Tracking the transition to agriculture in Southern Europe through ancient DNA analysis of dental calculus

Claudio Ottoni[a,1,2], Dušan Borić[b,c], Olivia Cheronet[d], Vitale Sparacello[e], Irene Dori[f], Alfredo Coppa[c,d,g], Dragana Antonović[h], Dario Vujević[i], T. Douglas Price[j], Ron Pinhasi[d], and Emanuela Cristiani[a,2]

[a]DANTE - Diet and Ancient Technology Laboratory, Department of Oral and Maxillo-Facial Sciences, Sapienza University of Rome, 00161 Rome, Italy; [b]The Italian Academy for Advanced Studies in America, Columbia University, New York, NY 10027; [c]Department of Environmental Biology, Sapienza University of Rome, 00185 Rome, Italy; [d]Department of Evolutionary Anthropology, University of Vienna, 1090 Vienna, Austria; [e]Department of Environmental and Life Sciences, University of Cagliari, 09042 Monserrato, Italy; [f]Soprintendenza Archeologia, Belle Arti e Paesaggio per le province di Verona, Rovigo e Vicenza, 37121 Verona, Italy; [g]Department of Genetics, Harvard Medical School, Harvard University, Cambridge, MA 02138; [h]Institute of Archaeology, 11000 Belgrade, Serbia; [i]Department of Archaeology, University of Zadar, 23000 Zadar, Croatia; and [j]Laboratory for Archaeological Chemistry, University of Wisconsin–Madison, Madison, WI 53706

Archaeological dental calculus, or mineralized plaque, is a key tool to track the evolution of oral microbiota across time in response to processes that impacted our culture and biology, such as the rise of farming during the Neolithic. However, the extent to which the human oral flora changed from prehistory until present has remained elusive due to the scarcity of data on the microbiomes of prehistoric humans. Here, we present our reconstruction of oral microbiomes via shotgun metagenomics of dental calculus in 44 ancient foragers and farmers from two regions playing a pivotal role in the spread of farming across Europe—the Balkans and the Italian Peninsula. We show that the introduction of farming in Southern Europe did not alter significantly the oral microbiomes of local forager groups, and it was in particular associated with a higher abundance of the species *Olsenella* sp. oral taxon 807. The human oral environment in prehistory was dominated by a microbial species, Anaerolineaceae bacterium oral taxon 439, that diversified geographically. A Near Eastern lineage of this bacterial commensal dispersed with Neolithic farmers and replaced the variant present in the local foragers. Our findings also illustrate that major taxonomic shifts in human oral microbiome composition occurred after the Neolithic and that the functional profile of modern humans evolved in recent times to develop peculiar mechanisms of antibiotic resistance that were previously absent.

ancient DNA | dental calculus | metagenomics | Southern Europe

Studies over the last decade testified to the paramount importance of microbes living in symbiosis with humans to host physiology and ecology—from pathogen susceptibility to stress response and metabolism (1). The study of the genetic composition of microbial communities—the microbiome—has revolutionized our understanding of the ecology and evolution of living organisms. The oral microbiome in particular is an extremely rich and diverse driver of health and disease in humans (2), exhibiting population-specific variation (3). Recent studies demonstrated that dental calculus, mineralized plaque accumulating on the tooth surface, provides a biomolecular record of microorganisms, host cells, and debris from food, environment, and craft activities (4, 5).

The ancient DNA (aDNA) retrieved from the mineral matrix of dental calculus can be used to reconstruct oral microbiomes of ancient humans and animals, and potentially infer diet, health, and lifestyles from the past (6–11). By comparing human oral microbiomes from archaeological dental calculus and modern plaque samples, previous studies detected shifts in microbial composition purportedly driven by dietary changes that occurred with the rise of farming ca. 11,000 y ago, during the Neolithic and, more recently, the Industrial Revolution (8, 9). However, recent evidence demonstrated that the microbial composition of dental plaque and calculus is different, hence the pitfalls of using modern plaque as a comparative dataset (7). Furthermore, oral microbiomes from

early human prehistory, in particular from ancient foragers and farmers, are still poorly understood with the analysis of only a handful of individuals (8, 9). Investigating the oral environment of ancient human populations in conjunction with their archaeological contexts and subsistence economies is key to a comprehensive understanding of the complex dynamics behind the evolution of human oral microbiomes.

Here, we use ancient dental calculus to reconstruct the oral microbiomes of 44 individuals chronologically spanning more than 10 millennia, from the Paleolithic to the Middle Ages, in two key regions bridging the spread of farming across Southern Europe—the Balkans and the Italian Peninsula (Fig. 1 and Dataset S1). A total of 28 individuals, dated to the Upper Paleolithic, Mesolithic, and Early Neolithic, along with two samples from the Bronze Age and the Medieval period, originate from the Danube Gorges area of the central Balkans, Romania, and Croatia. The Danube Gorges region provides key evidence of a long-lasting Late Glacial and

Significance

The oral microbial community living in symbiosis with humans is a rich and diverse driver of health and disease that is strongly influenced by our ecology and lifestyle. However, its evolution across human prehistory remains elusive. By analyzing the DNA entrapped in archaeological dental calculus, we characterize the oral microbiomes of 44 prehistoric foragers and farmers from Southern Europe. We demonstrate that the genome of an oral bacteria diversified geographically and recorded one of the most dramatic changes in our biological and cultural history, the spread of farming. The transition to agriculture did not alter significantly the oral microbiome of ancient humans, whereas more significant changes occurred later in history, including the development of peculiar antibiotic resistance pathways.

www.manaraa.com

Fig. 1. Geographic provenance of the dental calculus samples analyzed in this study.

Early Holocene occupation along a riverine environment. The unique Mesolithic and Early Neolithic archaeological record of mortuary practices and the wealth of recently obtained biomolecular evidence from this region reveals an important case for contacts and admixture between local foragers with western and eastern Eurasian hunter-gatherer genetic ancestries and farming groups of northwestern-Anatolian ancestry who spread along the Danubian route from the mid-seventh millennium BC (12–14).

A total of 10 individuals chronologically spanning the period from the Late Paleolithic to the Bronze Age were sampled from the funerary human record of Grotta Continenza, in the Fucino Basin of central Italy. Since the last phases of the Late Epigravettian, Grotta Continenza was repeatedly used as a dwelling and a burial place (15), thereby representing one of a few places in Italy that yielded different phases of forager and early farmer occupation and associated mortuary remains in a single stratigraphic sequence. A recent study revealed the typical Western hunter-gatherer ancestry of Mesolithic humans from Grotta Continenza (16).

Six further individuals in our samples originate from the sites of Arene Candide and Arma dell'Aquila in the Finalese area of Liguria, northwestern Italy. The area is the testimony of the western diffusion of the Neolithic along a Mediterranean route reaching parts of Northern Italy around 5,800 to 5,600 BC, with the establishment of the Impresso-Cardial chrono-cultural complex (ICC), followed by the Square-Mouthed Pottery Culture (SMP) beginning ca. 5,000 BC (17). Four individuals come from Arene Candide Cave and two from the nearby site of Arma dell'Aquila; five are directly dated to the SMP Neolithic and one to the ICC (18).

By documenting the taxonomic composition and the functional activity of the human oral microbiota in a large number of samples prior to and after the adoption of agriculture, we tested whether the transition to farming that started with the Neolithic changed the human oral microbiome in the Balkans and Italy. Furthermore, by comparing the ancient microbiomes of this study with those reconstructed from ancient and modern dental calculus samples in the literature, we tested whether human oral flora changed over time, from prehistory until the present. Finally, the spatiotemporal coverage of oral microbial variation provided by our study made it possible to identify a commensal species that accompanied human groups during the Neolithization process, *Anaerolineaceae bacterium oral taxon 439*. We examined the

phylogeographic patterns of variation in the genomes of this commensal species demonstrating its potential to infer migratory trajectories and biological interactions in human populations from the past.

Results

DNA Preservation. Between 26 and 61 million DNA reads were generated after sequencing for each sample. Of these, after adapter trimming and quality filtering, 15 to 36% could be assigned to Bacteria and Archaea taxa with Kraken2 (Dataset S2). The results of the Sourcetracker analysis showed that except for the sample VK1 from Croatia, which was dominated by soil microorganisms, over 90% of the classified reads belonged to oral taxa, while extremely low fractions were identified as stemming from the laboratory controls, skin, and soil microbiomes (Fig. 2 and Dataset S3). This pattern was confirmed in a nonmetric multidimensional scaling of Bray–Curtis distances calculated from normalized abundance reads at the species level, in which all the ancient samples investigated in this study, except for VK1, clustered with oral microbiomes from the literature (*SI Appendix*, Fig. S1).

Alignment of the reads to the reference sequence of the most abundant species detected (Dataset S4) revealed terminal deamination patterns ranging on average 17 to 46%, with the lowest values observed in the Medieval sample LV29, as expected based on biomolecular diagenetic trajectories (19). We found that the Paleolithic and Mesolithic samples from Grotta Continenza, regardless of their older chronology, consistently showed lower deamination values and higher DNA fragment length distributions across different microbial species (*SI Appendix*, Fig. S2). This indicates an overall higher degree of DNA preservation in Grotta Continenza, which is most likely due to the peculiar taphonomic conditions of the site, a cave at 710 m above sea level. Preservation of host human DNA in the dental calculus samples appeared to be low, with yields ranging 0.01 to 0.79% (below 0.1% in over 70% of the samples, Dataset S4). However, the fairly low deamination rates detected in some samples indicates that human DNA in some instances may stem from contaminant sources that could not be removed by the ultraviolet (UV)- and ethylenediaminetetraacetic acid (EDTA)-based decontamination procedures adopted, possibly originating from postexcavation procedures (20). Overall, this confirms the low efficiency of shotgun sequencing to retrieve significant proportions of host DNA molecules in ancient dental calculus (21).

To rule out any potential taxonomic skew associated with loss of AT-rich DNA fragments in archaeological calculus samples, as previously reported (22), we screened the GC content to fragment length in four highly abundant and discriminant microbial species across different temporal bins of samples from this study and the literature. We found that the GC content of shorter reads was not significantly higher than the overall mean value in the chronological samples investigated, either prehistoric and historic, even in the low GC-content archaeal species *Methanobrevibacter oralis* (*SI Appendix*, Fig. S3).

Taxonomic Changes in Human Oral Microbiomes over Time. To characterize the oral microbiomes of the Mesolithic and Neolithic individuals from the Danube Gorges and Italy within a wider framework of oral microbiome variation, we used a comparative dataset of prehistoric, historic, and modern dental calculus of human samples, as well as modern chimpanzees from the literature (Dataset S5). A principal component analysis (PCA) of microbial species abundances (Fig. 3 and Dataset S6) showed that modern human plaque samples clustered in the bottom left of the plot, modern chimpanzee dental calculus in the bottom right, while ancient and modern human calculus samples were grouped in the upper part of the plot. Of these, the modern samples were on the left, the ancient samples from the
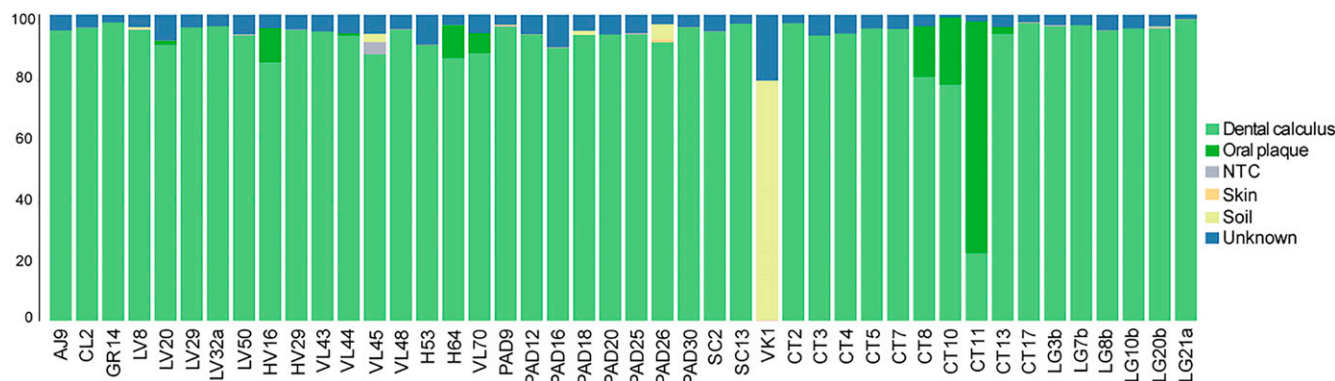
Fig. 2. Stacked bar plot of Sourcetracker analysis showing the proportion of Kraken2 classified reads at the species level stemming from modern dental calculus, modern plaque, skin, soil, and laboratory controls (NTC) in the dental calculus samples analyzed here.

Danube Gorges and Italy, together with other prehistoric samples from the literature, clustered mostly on the right, whereas historical samples were more scattered from left to right. The plot of PCA loadings indicated that Anaerolineaceae bacterium oral taxon 439, *M. oralis*, *Desulfomicrobium orale*, and *Desulfobulbus oralis*, among others, were responsible for the positioning of the prehistoric samples in the top right of the plot (including most of the Mesolithic and Neolithic individuals from the Danube Gorges and Italy, and Neanderthal samples, among others) (*SI Appendix*, Fig. S4).

To test whether any changes in microbial composition occurred over time, we contrasted the microbial species abundances in chronological and geographic groups through permutational multivariate analysis of variance (PERMANOVA) (Dataset S7). We found no significant differences between the Late Mesolithic foragers and the Neolithic farmers in the Danube Gorges ($P = 0.1$, adonis test). The analysis of differential species abundances with DESeq2 and Statistical Analysis of Metagenomic Profiles (STAMP) showed that the Neolithic farmers possessed a higher frequency *Olsenella* sp. oral taxon 807 (Dataset S8 and *SI Appendix*, Figs. S5 and S6). Similarly, in Italy, the oral microbiome of the Paleolithic and Mesolithic foragers from Grotta Continenza was not significantly different from that of the Neolithic farmers from Liguria, at the Finalese sites ($P = 0.037$). The analysis of differential species abundances showed that *Olsenella* sp. oral taxon 807 and Anaerolineaceae bacterium oral taxon 439 were

more abundant in the Neolithic farmers, whereas *Streptococcus sanguinis* was higher in the Mesolithic foragers (Dataset S8 and *SI Appendix*, Figs. S5 and S6). When contrasting the chronological groups of this study geographically, we found significant differences ($P < 0.001$, Datasets S7 and S8) between the foragers from Italy and the Danube Gorges, thus indicating geographic structuring in the oral microbiome composition. Conversely, the farmer groups appeared to possess more homogenous oral microbial compositions in the two regions ($P = 0.19$).

When testing the microbial species abundances against present-day calculus and historic samples dated to the eighteenth to nineteenth century AD, all the prehistoric samples of this study appeared to be significantly different ($P < 0.01$, adonis test, Dataset S7). In particular, both the Italian and the Danube Gorges Neolithic farmers showed a higher frequency of Anaerolineaceae bacterium oral taxon 439 and *Olsenella* sp. oral taxon 807 (Dataset S8 and *SI Appendix*, Figs. S5 and S6), whereas *Rothia aeria* and *Lautropia mirabilis* were among the predominant species in present-day samples. Anaerolineaceae bacterium oral taxon 439 was found to be differentially abundant also in the Mesolithic foragers from the Danube Gorges, together with *D. orale*, while the historic sample was mostly characterized by the higher abundance of the archaea *M. oralis*.

Finally, we compared our ancient samples against a group of historic wild chimpanzees, as representatives of typical foraging behavior. In all instances, the oral microbiome of prehistoric

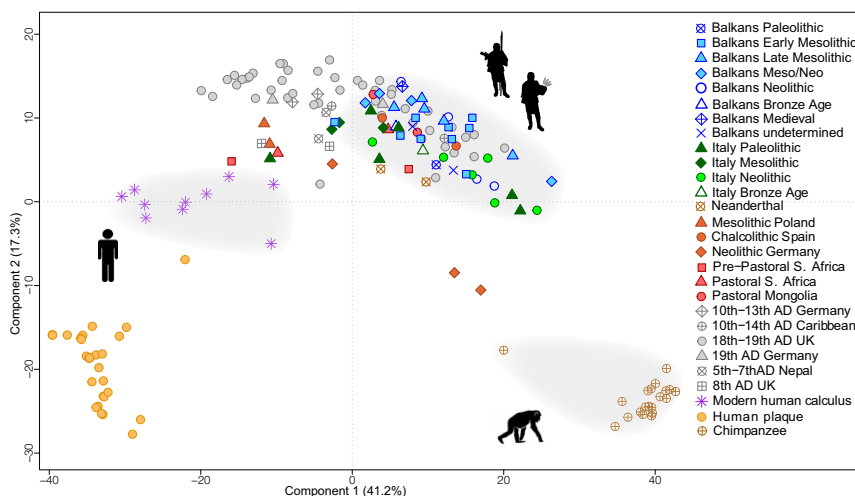**Fig. 3.** PCA of oral microbial species abundances from this study and the literature.

Ottoni et al.
Tracking the transition to agriculture in Southern Europe through ancient DNA analysis of dental calculus

PNAS | 3 of 11
https://doi.org/10.1073/pnas.2102116118
www.manaraa.com

foragers and farmers from Italy and the Danube Gorges was significantly different ($P < 0.01$, adonis test). In particular, we detected in the chimpanzees a lower frequency of Anaerolineaceae bacterium oral taxon 439, *L. mirabilis*, and oral *Streptococcus* species (Datasets S7 and S8 and *SI Appendix*, Figs. S5 and S6) and a higher frequency of *Porphyromonas gingivalis* and *Lawsonella clevelandensis*.

To rule out any potential biases associated with oral geography (23), we compared microbial composition in groups of calculus samples based on tooth type. By contrasting exclusively anterior teeth (incisors/canines), we still found significant differences between the prehistoric samples of this study and the modern and historic (eighteenth to nineteenth century) samples from the literature (7) ($P < 0.01$, adonis test, Dataset S7). Significant differences were also found when contrasting microbial abundances in molars between the prehistoric samples and the historic eighteenth to nineteenth century sample ($P < 0.01$). On the other hand, no difference was found between ancient calculus from molars and anterior teeth of the samples from this study ($P = 0.175$). This suggests that oral geography did not affect significantly the results of our tests, even though more systematic studies based on a larger number of samples and tooth types may help in the future to address this issue in more depth.

Functional Analysis. To explore the metabolic functional profiles of the calculus samples and test potential differences in metabolic pathways categories in the chronological samples that we investigated, we used the SEED subsystem classification (24). A PCA of the SEED profiles (Fig. 4*A* and Dataset S9) showed that the ancient foragers and farmers of this study clustered between the chimpanzees and the modern calculus samples. By testing the differences in the metabolic profiles through PERMANOVA, we found no differences between the foragers and farmers in either the Danube Gorges ($P = 0.174$, adonis test) or Italy ($P = 0.071$) (Dataset S10). On the other hand, they were all functionally different from both the chimpanzees and the modern humans ($P < 0.01$), though in some instances we could not reject data dispersion effects due to heterogeneity of variances. As already observed in the taxonomic analysis, the eighteenth to nineteenth century AD historic calculus samples were more scattered in the PCA from the right to the left end of the plot.

The analysis of the PCA loadings, together with DESeq2 and sparse Least Square Discriminant Analysis (sPLSDA), revealed the enrichment in the oral microbiomes of the Mesolithic foragers and Neolithic farmers from the Danube Gorges and Italy, as well as the chimpanzees, of the pathways Trans-envelope signaling system VreARI, Sigma54-dependent transcription related gene cluster and protein secretion system type I, which are associated with bacterial pathogenesis and virulence and environmental and physiological stress (25–27) (Dataset S11 and *SI Appendix*, Fig. S7). To a lesser extent, the Mesolithic foragers from the Danube Gorges and Italy showed the enrichment of the utilization systems for glycans and polysaccharides pathway, detected also in the chimpanzees. This pathway is associated with carbohydrate-active enzymes that enable the breakdown of dietary glycans, including, among others, plant storage and cell wall polysaccharides (28).

In modern calculus, we detected a higher representation of pathways associated with lipoprotein transporters to outer membranes. The outer membrane in gram-negative bacteria constitutes a permeability barrier against antibiotics (29), and lipoproteins are involved in several cell mechanisms, including antibiotic efflux pumps (30). Antibiotic microbial resistance (AMR) pathways have been documented in the human oral microbiome (31) and were detected in ancient Medieval human calculus samples (11). To test whether AMR pathways in the human oral microbiota changed over time in our comparative dataset, we used the Basic Local

Alignment Search Tool (BLAST) and the Comprehensive Antibiotic Resistance Database (CARD) (10, 32). The top match of each read was assigned to the respective Antibiotic Resistance Ontology (ARO) accession number and AMR gene family, as previously described (10), and the reads abundance of each family was normalized for the number of reads matching the microbial housekeeping gene *recA* (Dataset S12). Differences in AMR gene family's abundance were tested among chronological groups from this study and the literature (Dataset S13 and *SI Appendix*, Fig. S8*A*), and the results were confirmed after normalization with two other housekeeping genes, *rpoB* and *gyrB* (*SI Appendix*, Figs. S8 *B and C* and S9). Three AMR gene families were detected exclusively in the present-day calculus sample (aminoglycoside phosphotransferase APH6, CfxA beta-lactamase, and Erm 23S ribosomal RNA methyltransferase), one was more abundant in the present-day and historic samples (tetracycline resistance ribosomal protection protein) and one was more abundant in the prehistoric and historic samples (vancomycin-resistant beta prime subunit of RNA polymerase, *rpoC* gene) (Fig. 4*B* and Dataset S13, $P < 0.05$ pairwise Wilcoxon rank sum test). This pattern of differential abundances was confirmed in the three different normalization datasets (*SI Appendix*, Fig. S9).

Phylogenetic Analysis of Anaerolineaceae bacterium oral taxon 439. The microbial taxonomic analysis that we conducted showed that the species Anaerolineaceae bacterium oral taxon 439, which belongs to the Chloroflexi phylum, was abundant in the oral microbiome of the ancient foragers and farmers from the Danube Gorges and Italy, for more than 10 thousand years, and, though in lower proportions, was also present in more recent times, from the Bronze Age to the eighteenth to nineteenth century and present-day humans (*SI Appendix*, Fig. S10). It was demonstrated that long-term host–microbe association may result in shared phylogeographic patterns, potentially informing on human migrations and interactions (9, 33, 34). To investigate the molecular diversification of Anaerolineaceae bacterium oral taxon 439 across time and space and detect potential phylogeographic signatures possessed by this oral commensal, we reconstructed full bacterial genomes over a broad temporal transect.

Through shotgun sequencing, we generated 3-fold (CT2) to 39-fold (GR14) full genomes in 34 samples investigated in this study (Dataset S14). By analyzing ancient and modern dental calculus raw sequencing data from the literature, we identified another 16 genomes at >fourfold coverage isolated from modern and ancient individuals, including a Neanderthal from El Sidron (Spain), as well as the reference genome (CP017039). We built a maximum-likelihood tree of 51 full genomes based on 12,306 threefold coverage single-nucleotide polymorphisms (SNPs) that were called in at least 95% of the samples. For the final tree, we kept only the samples of this study for which well ascertained chronology was available (Dataset S1). The tree was rooted to the bacterial genome isolated from El Sidron Neanderthal (ca. 49,000 calibrated [cal] y B.P.), and the topology revealed a strong chronological and geographic clustering in particular for the prehistoric samples (Fig. 5).

We identified four main clusters (bootstrap value >95%). All the ancient bacterial genomes isolated from ancient foragers clustered in distinct geographical and chronological groups (i-iii), represented by samples from Paleolithic and Mesolithic Italy (i), Early Mesolithic (ii), and Late Mesolithic–Mesolithic/Neolithic transition (iii) Danube Gorges. Two samples from Late Paleolithic and Early Mesolithic Italy (CT2, CT5) and the Epipaleolithic sample from the site of Climente II in Romania (CL2) clustered in more basal positions. All the Neolithic samples from the Danube Gorges and the Finalese sites belonged to a completely distinct branch (iv), which also included all the post-Neolithic individuals that we investigated, from the Chalcolithic (Spain), the Bronze
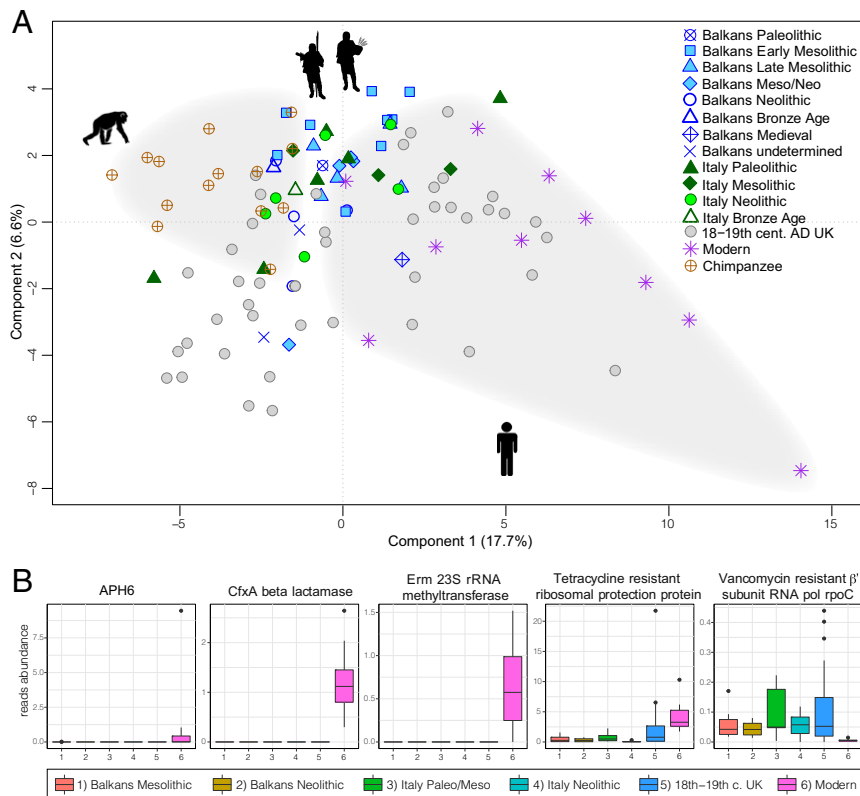
Fig. 4. PCA of microbial SEED categories in the oral microbiomes from this study and the literature (*A*), and boxplots showing the abundance of reads matching five AMR gene families (normalized for the abundance of reads matching the housekeeping gene *recA*) in the oral microbiomes of the prehistoric samples analyzed in this study and the eighteenth to nineteenth century and present-day samples from the literature (*B*).

Age (Padina in the Danube Gorges and Grotta Continenza in Italy), Middle Ages (Lepenski Vir), and more recent historical times until present-day. The same topology was observed in a maximum-likelihood tree built at more stringent conditions, using 2,098 fivefold coverage SNPs called in at least 95% of the samples (*SI Appendix*, Fig. S11), thus proving the consistency of the clusters detected. Furthermore, a recent study of ancient Anaerolineaceae bacterium oral taxon 439 genomes in Japan showed a similar pattern of differentiation in distinct clades of lineages originating from hunter-gatherers and agriculturalists, thus testifying to the validity of our phylogenetic reconstruction (35).

The analysis with TempEst (36) showed that the phylogenetic distances represented in the tree were poorly correlated with the sampling time ($R^2 = 0.05$). This is most likely due to homoplasies and recombination, as indicated by pairwise homoplasy index (PHI) test ($P < 0.01$) (37) and the analysis with ClonalFrame (relative effect of recombination and mutation r/m = 1.2, *SI Appendix*, Fig. S12) (38).

Biomolecular Evidence of Plants in Dental Calculus. Identification and validation of dietary sources from shotgun sequencing data of ancient dental calculus is challenging, and a number of validation criteria are necessary to authenticate dietary DNA signals (39). Given the fairly high degree of preservation and endogenous microbial oral content observed in our ancient samples, we tested whether DNA originating from plants (Viridiplantae) and animals (Metazoa) informing on food consumption or potential craft activities could be identified and authenticated.

Taxonomic classification and inspection of Eukaryote taxa was done with Kraken2 and two independent databases (the National Center for Biotechnology Information (NCBI) nucleotide (NT)

database and the custom database used for microbial analysis containing also organelles DNA). Following multiple means of verification and stringent authentication criteria (see *Materials and Methods*), no animal DNA could be identified. The plant molecular record appeared to be better preserved as we could successfully retrieve DNA molecules in two individuals from Vlasac, VL45 and VL48. We detected up to 340 reads mapping to the genome of the common hazel (*Corylus avellana*), 171 reads mapping to the genome of birch (*Betula pendula*) in VL45, and up to 74 reads mapping to the chloroplast genome of elderberry (*Sambucus nigra*) in VL48 (Dataset S15). All the animal and plant taxa that could not be authenticated and were filtered out in the analysis are reported in Dataset S16.

Discussion

The metagenomic dataset of ancient human dental calculus from prehistoric foragers and farmers that we documented fills a major chronological gap of knowledge concerning the evolution of human oral microbiomes. The microbial taxonomic analysis that we conducted demonstrated that the transition from foraging to farming subsistence strategies was not associated with significant changes in the human oral flora in the Balkans and Italy, at least when adopting a significance threshold of 0.01. In particular, in the Danube Gorges, this is possibly due to the fact that the dietary transition to agriculture was very gradual. Isotopic evidence (40) along with residue preserved on the earliest pottery sherds revealed a continuous reliance on the freshwater resources during the Neolithization period (41). Furthermore, multidisciplinary evidence, from material culture to mortuary practices and ancient genomes (13, 42), testifies to a scenario of deep biological mingling and cultural exchange between foragers and farmers across the Mesolithic/Neolithic transition in the
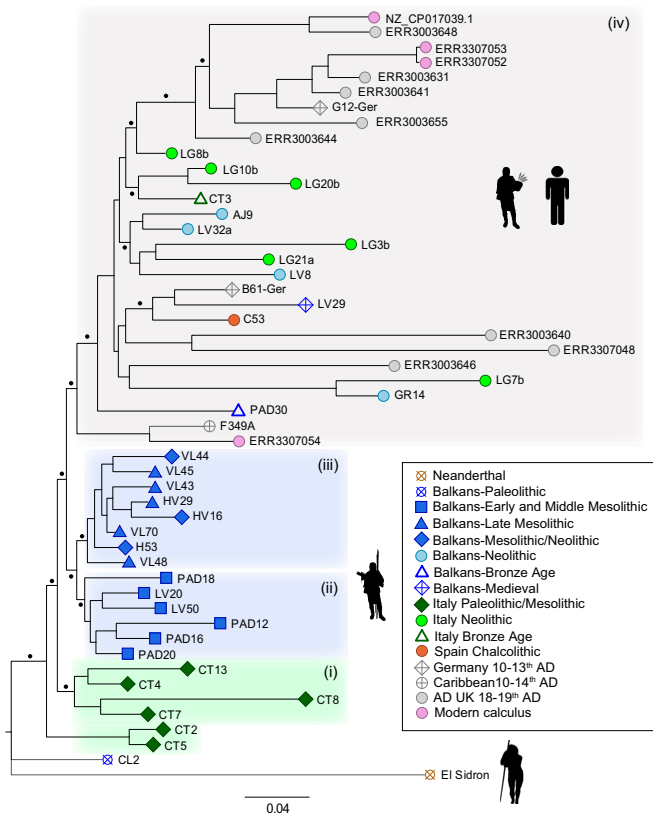
Ottoni et al.
Tracking the transition to agriculture in Southern Europe through ancient DNA analysis of dental calculus

PNAS | 5 of 11
https://doi.org/10.1073/pnas.2102116118

ANTHROPOLOGY

www.manaraa.com

Fig. 5. Maximum-likelihood tree of 12,306 threefold coverage SNPs of the oral commensal species Anaerolineaceae bacterium oral taxon 439 from human dental calculus of 47 ancient and 4 present-day individuals. A dot indicates the nodes with bootstrap value >95%.

Danube Gorges, which may have resulted in limited and gradual shifts in the oral environments. However, due to the high residual variation observed when testing differences with the Neolithic samples (>0.8, Dataset S7), we acknowledge that larger samples sizes in the future may help to obtain more robust tests and assess the differences between oral microbiome composition of prehistoric foragers and farmers from the regions investigated with higher significance.

The taxonomic analysis that we conducted with DESeq2 revealed that the Neolithic farmers most likely contributed with higher abundances of *Olsenella* sp. oral taxon 807 in both the Danube Gorges and Italy. Furthermore, the influx of farmers from the Near East may have determined an overall homogenization of oral microbiomes in the Danube Gorges and Italy, as witnessed by the PERMANOVA test and the lack of differential species in the Neolithic farmers from these two regions in DESeq2 (Datasets S7 and S8). Several factors, including the peculiar sociocultural contexts of the Neolithization process potentially involving different or additional source populations at the regional level (16), may have influenced and changed the human oral microbiomes to different degrees of significance. Due to the small size of the Neolithic samples analyzed here and the limited geographic coverage of this study restricted to the Balkans and Italy, we believe that future investigations conducted at the regional scale on larger samples covering wider geographic areas will help to examine potential changes in the oral flora associated with the Neolithization process in more depth.

More significant differences in oral microbiome composition were found when comparing the prehistoric samples from this study, either foragers and farmers from the Danube Gorges and

Italy, with the eighteenth to nineteenth century and modern human calculus samples. In particular, we observed a decrease in frequency from prehistory until the present day of the Chloroflexi species Anaerolineaceae bacterium oral taxon 439, which appeared to be a distinctive feature of the ancient foragers and farmers from the Danube Gorges and Italy. The tree of the genomes of this oral commensal that we reconstructed over a chronological span of more than 10,000 y (Fig. 5) showed a peculiar geographic and chronological clustering of the foragers and the Neolithic farmers from Italy and the Balkans. Farming in the Danube Gorges is consolidated starting from the Early Neolithic (ca. 5,950 to 5,500 BC) (40). To date, the site of Lepenski Vir provides the most compelling evidence of continuous occupation from the Mesolithic to the Early Neolithic, though significant cultural changes were already under way at the Mesolithic/Neolithic transition phase at this as well as other sites, such as Vlasac. Our results demonstrate that the oral commensal bacterium of two Neolithic individuals from Lepenski Vir radiocarbon dated to ca. 6,000 to 5,700 cal BC (LV8, LV32a, cluster iv), was genetically distinct from that of an earlier individual dated to the Mesolithic at the same site (LV20, cluster ii) but more closely related to a later Medieval sample (LV29, cluster iv). Similarly, the Neolithic farmers from the Finalese sites in Italy possessed bacterial genomes that were more closely related to a Bronze Age sample from Grotta Continenza (CT3) than to the Paleolithic and Mesolithic foragers from the same site.

Genomic evidence demonstrated that Neolithic farmers from Lepenski Vir possessed Near Eastern ancestries that contributed to the genomic makeup of modern European populations (43). This suggests that cluster iv was carried by Neolithic farmers from the Near East and spread throughout Europe during the Neolithization process via the Danubian and Mediterranean routes, as witnessed by the farmers from the Finalese sites in Liguria. Interestingly, despite the evidence of cultural changes at the Mesolithic/Neolithic transition at Vlasac, all the individuals dated to the Mesolithic/Neolithic transition at Vlasac possessed Anaerolineaceae bacterium oral taxon 439 of cluster iii, suggesting the lack of biological contributions from Neolithic farmers (e.g., through vertical or horizontal transmission of oral commensals), as also witnessed by ancient genomic evidence (43). Finally, we did not find any Anaerolineaceae bacterium oral taxon 439 bacterial genomes dated from the Neolithic until present day in the clusters of the Mesolithic foragers, thus suggesting that the Neolithic oral commensal replaced the Mesolithic variant and was transferred across generations until present day, possibly due to selective or demographic processes. Interestingly, a similar pattern of phylogeographic diversification was observed in Anaerolineaceae bacterium oral taxon 439 originating from dental calculus of ancient hunter-gatherers and agriculturalists from Japan (35), thus proving that phylogenetic investigation of this ancient oral commensal represents an informative tool to reconstruct human migrations in the past. Further analysis of ancient and modern Anaerolineaceae bacterium oral taxon 439 genomes will help to address more specifically the lineage turnover that we observed since the Neolithic.

Oral Chloroflexi are anaerobic heterotrophs, which were observed to proliferate in subgingival pockets of periodontal disease (44). They have a fermentative metabolism based on carbohydrates and proteinaceous carbon sources, possibly scavenging lysed material from elevated bacterial turnover and tissue breakdown (45). Interestingly, the Chloroflexi species Anaerolineaceae bacterium oral taxon 439 was particularly abundant in the Neolithic farmers from Liguria (SI Appendix, Fig. S8), who showed high incidence of periodontitis and other dentoalveolar pathological conditions (46, 47) (SI Appendix, Supplementary Information Text). Paleopathological analysis also revealed the poor health conditions and developmental disturbances of the farmers from the Finalese area, with evidence of lesions suggestive of osteoarticular tuberculosis in

several individuals (46, 48), including those studied here (*SI Appendix, Supplementary Information Text*). This may suggest a possible association of Anaerolineaceae bacterium oral taxon 439 with different degrees of nutritional and environmental stress of prehistoric lifestyles and conditions of dysbiosis that led to proliferation of species previously described as potential oral pathogens (e.g., *D. orale*, *Olsenella* sp. oral taxon 807) (49, 50). Nevertheless, Anaerolineaceae bacterium oral taxon 439 has only recently been described in humans, and its ecology and physiology are largely unknown (51). For this reason, the evolutionary factors leading to its reduction in frequency in modern humans remain elusive, and more data in the future will help to address this issue.

The functional analysis that we conducted confirmed the lack of significant differences between foragers and farmers in both the Danube Gorges (adonis test, $P = 0.174$, Dataset S10) and Italy ($P = 0.07$). However, a significant shift in functional activity was found when contrasting the dental calculus microbiomes of the prehistoric samples with those from present-day humans ($P < 0.01$). This shift was in particular due to the enrichment in pathways associated with bacterial pathogenesis, virulence, and environmental and physiological stress that were also detected in wild chimpanzees (trans−envelope signaling system VreARI, Sigma54−dependent transcription related gene cluster and protein secretion system type I, Dataset S10 and *SI Appendix*, Fig. S7), which may be indicative of a higher exposure to oral infections associated with prehistoric foraging and farming lifestyle.

Differently, present-day samples appeared to be enriched in pathways associated with lipoprotein transporters to outer membranes, which represent a permeability barrier against antibiotics. We believe that this pattern of differential metabolic pathways may be the result of changes in the functional activity of the human oral microbiome since the mass production of antibiotics in the 1940s (10, 52). This was confirmed by our survey of AMR genes, which showed that the oral microbiome of present-day calculus samples possessed antibiotic resistance gene families that were totally absent or less abundant in the prehistoric and historic (eighteenth to nineteenth century) samples (Fig. 4B). These gene families encode aminoglycoside modifying enzymes (APH6) targeting streptomycin, beta-lactamases targeting beta-lactam antibiotics such as penicillin, and proteins conferring resistance to macrolides, lincosamides and streptogramin b (Erm, which are part of the RNA methyltransferase family), and tetracycline. Conversely, the higher proportion of reads matching the *rpoC* family (beta prime subunit RNA polymerase) in the ancient samples may suggest that the mechanism conferring resistance to vancomycin was naturally acquired by human oral microbiomes from the environment (10) already in human prehistory. The lower abundance of reads matching the *rpoC* family in the present-day calculus sample suggests that the adoption of industrial antibiotics may have partially suppressed this "natural" mechanism of resistance, at least in the present-day sample investigated here. We are confident that this and other dynamics behind the acquisition and loss of antibiotic resistance mechanisms may be better addressed in the future by generating more sequencing data from ancient and modern dental calculus.

Interestingly, we also detected in the ancient forgers from the Danube Gorges and Italy a higher functional activity associated with the breakdown of dietary glycans, including plant storage and cell wall polysaccharides (Dataset S11 and *SI Appendix*, Fig. S7). This pathway was also detected in the chimpanzees, which primarily adopt plant-based feeding strategies (53). Recent microbotanical evidence proved that plants represented a substantial fraction of the diet of Mesolithic foragers from the Danube Gorges and Grotta Continenza (54, 55), and this may be reflected by a higher representation in their oral microbiomes of functional activities associated with the breakdown of dietary glycans.

Our metagenomic analysis detected DNA molecules matching numerous plant and animal taxa; however, they could not be

authenticated based on the stringent criteria applied. In fact, spurious hits resulting in false positive signals, as well as limitations and assumptions associated with the use of modern reference sequences, make dietary analysis of shotgun dataset problematic (39). In particular, the presence of contaminating plant taxa in ancient biological samples, especially Triticea, is often debated (56) and represents a major hurdle for tracking food biomolecular records in dental calculus. Nonetheless, we were able to authenticate the presence of DNA from the common hazel (*C. avellana*), elderberry (*S. nigra*), and birch (*B. pendula*). These three taxa are well documented in the Danube Gorges and in macrofossil records of the region as well as throughout the European Mesolithic (57, 58). Hazelnut and elderberry are edible species, shell and seed remains of which were found at Vlasac (42, 59), hence they are not "unexpected" based on the local archaeological context (39). Their presence in dental calculus suggests their possible use as food resources, providing vitamins and minerals that complemented fish and meat nutrients. Birch is also well documented in the local plant macrofossil record (60). Birch pitch was used since the Middle Pleistocene (61) as an adhesive and hafting agent that was chewed before use (62), thus possibly leaving a molecular record in the dental calculus. The use of the mouth as third hand in extra-masticatory activities involving wood working cannot be excluded as another pathway of inclusion of wood plant particles in dental calculus (63).

The multiple validation criteria that we applied, corroborated by the macrofossil evidence in the archaeological contexts, indicate that the biomolecular record that we detected is authentic. However, we acknowledge that dietary analysis of shotgun sequencing data of dental calculus remains extremely challenging. Systematic biomolecular analysis of plant and animal taxa may be prone to several flaws, as recently demonstrated (39), and in the lack of further biomolecular, macrofossil, or microfossil evidence, authentication of taxa identified can be difficult to assess. Targeted-enrichment methods may represent a more appropriate means to systematically analyze dietary biomolecular information in ancient dental calculus.

Overall, the evidence that we gathered suggests that only minor changes in the human oral microbiome occurred with the onset of the Neolithic. A more significant shift in the taxonomic composition and in the functional profile of the human oral flora occurred later, as the microbiomes reconstructed from dental calculus of prehistoric foragers and farmers were significantly different from those of modern populations. In particular, prehistoric oral microbiomes were characterized by a higher frequency of a Chloroflexi species, Anaerolineaceae bacterium oral taxon 439, an oral commensal which recorded in its genome the spread of the Neolithic through the Balkans and Italy starting from the mid-seventh millennium BC. Furthermore, modern oral microbiota functionally evolved to develop pathways associated with peculiar mechanisms of resistance to antibiotics, most likely in response to their massive adoption in modern society since the past century.

When the shift in the taxonomic composition of oral microbiomes occurred and whether it was related to peculiar changes in diet and lifestyle during human evolution remains to be understood. The oral cavity is a complex environment influenced by multiple factors during an individual's lifespan—from health to diet, host genome, and interaction with the environment. In the future, filling the chronological and geographical gap with post-Neolithic samples and a larger calculus dataset from present-day populations may help in addressing the question of when and why the human oral microbiome changed.

Materials and Methods

aDNA Laboratory Methods. Supragingival dental calculus for aDNA analysis was available for 44 individuals (Dataset S1). Laboratory analyses were conducted in the dedicated aDNA facilities of the Department of Anthropology of the University of Vienna (Austria), following standard precautions

Ottoni et al.
Tracking the transition to agriculture in Southern Europe through ancient DNA analysis of dental calculus

PNAS | 7 of 11
https://doi.org/10.1073/pnas.2102116118

www.manaraa.com

for access to the facilities and decontamination (64). Prior to DNA extraction, 2 to 24 mg dental calculus was washed in 0.5 mL EDTA 0.5 M pH 8 for 15 min at room temperature in a thermoshaker to remove surface contamination. Dental calculus samples were decanted through centrifugation, dried and further decontaminated through UV irradiation in a cross-linker, and then gently broken in pieces with a microspatula when needed. Following decontamination, samples were incubated in 1 mL EDTA 0.5 M pH 8 and 0.25 mg/mL Proteinase K at 55 °C for 12 h and at 37 °C for 36 h in a thermoshaker. After lysis, undissolved dental calculus was decanted through centrifugation, and DNA was extracted with Roche Assembly Tubes silica columns (65), with slight modifications. Four extraction blanks were processed in parallel with the archaeological samples. A total of 20 μL of each DNA extract was converted in double-stranded genomic libraries (66), 18 μL of which was split in three double-indexing amplification reactions of 25 cycles (50 μL volume with 2.5 U PfuTurbo Cx, 200 μL each dNTP, and 200 nM primers P5 and P7) that were finally pooled and purified with Minelute silica columns (Qiagen). To remove heteroduplexes due to overamplification, 2 μL of purified libraries were used for reconditioning PCR (one cycle) (67) in 25 μL AccuPrime Pfx Supermix and primers IS5 and IS6 (200 nM each) and purified with Minelute columns (Qiagen). Extraction, library, and amplification controls were processed in parallel with the samples. Concentrations and library profiles were checked using a DNA1000 Bioanalyzer chip (Agilent) and a Qubit double-stranded (ds) DNA high sensitivity (Thermofisher).

All libraries were pooled at equimolar concentrations (a 10-fold dilution of the extraction and library controls was added to the pool) and sequenced in five lanes of a HiSEq. 4000 (2 × 150 base pair [bp]) at the Norwegian Sequencing Centre core facility to obtain a final sequencing depth of about 30 million reads per sample (Dataset S2).

Metagenomic Analysis and Alignment to Reference Sequences. Trimming of adapter sequences, reads quality filtering, and paired reads merging was conducted with AdapterRemoval (68) (–minlength 30 –minquality 25 –trimns –trimqualities –collapse). Sequence duplicates were removed with PrinsEq (69). We used Kraken2 and a custom database updated to November 2020 of bacterial, viral, archaeal, and organelle genomes from the NCBI Reference Sequence (RefSeq) database (https://www.ncbi.nlm.nih.gov/refseq/) for taxonomic classification of merged deduplicated reads, as previously described (6). All the reference genomes were masked in Kraken2 for low-complexity regions with Dustmasker, to reduce the impact of spurious classifications. The same workflow was adopted for taxonomic classification of sequencing data of dental calculus, plaque, and other shotgun metagenomic datasets from the literature used for the comparative analyses (Dataset S5). To explore the oral microbiomes of nonhuman primates with typical foraging behavior, we also included sequencing data from the literature of dental calculus originating from skeletal remains of modern wild chimpanzees (<100 y B.P.) from the Gombe National Park (n = 19, Tanzania) (70) and Sierra Leone (n = 1) (9). The low numerosity of Neanderthal oral microbiomes investigated as of the date of this writing [n = 3, from different regions and with purported different dietary habits (9)] did not allow us to make direct robust tests against Neanderthal foragers. More extensive insights on the relationships between foraging behavior and oral microbiome variation may be gained in the future with the analysis of more Neanderthal dental calculus samples.

To assess the authenticity of most abundant microbial species detected by Kraken2, we aligned the reads against the respective reference sequences deposited in the NCBI RefSeq database (reported as "reference" or "representative" genomes in the RefSeq category) with the Burrows–Wheeler Alignment (bwa) tool and the aln algorithm at high stringency (–n 0.1). Postmortem damage in the form of cytosine deaminations was estimated with mapDamage (71). We investigated the edit distance distribution by calculating the parameter −Δ% (negative difference proportion) in each sample (Dataset S4) (72). To do that, we generated a bed file of edit distances to the reference of the mapped reads with bedtools bamtobed -tag NM, and we calculated the −Δ% in R. We regarded taxonomic assignations with a −Δ% > 0.8 as authentic, to account for the possibility of cross alignments due to horizontal gene transfer and the presence of closely related microbial species in the sample, as recently proposed (62). For some species, edit distances >0.8 were obtained after filtering the mapped reads with PMDtools for postmortem damage (PMD) (–threshold = 1) and mapping quality >30 (–requiremapq = 30), suggesting that even with stringent alignment criteria (bwa –n 0.1), nonspecific reads from closely related contaminant taxa may be aligned. Average fragment length of reads mapping to microbial species was calculated with samtools stats. Cytosine deamination rates of 5′ terminal position, and average fragment lengths of the ancient samples analyzed in this study were visualized in boxplots with ggplot in R (*SI Appendix*, Fig. S2).

To screen the presence of human host DNA in the archaeological dental calculus, we aligned the raw reads against the human genome reference sequence (GRCh38) with bwa (–n 0.01 –o 2), and the rate of cytosine deaminations was estimated with mapDamage.

Recently, genomic GC content of microbial species was suggested to be a potential source of taxonomic skew in the analysis of archaeological microbiomes due to a systematic loss of AT-rich short fragments, which is expected to be more significant in low GC-content microbial genomes (22). To screen the presence of biases associated with differential taxonomic preservation in our dataset, we calculated the GC content to fragment length in four species that were detected as highly abundant and discriminant for the chronological groups investigated in the taxonomic analysis that we conducted (as resulting from the PCA, DeSEQ2, STAMP), *Anaerolineaceae* bacterium oral taxon 439, *M. oralis*, *Olsenella* sp. oral taxon 807, and *R. aeria*. To do that, we first extracted and converted to fasta the reads assigned to each microbial species by Kraken2 with extract_kraken_reads.py in KrakenTools (https://github.com/jenniferlu717/KrakenTools). To compute summary statistics, including the GC-content, for each read we used Infoseq, by the Emboss Team. Finally, we used the package lattice in R to plot the GC content against the fragment length with bwplot.

Taxonomic Data Analysis. The Kraken output was used to estimate read abundances for each species with Bracken (73). Since we used a database of complete bacterial and archaeal NCBI RefSeq genomes, we normalized the species read abundances of all the datasets, from this study and from the literature, for genome length by dividing them for the size (in Gb) reported in the NCBI. We generated a table of genome lengths from the NCBI browser (https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/) for the taxa represented in our RefSeq custom database. This table was then used for normalization with a custom python script (https://github.com/claottoni). The genome-length normalized read counts of the entire dataset were also normalized for library size through total sum scaling in R. Normalized abundance tables of Bacteria and Archaea were filtered to include only taxa with frequency >0.02% in the full dataset of samples from this study and the literature (Dataset S6). To retrieve full taxonomic data of species abundances (from species up to phylum), we used taxonomy-ranks (https://pypi.org/project/taxonomy-ranks/).

We estimated the proportion of reads in archaeological dental calculus stemming from other microbiomes with Sourcetracker and the normalized bacterial and archaeal species reads abundance before filtration of low-abundance taxa. Modern human dental calculus and plaque, skin, and soil from the European Nucleotide Archive (ENA) (https://www.ebi.ac.uk/ena) and the laboratory controls generated in this study were used as sources, whereas all the ancient dental calculus samples were used as sink (Fig. 2 and Dataset S3). A nonmetric multidimensional scaling of Bray–Curtis distances from species abundances of the various microbiomes was done with vegan in R (*SI Appendix*, Fig. S1). Given the high fraction of oral microbial taxa (>90%), no filtration of laboratory controls, soil, or skin taxa was conducted on the ancient dental calculus samples from this study.

A PCA of species abundances from oral microbiomes (dental calculus and plaque) from this study and the literature (Fig. 3 and Dataset S6) was conducted with the mixomics library in R. To account for the compositional nature of microbiome data (74), we applied a centered-log ratio (CLR) transformation after converting the normalized reads in counts per million and using a +1 offset in the abundance tables. We identified chronological samples based on numerosity (≥3 individuals), namely the Mesolithic and Neolithic sample from the Danube Gorges, Paleolithic/Mesolithic from Grotta Continenza, Neolithic from the Finalese sites in Liguria, the historic sample (eighteenth to nineteenth century AD) from Radcliffe (UK), modern calculus from Spain, (7) and modern plaque from the Human Microbiome Project (HMP). To test differences among the chronological groups identified, we made PERMANOVA tests with adonis using euclidean distance matrices of the CLR-transformed species abundances, 999 permutations (package vegan in R), and P < 0.01 considered significant. The differences between groups were further tested by adopting presence/absence distance matrices with Jaccard distances (Dataset S7). The homogeneity of variances in the groups compared in the PERMANOVA was tested with betadisper and the anova function in R, resulting in all instances in nonsignificant differences except for the Mesolithic and the Radcliffe samples. Adonis was also used to test groups of calculus samples based on oral geography. For this analysis, with particular regards to the eighteenth to nineteenth century historic sample from Radcliffe (7), we took into account only the calculus samples originating from a single tooth.

To detect the most differentially abundant species among chronological groups, we used DESeq2 and generated heatmaps with the package pheatmap in R (Dataset S8 and *SI Appendix*, Fig. S6). Differential abundances were also explored with STAMP (75), where frequency tables were analyzed

www.manaraa.com

with White's nonparametric two-sided *t* test, bootstrapping to determine the difference between proportions and 95% cutoff, and Storey's false discovery rate (FDR), as previously adopted (7), with significance level set as corrected *P* values (q values) ≤ 0.05 and effect size ≥ 1 (*SI Appendix*, Fig. S5).

Functional Analysis and AMR Analysis. To investigate SEED functional profiles, we classified the quality filtered and collapsed reads from the dental calculus of the ancient samples from this study, the eighteenth to nineteenth century, and the modern samples from the literature (7) against the NCBI-NR database with Diamond BlastX (76). SEED categorization was obtained in Megan with daa-meganizer and the mapping file megan-map-Jul2020-2-X.db. SEED profile tables were exported at the second level of categorization from Megan (Dataset S9). To conduct the PCA, we applied the CLR transformation and a minimal offset +0.1. To be consistent with the taxonomic analysis, SEED categories with relative abundance <0.02% in the full dataset of samples from this study and the literature were removed, resulting in 81 final categories. We tested differences in the SEED profiles of the chronological samples with PERMANOVA tests (adonis in R package vegan) (Dataset S10), as described in the *Taxonomic Data Analysis*, and identified differential SEED categories with DESeq2 (Dataset S11). Further identification of discriminant SEED categories was done with the sPLSDA in the Mesolithic foragers from the Danube Groges and Italy (*SI Appendix*, Fig. S7). The analysis was done in R with the package mixomics, and it was not done on the Neolithic samples due their low sample size (*n* < 6). To do that, we followed the steps described in a recent study for finding the optimal number of dimensions and tuning the PCA (7).

To conduct the AMR investigation, we first used extract_kraken_reads.py from KrakenTools to extract and convert to fasta the reads of the ancient and modern calculus samples from this study and the literature (7, 9, 22, 35) classified as bacteria in Kraken2 (using the options –t 2 –include-children). We aligned the extracted reads against the CARD database with blastn (32). Then, we adopted a strategy based on the top bitscore match to retrieve the ARO accession number and the AMR gene family (10). To account for potential taphonomic conditions leading to preservation biases across samples of different chronology, we used blastn to align the reads against the nucleotide sequences of three microbial housekeeping genes (*recA, rpoB, gyrB*) deposited in the NCBI. We downloaded the sequences with with esearch and efetch from EDirect package of NCBI. For each calculus sample, the number of reads matching the AMR gene families was normalized for the number of reads matching the three housekeeping genes and converted to counts per thousand (times 1,000). We investigated only the most abundant gene families by removing those represented below 0.02% in the whole dataset. This filtration resulted in 43 gene families that were further analyzed to generate boxplots in the chronological samples with ggplot in R (*SI Appendix*, Fig. S8). Pairwise Wilcoxon rank sum tests were done in R with dplyr and the Benjamini–Hochberg *P*-value adjustment method (Dataset S13). Four gene families were significantly different between all the ancient samples and the present-day sample (*P* < 0.05) and for this reason were taken into further consideration and discussed (Fig. 4*B* and *SI Appendix*, Fig. S9). The differential abundance of these AMR gene families was also observed in other prehistoric and historic samples from the literature (7, 9, 22, 35), thus suggesting that methodological factors associated with laboratory procedures were most likely not responsible for the differences observed. The *rpoC* gene family was also discussed given its higher abundance in the ancient samples, except for those from Weyrich's study (9) (Datasets S12 and S13 and *SI Appendix*, Fig. S8).

Plants Biomolecular Analysis and Authentication. Shotgun metagenomic datasets have recently been used for ancient dietary analyses but are often problematic due to spurious hits causing false positive signals and as a result of limitations and assumptions associated with the use of modern reference sequences that may not reflect ancient taxa.

To investigate the biomolecular record of potential Eukaryote dietary sources, we applied a number of validation criteria, as recently suggested (39). First, we made the taxonomic classification with Kraken2 and two independent databases, the NCBI-NT database and the custom microbial database that also included metazoa, plant, and fungal mitochondrial and plastid genomes. We visualized the output reports with Pavian (77) and generated Eukaryote species abundance tables. Second, we removed the species present in the laboratory controls (extraction and library blanks) and in the calculus sample with low oral microbial content, VK1. This sample, despite originating from a different archaeological site and area (Croatia), was used as control for environmental contaminants due to its high soil microbial content (Fig. 2 and Dataset S3) and for potential contaminants introduced in the collection and storage of the sample for DNA analyses (conducted by the same person for all

the samples, E.C.), as recently suggested (39). After filtration, we focused on the identified taxa accounting for >0.5% of all reads classified in the animal (Metazoa) and plant (Viridiplantae) kingdoms. Finally, validation of species classifications was further assessed by aligning the reads to the reference genomes of the identified candidate species (when full genome assemblies were not available, organelle reference sequences were used, as in *S. nigra*). The number of mapped reads was estimated after filtering for postmortem damage (–threshold = 1) and mapping quality >30 (–requiremapq = 30) with PMDtools. We regarded as authentic only the species showing postmortem damage and a declining edit distance distribution, as recently described (62, 72). In species for which genome assemblies were available, we also considered as further means of authentication the average read coverage per contig (*SI Appendix*, Fig. S13). Reads mapping to contigs showing abnormally high coverage values (up to 100 times higher than other contigs) were removed from the final read count (e.g., contig GCA_901000735 in *C. avellana*, and GCA_900184695 in *B. pendula*), as most likely resulting from spurious alignments due to conserved genomic regions (39).

No animal species identified by Kraken2 passed the authentication criteria that we established. Given the high consumption of aquatic resources in the Mesolithic human communities of the Danube Gorges, as suggested by stable isotope evidence (40, 41), we tested in more depth the potential presence of fish DNA in the dental calculus of the individuals analyzed in this study by making a competitive alignment of the reads against all the organelle sequences of Acipenseridae and Cyprinidae species, which are among the most abundant in the Danube River. A low number of mapped reads (<10) was found in most samples, with the exception of PAD9, in which 26 reads aligned to 25 Cyprinidae species organelles DNA. Given the low amount reads, no postmortem damage assessment could be done. Hence, in the absence of validation criteria, we disregarded the results.

Phylogenetic Analysis. Sequencing data of dental calculus samples from this study and the literature were aligned to the reference genome of the oral bacterial species Anaerolineaceae bacterium oral taxon 439 (NZ_CP017039), which was found to be the most abundant and recurrent microbial species in the prehistoric foragers and farmers from the Danube Gorges and Italy. Postmortem damage and −Δ% (negative difference proportion) were estimated to authenticate the alignments, as described in *Results*. The individuals showing >70% of threefold genome coverage were selected for downstream phylogenetic analysis (Dataset S14). In total, 51 genomes were used for phylogenetic tree reconstruction. Of the ancient Danube Gorges samples, two individuals from Schela Cledovei (SC2, SC13) for which chronology could not be fully ascertained were removed.

After rescaling quality scores of nucleotide variants in the reads originating from postmortem deamination processes with mapDamage, variant calling was performed with samtools and bcftools. To generate the alignment of nucleotide variants across all the samples investigated, we used snpToolkit (78) (https://github.com/Amine-Namouchi/snpToolkit/), which allows to filter and annotate SNPs from vcf files. The SNPs were filtered according to 1) quality score ≥30 (-q 30), 2) depth of coverage ≥3 (-d 3), 3) variant allele reads ratio ≥90% (-r 0.9), and 4) 3-bp distant contiguous SNPs (-f 3). Furthermore, we screened the Anaerolineaceae bacterium oral taxon 439 reference genome for repetitive regions with RepeatScout (79) and Tandem Repeat Finder (80) and annotated them with RepeatMasker (http://www.repeatmasker.org). The repetitive regions identified were excluded from the snpToolkit annotation (-e option), together with the regions corresponding to ribosomal RNA and transposases, as reported in the gbff file of the reference sequence. The annotation files generated for all the samples investigated were used to create an alignment of concatenated SNPs with snpToolkit combine, using the bam option to inspect the bam files and account for the presence of poorly covered positions (below the threshold of three reads, registered as missing data in the alignment file).

We used snpToolkit to calculate also the number of variants present at ratios (parameter -r) of aligned reads between 0.4 and 0.6 and above 0.8 (we did not account for ratios lower than 0.4, which could more likely stem from artifact variation, e.g., due to cytosine deamination or sequencing errors). They were used to estimate the fraction of chimeric variants possibly due to the presence of mixed Anaerolineaceae bacterium oral taxon 439 strains in the individual's oral microbiome (Dataset S14). We found that the ratio of chimeric variants was high (>0.5) in eight individuals. Importantly, we note that when more than one strain is present; the strategy of variants filtration adopted for building the tree would account only for the genetic variants shared by the strains, resulting in a more ancestral phylogenetic position of the individual in the tree, as described previously (81). Hence, we find it unlikely that potential mixed strains affected the overall topology of the

Ottoni et al.
Tracking the transition to agriculture in Southern Europe through ancient DNA analysis of dental calculus

PNAS | 9 of 11
https://doi.org/10.1073/pnas.2102116118

www.manaraa.com

tree and that the individuals carrying chimeric sequences could belong to a clade different from the one assigned.

The SNP alignment was used to reconstruct a maximum-likelihood phylogenetic tree with IQ-TREE (82), using ModelFinder (83) (–m MFP, ModelFinder Plus as default behavior in IQ-TREE) to choose the appropriate substitution model. Branch supports were assessed by performing 1,000 replicates of ultrafast bootstrap approximation (84) and Shimodaira–Hasegawa (SH)-like approximate likelihood ratio test (85). For phylogenetic inference, we used a reduced dataset of threefold coverage SNPs that were called in at least 95% of the samples (12,306 SNPs), which was generated with trimAl (86). We used the bacterial genome isolated from the Neanderthal of El Sidron (Spain), which was the oldest sample available (49,000 y), to root the tree. The output was visualized with FigTree (Fig. 5) (http://tree.bio.ed.ac.uk/software/figtree/).

To corroborate the topology of the tree, we built a second maximum-likelihood tree using only the SNPs with fivefold coverage called in 95% of the samples. Despite the more stringent conditions and the lower number of SNPs (2,098), the main clusters of the tree were reproduced, except for the samples F349 and ERR3307054 that fell in more ancestral positions due to the lower resolution achieved (*SI Appendix*, Fig. S11). F349 originates from a different geographic context (Caribbean), and more samples from that region (and more widely the Americas) may help in the future to increase the phylogenetic resolution of this lineage.

We evaluated the temporal signal of the tree with TempEst (36), using the chronology information available for each individual as sampling time. The analysis showed a low temporal signal (root-to-tip genetic distance against sampling time $R^2 = 0.06$); hence, we tested to what extent recombination occurred across the reconstructed phylogeny by conducting a PHI test with PhiPack (37) and a maximum-likelihood estimate of recombination parameters with ClonalFrameML (38). To do that, we generated an alignment of full genomes that included the SNPs used for the tree reconstruction. We first filtered the vcf file of each sample with vcftools and the list of SNPs used in the phylogenetic tree (options –positions and –recode of vcftools). Then, we used the vcf files to generate full fasta bacterial genomes for each individual with gatk FastaAlternateReferenceMaker and the reference genome (NZ_CP017039). By doing that, we obtained fasta sequences covering the full length of the genome, with no gaps. For this reason, to incorporate the gaps corresponding to the low coverage regions, for each individual we generated a list of missing sites (below threefold coverage) from the output of snpToolkit. Then, we used awk to incorporate the missing sites in the full genome fasta file by replacing the corresponding nucleotide with a question mark. The fasta files of each sample were merged in a single alignment file that was used for recombination analysis in ClonalFrameML. For the latter, we made a first simple run with default values to estimate the model parameters with the Baum–Welch Expectation-Maximization algorithm, R/θ (relative rate of recombination to mutation), δ (mean DNA import length), and ν (mean divergence of imported DNA). In the second run, we used the parameters estimated in the first run (-initial_values), we set the constraint

on the variability of recombination parameters (-embranch_dispersion) at 0.1, and estimated the model parameters in each branch (-embranch). A graphical output of the analysis (*SI Appendix*, Fig. S12) was generated with the R script cfml_results.R available in the ClonalFrameML repository (https://github.com/xavierdidelot/ClonalFrameML). We estimated the relative effect of recombination and mutation (87) as r/m = (R/θ) × δ × ν = 1.2, thus indicating a similar contribution of recombination and mutation to the genetic variation represented by the tree. The relative effect of recombination and mutation calculated for each branch (*SI Appendix*, Fig. S12) indicated that recombination events were particularly high in the external branches of the tree (r/m >1) and lower in the internal branches (r/m < 1). We also tested the presence of recombination and homoplasy through a PHI test with 1,000 permutations in PhiPack (37).

1. B. Koskella, L. J. Hall, C. J. E. Metcalf, The microbiome beyond the horizon of ecological and evolutionary theory. *Nat. Ecol. Evol.* **1**, 1606–1615 (2017).

2. R. J. Lamont, H. Koo, G. Hajishengallis, The oral microbiota: Dynamic communities and host interactions. *Nat. Rev. Microbiol.* **16**, 745–759 (2018).

3. M. R. Mason, H. N. Nagaraja, T. Camerlengo, V. Joshi, P. S. Kumar, Deep sequencing identifies ethnicity-specific bacterial signatures in the oral microbiome. *PLoS One* **8**, e77287 (2013).

4. C. Warinner, C. Speller, M. J. Collins, A new era in palaeomicrobiology: Prospects for ancient dental calculus as a long-term record of the human oral microbiome. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130376 (2015).

5. A. Radini et al., Medieval women's early involvement in manuscript production suggested by lapis lazuli identification in dental calculus. *Sci. Adv.* **5**, eaau7126 (2019).

6. C. Ottoni et al., Metagenomic analysis of dental calculus in ancient Egyptian baboons. *Sci. Rep.* **9**, 19637 (2019).

7. I. M. Velsko et al., Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage. *Microbiome* **7**, 102 (2019).

8. C. J. Adler et al., Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat. Genet.* **45**, 450–455, 455e1 (2013).

9. L. S. Weyrich et al., Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* **544**, 357–361 (2017).

10. J. C. Brealey et al., Dental calculus as a tool to study the evolution of the mammalian oral microbiome. *Mol. Biol. Evol.* **37**, 3003–3022. (2020).

11. C. Warinner et al., Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46**, 336–344 (2014).

12. D. Borić, T. D. Price, Strontium isotopes document greater human mobility at the start of the Balkan Neolithic. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 3298–3303 (2013).

13. M. Feldman et al., Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia. *Nat. Commun.* **10**, 1218 (2019).

14. D. Borić et al., High-resolution AMS dating of architecture, Boulder artworks and the transition to farming at Lepenski Vir. *Sci. Rep.* **8**, 14221 (2018).

15. G. Boschian, M. Serradimigni, M. Colombo, S. Ghislandi, R. Grifoni Cremonesi, Change fast or change slow? Late Glacial and Early Holocene cultures in a changing environment at Grotta Continenza, Central Italy. *Quat. Int.* **450**, 186–208 (2017).

16. M. L. Antonio et al., Ancient Rome: A genetic crossroads of Europe and the Mediterranean. *Science* **366**, 708–714 (2019).

17. D. Binder et al., Modelling the earliest north-western dispersal of Mediterranean Impressed Wares: New dates and Bayesian chronological model. *Documenta Praehistorica* **44**, 54–77 (2017).

18. V. S. Sparacello et al., Dating the funerary use of caves in Liguria (northwestern Italy) from the Neolithic to historic times: Results from a large-scale AMS campaign on human skeletal series. *Quat. Int.* **536**, 30–44 (2020).

19. S. Sawyer, J. Krause, K. Guschanski, V. Savolainen, S. Pääbo, Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131 (2012).

20. P. Korlević et al., Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).

21. A. T. Ozga et al., Successful enrichment and recovery of whole mitochondrial genomes from ancient human dental calculus. *Am. J. Phys. Anthropol.* **160**, 220–228 (2016).

22. A. E. Mann et al., Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. *Sci. Rep.* **8**, 9822 (2018).

23. J. L. Mark Welch, S. T. Ramírez-Puebla, G. G. Borisy, Oral microbiome geography: Micron-scale habitat and niche. *Cell Host Microbe* **28**, 160–168 (2020).

10 of 11 | PNAS
https://doi.org/10.1073/pnas.2102116118

Ottoni et al.
Tracking the transition to agriculture in Southern Europe through ancient DNA analysis

24. R. Overbeek et al., The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**, 5691–5702 (2005).

25. M. J. Kazmierczak, M. Wiedmann, K. J. Boor, Alternative sigma factors and their roles in bacterial virulence. *Microbiol. Mol. Biol. Rev.* **69**, 527–543 (2005).

26. M. A. Llamas et al., A novel extracytoplasmic function (ECF) sigma factor regulates virulence in Pseudomonas aeruginosa. *PLoS Pathog.* **5**, e1000572 (2009).

27. E. R. Green, J. Mecsas, Bacterial secretion systems: An overview. *Microbiol. Spectr.* **4**, 1–32 (2016).

28. J. A. Briggs, J. M. Grondin, H. Brumer, Communal living: Glycan utilization by the human gut microbiota. *Environ. Microbiol.* **23**, 15–35 (2020).

29. K. M. Lehman, M. Grabowicz, Countering gram-negative antibiotic resistance: Recent progress in disrupting the outer membrane with novel therapeutics. *Antibiotics (Basel)* **8**, 163 (2019).

30. M. Grabowicz, T. J. Silhavy, Redefining the essential trafficking pathway for outer membrane lipoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4769–4774 (2017).

31. M. O. A. Sommer, G. Dantas, G. M. Church, Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128–1131 (2009).

32. B. Jia et al., CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).

33. D. Falush et al., Traces of human migrations in Helicobacter pylori populations. *Science* **299**, 1582–1585 (2003).

34. Y. Moodley et al., The peopling of the Pacific from a bacterial perspective. *Science* **323**, 527–530 (2009).

35. R. Eisenhofer, H. Kanzawa-Kiriyama, K.-I. Shinoda, L. S. Weyrich, Investigating the demographic history of Japan using ancient oral microbiota. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190578 (2020).

36. A. Rambaut, T. T. Lam, L. Max Carvalho, O. G. Pybus, Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).

37. T. C. Bruen, H. Philippe, D. Bryant, A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665–2681 (2006).

38. X. Didelot, D. J. Wilson, ClonalFrameML: Efficient inference of recombination in whole bacterial genomes. *PLoS Comput. Biol.* **11**, e1004041 (2015).

39. A. E. Mann et al., Do I have something in my teeth? The trouble with genetic analyses of diet from archaeological dental calculus. *Quat. Int.*, 10.1016/j.quaint.2020.11.019 (2020).

40. R. Schulting, D. Borić, "A tale of two processes of Neolithisation" in *The Neolithic of Europe*, P. Bickle, V. Cummings, D. Hofmann, J. Pollard, Eds. (Oxbow Books, 2017), pp. 82–106.

41. L. J. E. Cramp et al., Regional diversity in subsistence among early farmers in Southeast Europe revealed by archaeological organic residues. *Proc. Biol. Sci.* **286**, 20182347 (2019).

42. D. Borić et al., Late Mesolithic lifeways and deathways at Vlasac (Serbia). *J. Field Archaeol.* **39**, 4–31 (2014).

43. I. Mathieson et al., The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018).

44. S. P. Szafranski et al., High-resolution taxonomic profiling of the subgingival microbiome for biomarker discovery and periodontitis diagnosis. *Appl. Environ. Microbiol.* **81**, 1047–1058 (2015).

45. A. G. Campbell et al., Diversity and genomic insights into the uncultured Chloroflexi from the human microbiota. *Environ. Microbiol.* **16**, 2635–2643 (2014).

46. V. Formicola, Q. Milanesi, C. Scarsini, Evidence of spinal tuberculosis at the beginning of the fourth millennium BC from Arene Candide cave (Liguria, Italy). *Am. J. Phys. Anthropol.* **72**, 1–6 (1987).

47. V. S. Sparacello, C. Panelli, S. Rossi, I. Dori, A. Varalli, "Archaeothanatology and palaeobiology of the burials and 'scattered human remains' from Arma dell'Aquila (Finale Ligure, Savona)" in *Gli Scavi All'Arma Dell'Aquila (Finale Ligure, Savona): Le Ricerche e i Materiali Degli Scavi Del Novecento*, P. Biagi, E. Starnini, Eds. (Società per la Preistoria e Protostoria della Regione Friuli-Venezia Giulia, 2018), vol. 15, pp. 143–181.

48. A. Canci, S. Minozzi, S. M. Borgognini Tarli, "Resti scheletrici umani" in *Il Neolitico Nella Caverna Delle Arene Candide (Scavi 1972-77)*, Collezioni di Monografie Preistoriche e Archeologiche X., S. Tiné, Ed. (Istituto Internazionale di Studi Liguri, 1999), pp. 304–312.

49. P. S. Langendijk, E. M. Kulik, H. Sandmeier, J. Meyer, J. S. van der Hoeven, Isolation of Desulfomicrobium orale sp. nov. and Desulfovibrio strain NY682, oral sulfate-reducing bacteria involved in human periodontal disease. *Int. J. Syst. Evol. Microbiol.* **51**, 1035–1044 (2001).

50. J. Obata et al., Identification of the microbiota in carious dentin lesions using 16S rRNA gene sequencing. *PLoS One* **9**, e103712 (2014).

51. S. J. McIlroy et al., Culture-independent analyses reveal novel anaerolineaceae as abundant primary fermenters in anaerobic digesters treating waste activated sludge. *Front. Microbiol.* **8**, 1134 (2017).

52. K. E. Thorpe, P. Joski, K. J. Johnston, Antibiotic-resistant infection treatment costs have doubled since 2002, now exceeding $2 billion annually. *Health Aff. (Millwood)* **37**, 662–669 (2018).

53. R. C. Power, D. C. Salazar-García, R. M. Wittig, M. Freiberg, A. G. Henry, Dental calculus evidence of Taï Forest Chimpanzee plant consumption and life history transitions. *Sci. Rep.* **5**, 15161 (2015).

54. E. Cristiani, A. Radini, M. Edinborough, D. Borić, Dental calculus reveals Mesolithic foragers in the Balkans consumed domesticated plant foods. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10298–10303 (2016).

55. A. Nava et al., Multipronged dental analyses reveal dietary differences in last foragers and first farmers at Grotta Continenza, central Italy (15,500-7000 BP). *Sci. Rep.* **11**, 4261 (2021).

56. M. W. Pedersen et al., Postglacial viability and colonization in North America's ice-free corridor. *Nature* **537**, 45–49 (2016).

57. D. Holst, Hazelnut economy of early Holocene hunter–gatherers: A case study from Mesolithic Duvensee, northern Germany. *J. Archaeol. Sci.* **37**, 2871–2880 (2010).

58. S. Mithen, N. Finlay, W. Carruthers, S. Carter, P. Ashmore, Plant use in the mesolithic: Evidence from Staosnaig, Isle of Colonsay, Scotland. *J. Archaeol. Sci.* **28**, 223–234 (2001).

59. M. Divišová, P. Šída, Plant use in the Mesolithic period. Archaeobotanical data from the Czech Republic in a European context–A review. *Interdisciplinaria Archaeologica: Natural Sciences in Archaeology* **6**, 95–106 (2015).

60. E. L. Sturtevant, U. P. Hedrick, *Sturtevant's Edible Plants of the World* (Dover Publications, 1972).

61. P. R. B. Kozowyk, M. Soressi, D. Pomstra, G. H. J. Langejans, Experimental methods for the Palaeolithic dry distillation of birch bark: Implications for the origin and development of Neandertal adhesive technology. *Sci. Rep.* **7**, 8033 (2017).

62. T. Z. T. Jensen et al., A 5700 year-old human genome and oral microbiome from chewed birch pitch. *Nat. Commun.* **10**, 5520 (2019).

63. A. Radini et al., Neanderthals, trees and dental calculus: New evidence from El Sidrón. *Antiquity* **90**, 290–301 (2016).

64. B. Llamas et al., From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *Sci. Technol. Archaeol. Res.* **3**, 1–14 (2017).

65. J. Dabney, M. Meyer, "Extraction of highly degraded DNA from ancient bones and teeth" in *Ancient DNA: Methods and Protocols*, B. Shapiro et al., Eds. (Springer, New York, 2019), pp. 25–29.

66. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, db.prot5448 (2010).

67. M.-T. Gansauge et al., Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, e79 (2017).

68. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).

69. R. Schmieder, R. Edwards, Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).

70. A. T. Ozga et al., Oral microbiome diversity in chimpanzees from Gombe National Park. *Sci. Rep.* **9**, 17354 (2019).

71. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).

72. R. Hübler et al., HOPS: Automated detection and authentication of pathogen DNA in archaeological remains. *Genome Biol.* **20**, 280 (2019).

73. J. Lu, F. P. Breitwieser, P. Thielen, S. L. Salzberg, Bracken: Estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104 (2017).

74. G. B. Gloor, J. M. Macklaim, V. Pawlowsky-Glahn, J. J. Egozcue, Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* **8**, 2224 (2017).

75. D. H. Parks, G. W. Tyson, P. Hugenholtz, R. G. Beiko, STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**, 3123–3124 (2014).

76. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).

77. F. P. Breitwieser, S. L. Salzberg, Pavian: Interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics* **36**, 1303–1304 (2020).

78. A. Namouchi et al., Integrative approach using Yersinia pestis genomes to revisit the historical landscape of plague during the Medieval period. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11790–E11797 (2018).

79. A. L. Price, N. C. Jones, P. A. Pevzner, De novo identification of repeat families in large genomes. *Bioinformatics* **21** (suppl. 1), i351–i358 (2005).

80. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).

81. M. Gan, Q. Liu, C. Yang, Q. Gao, T. Luo, Deep whole-genome sequencing to detect mixed infection of mycobacterium tuberculosis. *PLoS One* **11**, e0159029 (2016).

82. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

83. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

84. B. Q. Minh, M. A. T. Nguyen, A. von Haeseler, Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).

85. S. Guindon et al., New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

86. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

87. J. F. C. Kingman, The coalescent. *Stochastic Process. Appl.* **13**, 235–248 (1982).

Ottoni et al.
Tracking the transition to agriculture in Southern Europe through ancient DNA analysis of dental calculus

PNAS | 11 of 11
https://doi.org/10.1073/pnas.2102116118

ANTHROPOLOGY

www.manaraa.com