, Yaka R, et al. 2016. The demographic development of the first farmers in Anatolia. *Curr Biol.* 26(19):2659–2666.
- Kirigin B, Mišć M, Barbarić V. 2010. Palagruža – the Island of Diomedes: summary excavation report 2002–2008. *Hesperia.* 25:65–91.
- Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* 15:356.
- Krzewińska M, Kılıç GM, Juras A, Koptekin D, Chyleński M, Nikitin AG, Shcherbakov N, Shuteleva I, Leonova T, Kraeva L, et al. 2018. Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads. *Sci Adv.* 4(10):eaat4457.
- 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(7666):214–218.
- Lazaridis I, Nadel D, Rollefson G, Merrett DC, 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(7617):419–424.
- Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513(7518):409–413.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and samtools. *Bioinformatics* 25(16):2078–2079.
- 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(7680):368–372.
- Lombardo M. 2014. Iapygians: the indigenous populations of Ancient Apulia in the fifth and fourth centuries B.C.E. In: Carpenter T, Lynch K, Robinson E, editors. *The Italic people of ancient Apulia: new evidence from pottery for workshops, markets, and customs.* Cambridge: Cambridge University Press. p. 36–68.
- Malaspina A-S, Lao O, Schroeder H, Rasmussen M, Raghavan M, Moltke I, Campos PF, Sagredo FS, Rasmussen S, Gonçalves VF, et al. 2014. Two ancient human genomes reveal Polynesian ancestry among the indigenous Botocudos of Brazil. *Curr Biol.* 24(21):R1035–R1037.
- Marcus JH, Posth C, Ringbauer H, Lai L, Skeates R, Sidore C, Beckett J, Furtwängler A, Olivieri A, Chiang CWK, et al. 2020. Genetic history from the Middle Neolithic to present on the Mediterranean island of Sardinia. *Nat Commun.* 11(1):939.
- Margaryan A, Lawson DJ, Sikora M, Racimo F, Rasmussen S, Moltke I, Cassidy LM, Jørsboe E, Ingason A, Pedersen MW, et al. 2020. Population genomics of the Viking world. *Nature* 585(7825):390–396.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embo J.* 17(1):10–12.
- Martiniano R, Caffell A, Holst M, Hunter-Mann K, Montgomery J, Müldner G, McLaughlin RL, Teasdale MD, van Rheeën W, Veldink JH, et al. 2016. Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nat Commun.* 7:10326.
- Martiniano R, Cassidy LM, Ó'Maoldúin R, McLaughlin R, Silva NM, Manco L, Fidalgo D, Pereira T, Coelho MJ, Serra M, et al. 2017. The population genomics of archaeological transition in west Iberia: investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genet.* 13(7):e1006852.
- 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(7695):197–203.
- Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528(7583):499–503.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernitsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20(9):1297–1303.

- Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* 2010(6):db.prot5448.
- Mittnik A, Wang C-C, Pfrengle S, Daubaras M, Zariņa G, Hallgren F, Allmäe R, Khartanovich V, Moiseyev V, Törv M, et al. 2018. The genetic prehistory of the Baltic Sea region. *Nat Commun.* 9(1):442.
- Molinaro L, Montinaro F, Yelmen B, Marnetto D, Behar DM, Kivisild T, Pagani L. 2019. West Asian sources of the Eurasian component in Ethiopians: a reassessment. *Sci Rep.* 9(1):18811.
- Monroy Kuhn JM, Jakobsson M, Günther T. 2018. Estimating genetic kin relationships in prehistoric populations. *PLoS One* 13(4):e0195491.
- Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardo 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(6457):eaat7487.
- Nava ML, Descoedres JP. 1990. Greek and Adriatic influences in Daunia in the Early Iron Age. In: Descoedres JP, editor. *Greek colonists and native populations*. Canberra: Humanities Research Centre; Oxford: Clarendon Press. p. 560–578.
- Nijboer AJ. 2006. La cronologia assoluta dell'età del Ferro nel Mediterraneo, dibattito sui metodi e sui risultati. In: Bartoloni G, Delpino F, editors. *Oriente e Occidente: Metodi e Discipline a Confronto*. Pisa: Istituti Editoriali e Poligrafici Internazionali. p. 527–556.
- Norman C. 2011. The tribal tattooing of Daunian women. *Eur J Archaeol.* 14(1–2):133–157.
- Norman C. 2016. Daunian women: costume and actions commemorated in stone. In: Budin SL, Turfa JM, editors. *Women in antiquity: Real women across the Ancient World*. London; New York: Routledge, Taylor & Francis Group. p. 901–912.
- Olalde I, Brace S, Allentoft ME, 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(7695):190–196.
- Olalde I, Mallick S, Patterson N, Rohland N, Villalba-Mouco V, Silva M, Dulias K, Edwards CJ, Gandini F, Pala M, et al. 2019. The genomic history of the Iberian Peninsula over the past 8000 years. *Science* 363(6432):1230–1234.
- Olalde I, Schroeder H, Sandoval-Velasco M, Vinner L, Lobón I, Ramírez O, Civit S, García-Borja P, Salazar-García DC, Talamo S, et al. 2015. A common genetic origin for early farmers from Mediterranean Cardial and Central European LBK cultures. *Mol Biol Evol.* 32(12):3132–3142.
- Omrak A, Günther T, Valdiosera C, Svensson EM, Malmström H, Kiesewetter H, Aylward W, Storå J, Jakobsson M, Götherström A. 2016. Genomic evidence establishes Anatolia as the source of the European Neolithic gene pool. *Curr Biol.* 26(2):270–275.
- Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, et al. 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499(7456):74–78.
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics* 192(3):1065–1093.
- Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet.* 2(12):e190.
- Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, Wilson Sayres MA, Ayub Q, McCarthy SA, Narechania A, Kashin S, et al. 2016. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat Genet.* 48(6):593–599.
- R Core Team. 2021. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: <https://www.R-project.org/>.
- Raveane A, Aneli S, Montinaro F, Athanasiadis G, Barlera S, Birolo G, Boncoraglio G, Di Blasio AM, Di Gaetano C, Pagani L, et al. 2019. Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe. *Sci Adv.* 5(9):eaaw3492.
- Reimer PJ, Austin WEN, Bard E, Bayliss A, Blackwell PG, Ramsey CB, Butzin M, Cheng H, Lawrence Edwards R, Friedrich M, et al. 2020. The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* 62(4):725–757.
- Rivollat M, Jeong C, Schiffels S, Küçükkalıççı İ, Pemonge M-H, Rohrlach AB, Alt KW, Binder D, Friedrich S, Ghesquière E, et al. 2020. Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. *Sci Adv.* 6(22):eaaz5344.
- Saag L, Laneman M, Varul L, Malve M, Valk H, Razzak MA, Shirobokov IG, Khartanovich VI, Mikhaylova ER, Kushniarevich A, et al. 2019. The arrival of Siberian ancestry connecting the Eastern Baltic to Uralic speakers further East. *Curr Biol.* 29(10):1701–1711.e16.
- Saag L, Varul L, Scheib CL, Stenderup J, Allentoft ME, Saag L, Pagani L, Reidla M, Tambets K, Metspalu E, et al. 2017. Extensive farming in Estonia started through a sex-biased migration from the Steppe. *Curr Biol.* 27(14):2185–2193.e6.
- Sallustio F, Cox SN, Serino G, Curci C, Pesce F, De Palma G, Papagianni A, Kirmizis D, Falchi M, Schena FP, et al. 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(7):940–948.
- 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(12):2576–2591.e12.
- Scaggion C, Carrara N. 2016. New studies on human skeletal remains from the ancient Herdonia (southeast Italy). Evidences of tuberculosis and brucellosis: two diseases connected with farm animals. *Antrocom* 12:85–96.
- Schiffels S, Haak W, Paajanen P, Llamas B, Popescu E, Loe L, Clarke R, Lyons A, Mortimer R, Sayer D, et al. 2016. Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nat Commun.* 7:10408.
- Schuenemann VJ, Peltzer A, Welteveld B, van Pelt WP, Molak M, Wang C-C, Furtwängler A, Urban C, Reiter E, Nieselt K, et al. 2017. Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nat Commun.* 8:15694.
- Sikora M, Pitulko VV, Sousa VC, Allentoft ME, Vinner L, Rasmussen S, Margaryan A, de Barros Damgaard P, de la Fuente C, Renaud G, et al. 2019. The population history of northeastern Siberia since the Pleistocene. *Nature* 570(7760):182–188.
- Skoglund P, Storå J, Götherström A, Jakobsson M. 2013. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J Archaeol Sci.* 40(12):4477–4482.
- Skourtanioti E, Erdal YS, Frangipane M, Balossi Restelli F, Yener KA, Pinnock F, Matthiae P, Özbal R, Schoop U-D, Guliyev F, et al. 2020. Genomic history of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus. *Cell* 181(5):1158–1175.e28.
- Tagliente M. 1986. I signori dei cavalli nella Daunia di età arcaica. *Annali Della Facoltà di Lettere e Filosofia Dell'Università Degli Studi di Perugia.* 23:1985–1986.
- Unterländer M, Palstra F, Lazaridis I, Pilipenko A, Hofmanová Z, Groß M, Sell C, Blöcher J, Kirsanow K, Rohland N, et al. 2017. Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. *Nat Commun.* 8:14615.
- Valdiosera C, Günther T, Vera-Rodríguez JC, Ureña I, Iriarte E, Rodríguez-Varela R, Simões LG, Martínez-Sánchez RM, Svensson EM, Malmström H, et al. 2018. Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. *Proc Natl Acad Sci U S A.* 115(13):3428–3433.
- van den Brink ECM, Beeri R, Kirzner D, Bron E, Cohen-Weinberger A, Kamaisky E, Gonen T, Gershony L, Nagar Y, Ben-Tor D, et al. 2017. A Late Bronze Age II clay coffin from Tel Shaddud in the Central Jezreel Valley, Israel: context and historical implications. *Lевант* 49(2):105–135.
- van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 30(2):E386–E394.
- Veeramah KR, Rott A, Groß M, van Dorp L, López S, Kirsanow K, Sell C, Blöcher J, Wegmann D, Link V, et al. 2018. Population genomic analysis of elongated skulls reveals extensive female-biased

- immigration in Early Medieval Bavaria. *Proc Natl Acad Sci U S A.* 115(13):3494–3499.
- Villalba-Mouco V, van de Loosdrecht MS, Posth C, Mora R, Martínez-Moreno J, Rojo-Guerra M, Salazar-García DC, Royo-Guillén JL, Kunst M, Rougier H, et al. 2019. Survival of Late Pleistocene hunter-gatherer ancestry in the Iberian Peninsula. *Curr Biol.* 29(7):1169–1177.e7.
- 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(1):590.
- Wang RJ, Al-Saffar SI, Rogers J, Hahn MW. 2021. Human generation times across the past 250,000 years. *bioRxiv* 2021.09.07.459333. doi: <https://doi.org/10.1101/2021.09.07.459333>.
- Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt H-J, Kronenberg F, Salas A, Schönherr S. 2016. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* 44(W1):W58–W63.
- Zalloua P, Collins CJ, Gosling A, Biagini SA, Costa B, Kardailsky O, Nigro L, Khalil W, Calafell F, Matisoo-Smith E. 2018. Ancient DNA of Phoenician remains indicates discontinuity in the settlement history of Ibiza. *Sci Rep.* 8(1):17567.
- Žegarac A, Winkelbach L, Blöcher J, Diekmann Y, Krečković Gavrilović M, Porčić M, Stojković B, Milašinović L, Schreiber M, Wegmann D, et al. 2021. Ancient genomes provide insights into family structure and the heredity of social status in the early Bronze Age of southeastern Europe. *Sci Rep.* 11(1):10072.