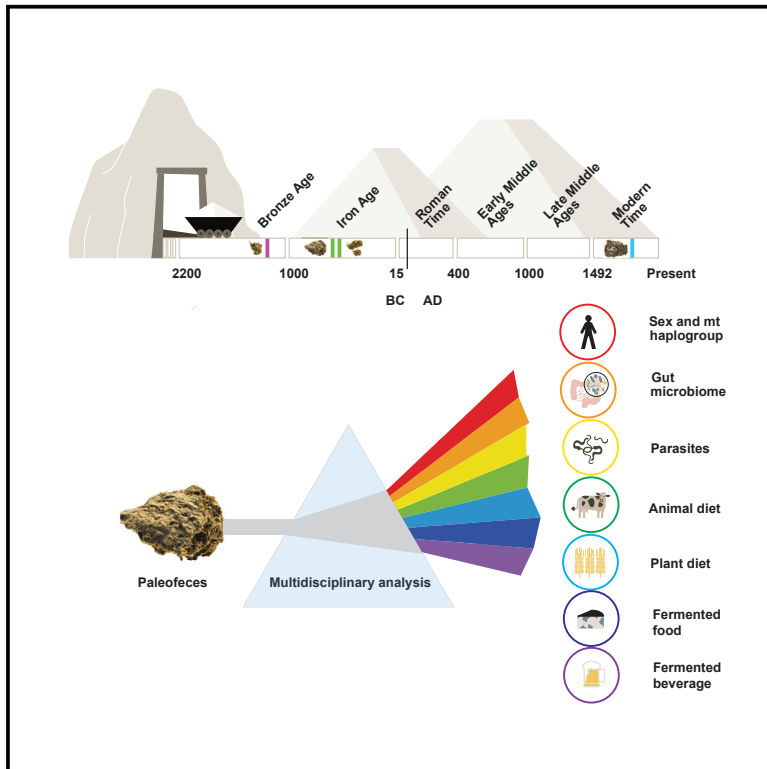


Current Biology

Hallstatt miners consumed blue cheese and beer during the Iron Age and retained a non-Westernized gut microbiome until the Baroque period

Graphical abstract



Authors

Frank Maixner, Mohamed S. Sarhan, Kun D. Huang, ..., Albert Zink, Hans Reschreiter, Kerstin Kowarik

Correspondence

frank.maixner@eurac.edu (F.M.),
kerstin.kowarik@nhm-wien.ac.at (K.K.)

In brief

Maixner et al. describe the gut microbiome and diet of European salt miners using paleofeces dating from the Bronze Age to the Baroque period. This analysis provides evidence for recent changes in the gut microbiome due to industrialization and for the consumption of fermented food and beverages in Iron Age Europe.

Highlights

- Gut microbiome and diet of European salt miners determined using paleofeces
- Until the Baroque, the microbiome resembled that of modern non-Westernized people
- Food-fermenting fungi in Iron Age feces indicates blue cheese and beer consumption

Article

Hallstatt miners consumed blue cheese and beer during the Iron Age and retained a non-Westernized gut microbiome until the Baroque period

Frank Maixner,^{1,12,13,16,17,*} Mohamed S. Sarhan,^{1,12,16} Kun D. Huang,^{2,3,16} Adrian Tett,^{2,4} Alexander Schoenafinger,^{1,5,12} Stefania Zingale,^{1,12} Aitor Blanco-Míguez,² Paolo Manghi,² Jan Cemper-Kiesslich,⁶ Wilfried Rosendahl,^{7,8} Ulrike Kusebauch,⁹ Seamus R. Morrone,⁹ Michael R. Hoopmann,⁹ Omar Rota-Stabelli,¹⁰ Thomas Rattei,⁴ Robert L. Moritz,⁹ Klaus Oegg,⁵ Nicola Segata,^{2,14,16} Albert Zink,^{1,16} Hans Reschreiter,^{11,16} and Kerstin Kowarik^{11,15,16,*}

¹Institute for Mummy Studies, EURAC Research, Viale Druso 1, 39100 Bolzano, Italy

²Department CIBIO, University of Trento, Via Sommarive 9, 38123 Povo (Trento), Italy

³Department of Sustainable Agro-Ecosystems and Bioresources, Fondazione Edmund Mach, Via Edmund Mach 1, 38010 San Michele all'Adige (TN), Italy

⁴CUBE (Division of Computational Systems Biology), Centre for Microbiology and Environmental Systems Science, University of Vienna, Althanstraße 14, 1090 Vienna, Austria

⁵Institute of Botany, University of Innsbruck, Sternwartestraße 15, 6020 Innsbruck, Austria

⁶Interfaculty Department of Legal Medicine & Department of Classics, University of Salzburg, Ignaz-Harrer-Straße 79, 5020 Salzburg, Austria

⁷Reiss-Engelhorn-Museen, Zeughaus C5, 68159 Mannheim, Germany

⁸Curt-Egelhorn-Zentrum Archäometrie, D6,3, 61859 Mannheim, Germany

⁹Institute for Systems Biology, 401 Terry Avenue North, Seattle, WA 98109, USA

¹⁰Center Agriculture Food Environment (C3A), University of Trento, 38010 San Michele all'Adige (TN), Italy

¹¹Prehistoric Department, Museum of Natural History Vienna, Burgring 7, 1010 Vienna, Austria

¹²Twitter: @EuracMummy

¹³Twitter: @FrankMaixner

¹⁴Twitter: @cibiocm

¹⁵Twitter: @KowarikKerstin

¹⁶These authors contributed equally

¹⁷Lead contact

*Correspondence: frank.maixner@eurac.edu (F.M.), kerstin.kowarik@nhm-wien.ac.at (K.K.)

<https://doi.org/10.1016/j.cub.2021.09.031>

SUMMARY

We subjected human paleofeces dating from the Bronze Age to the Baroque period (18th century AD) to in-depth microscopic, metagenomic, and proteomic analyses. The paleofeces were preserved in the underground salt mines of the UNESCO World Heritage site of Hallstatt in Austria. This allowed us to reconstruct the diet of the former population and gain insights into their ancient gut microbiome composition. Our dietary survey identified bran and glumes of different cereals as some of the most prevalent plant fragments. This highly fibrous, carbohydrate-rich diet was supplemented with proteins from broad beans and occasionally with fruits, nuts, or animal food products. Due to these traditional dietary habits, all ancient miners up to the Baroque period have gut microbiome structures akin to modern non-Westernized individuals whose diets are also mainly composed of unprocessed foods and fresh fruits and vegetables. This may indicate a shift in the gut community composition of modern Westernized populations due to quite recent dietary and lifestyle changes. When we extended our microbial survey to fungi present in the paleofeces, in one of the Iron Age samples, we observed a high abundance of *Penicillium roqueforti* and *Saccharomyces cerevisiae* DNA. Genome-wide analysis indicates that both fungi were involved in food fermentation and provides the first molecular evidence for blue cheese and beer consumption in Iron Age Europe.

INTRODUCTION

Paleofeces are naturally preserved ancient feces found in dry caves, desert areas, waterlogged environments, and frozen habitats. Specific environmental processes such as desiccation or freezing prevent their deterioration in mummies, ancient latrines, bogs, and soils.¹ Previous studies have shown that paleofecal material still contains plant macro- and microfossils, parasite

eggs, and even ancient biomolecules (DNA, proteins, metabolites).² Ancient paleofeces have therefore recently been used as a source of information to study prehistoric nutrition patterns^{3–5} and health^{6,7} and to analyze single representatives^{8,9} or the overall composition of the intestinal microbiome of our ancestors.^{10–12}

One of the few archaeological sites where well-preserved paleofeces can be found is the protohistoric salt mines of the

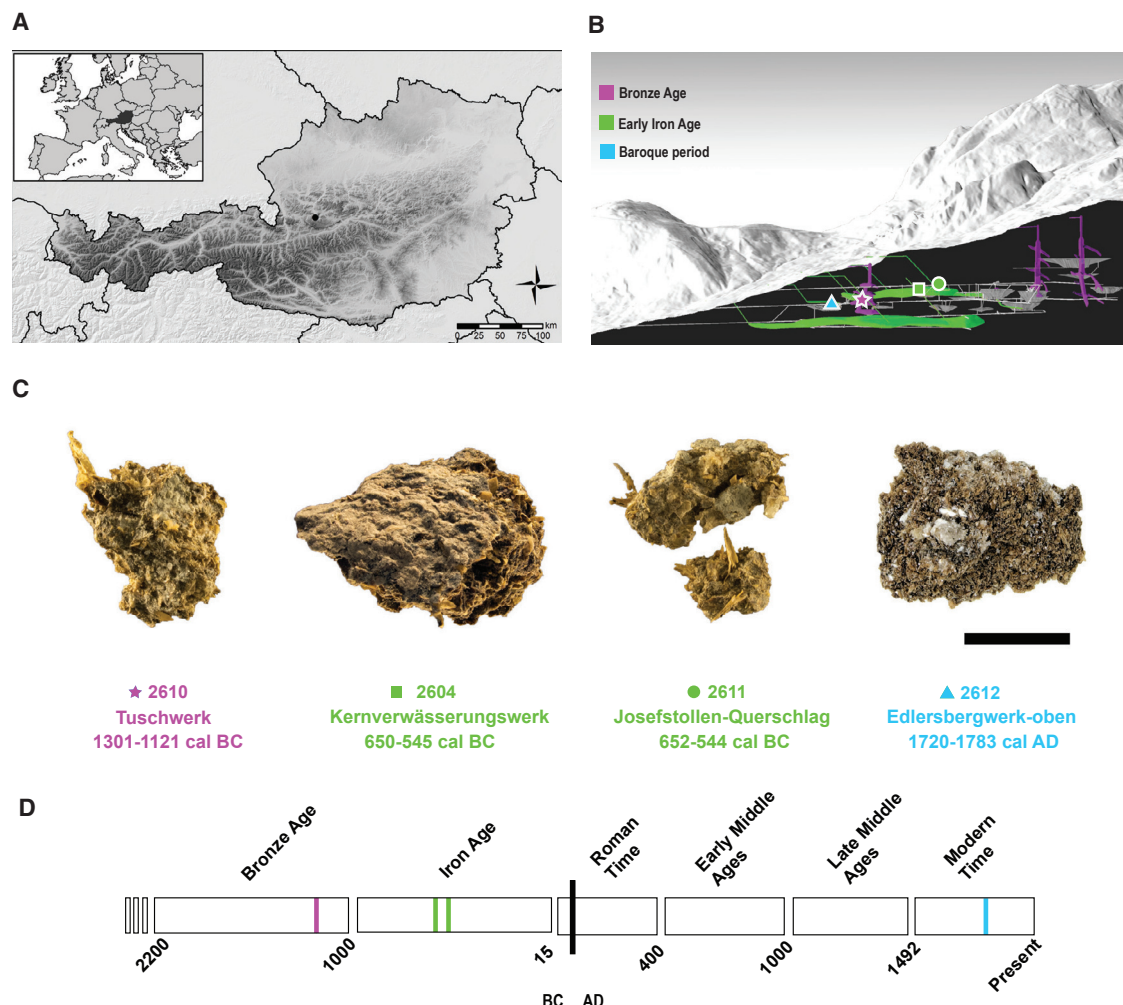


Figure 1. The Hallstatt salt mine and radiocarbon-dated paleofeces samples used in this study

(A) The salt mines are located in Upper Austria.

(B) Finding sites of the four paleofeces samples in the Bronze Age, Iron Age, and Baroque mining area. The symbol color corresponds to the radiocarbon date of the paleofeces.

(C) Macroscopic appearance of the four paleofeces samples. The scale bar corresponds to 1 cm of length. The sample description includes the sample ID, the mine workings name, and the radiocarbon date. The provided radiocarbon date range corresponds to the Cal 2-sigma values with the highest probability.

(D) Temporal assignment of the radiocarbon-dated paleofeces to the major European time periods from the Bronze Age onward.

See also [Figure S1](#) for details of the salt crystals surrounding sample 2612. [Data S1A](#) and [S1B](#) provide additional information about the samples and the radiocarbon dating.

Austrian UNESCO World Heritage area Hallstatt-Dachstein/Salzammergut. Protohistoric salt mines offer ideal preservation conditions for organic materials. The high salt concentrations and the constant annual temperature of around 8°C inside the isolated mine workings preserve organic artifacts very well. The Hallstatt salt mines located in the Eastern Alps ([Figure 1](#)) offer one of the world's oldest and most continuous record of underground salt mining.^{13–15} Large-scale underground mining in the Hallstatt salt mountains dates back at least to the 14th century BC (late Bronze Age). Several protohistoric (Bronze, Iron Age) and historic (14th century AD to present) mining phases are well documented. The site also gave name to the early period of the Iron Age in Europe, the so-called Hallstatt Period (800 to 400 BC). Dense layers of production waste reaching several

meters of thickness were excavated from the protohistoric Bronze Age and Iron Age mine workings of Hallstatt, uncovering thousands of wooden tools and construction elements, implements made from fur, rawhide, hundreds of woolen textile fragments, grass, bast ropes, and human excrements.¹⁶ These objects provide insights into the daily life of a Bronze Age and Iron Age mining community ranging from mining technology, organization of production, and resource management to human health, dietary habits, social organization of production processes, and social status within a mining system. These aspects have been studied extensively in Hallstatt based on a combination of data sources encompassing the protohistoric mine working, Bronze Age meat-curing facilities, and large Iron Age cemetery.^{14,16,17}

Here we focus on the question of the structure and evolution of dietary habits as well as the human gut microbiome in one of Europe's most important early production communities. We used microscopic, metagenomic, and proteomic analysis to characterize nutrition patterns of the protohistoric (Bronze Age, early Iron Age) and historic (Baroque period, 18th century AD) miners and metagenomic analysis to determine the structure and evolution of the gut microbiome. Our findings will enhance the understanding of early European dietary habits (especially the production and consumption of processed foodstuffs) and provide further evidence of the recent change in gut microbiome structure as result of industrialization and Westernization processes.

RESULTS

Paleofeces from Bronze Age to Baroque Period contain ancient endogenous DNA

In this study, we initially subjected four paleofeces samples, collected from Bronze Age and Iron Age Hallstatt mine workings, to radiocarbon dating, then to in-depth microscopic and molecular analysis (Figures 1A–1C; STAR Methods; Data S1A). The four paleofeces samples can be macroscopically differentiated into three samples containing a high amount of fibrous plant material (2610, 2604, 2611) and one more homogeneous sample (2612) that does not contain any visible larger plant fragments (Figures 1 and S1A). Radiocarbon analyses date the roughly structured samples to the Late Bronze Age (2610) and Iron Age (2604, 2611), which is in perfect accordance with the proposed period of usage of the mine workings where the paleofeces have been found.¹⁵ In contrast, the fine-textured paleofeces 2612 sampled in an Iron Age minedates to the Baroque period (18th century AD) (Figure 1C; Data S1B). For this part of the salt mine, however, it is historically documented that the mine workings had started to be reused from the beginning of 18th century onward.¹⁸ Independently of the paleofeces' age, their storage time since excavation (some samples were recovered in the year 1983), or the mode of excavation (wet sieving versus direct sampling) (Figure S1), we could retrieve biomolecules (DNA and protein) from all samples for the subsequent molecular analysis (Data S1 and S2; STAR Methods). Proteomics analysis provided the first evidence for the presence of endogenous biomolecules in the paleofeces material. The most abundant peptides were assigned to human intestinal tract proteins that are involved in food digestive processes (Data S2B, S2D, S2F, and S2H). The DNA of the paleofeces material was further subjected to a deep shotgun sequencing approach resulting in 57,130,584 to 221,314,691 quality-filtered reads (Data S1C). A first taxonomic overview using DIAMOND against the NCBI NR database revealed that the majority of reads in the samples are assigned to *Bacteria* (93.9% to 78.9% of all assigned reads), with *Firmicutes* and *Bacteroidetes* being the most abundant phyla of this kingdom (Figures S2A–S2D). Less than 7.5% of the reads were eukaryotic, with up to 6.7% fungal reads in sample 2604. The *Metazoa* and *Viridiplantae* reads, important for the molecular reconstruction of the diet, comprised 0.5% to 0.01% of all assigned reads. Further analysis of the human DNA in the paleofeces revealed an endogenous DNA content between 0.26% and 0.06%, sufficient for molecular sex and mitochondrial haplogroup assignment

(Data S1D). The highly fragmented human reads display a very low deamination pattern at the 5' ends (Figures S2E–S2H). In the most recent sample (2612), the reads appear even less fragmented and display almost no DNA damage. Considering the age of these samples, the DNA damage is exceptionally low. This high preservation is most likely due to the rapid desiccation of the samples in the salt mine, which may result in reduced hydrolytic damage of the biomolecules. Our analyses show that the four paleofeces come from male individuals that carry distinct mitogenomes with low contamination estimates (1% to 2%), indicating that each sample represents unique unmixed ancient feces.

Ancient paleofeces display a gut microbiome structure similar to modern non-Westernized individuals

We compared the microbiome structure of the paleofeces to a large number of contemporary metagenomes ($n = 823$) (Data S1E). Principal coordinate analysis (PCoA) performed on a species-level taxonomic composition shows that the paleofeces from the Bronze Age to the Baroque period cluster with stool samples from contemporary non-Westernized individuals (Figure 2A) with diets mainly consisting of unprocessed foods and fresh fruits and vegetables.¹⁹ This clustering is similarly observed for encoded metabolic pathways (Figure 2B). All the paleofecal samples were distinct from the oral and, more importantly, from the soil samples, suggesting little evidence of soil contamination, which is sometimes observed in ancient metagenomics studies.²⁰ The source prediction analysis further supports the sample preservation (Data S1F).

To further assess the paleofeces samples, we analyzed the prevalence of the top 15 most abundant species in the paleofeces compared to 8,968 gut microbiomes of healthy Westernized and non-Westernized adults (Data S1G). As a result, 13 out of the 15 most abundant species were identified to be associated with human gut environment, of which 11 species were found to be more prevalent in modern non-Westernized compared to Westernized populations. Five of these species, *Bifidobacterium angulatum*, *Lactobacillus ruminis*, *Catenibacterium mitsuokai*, *Prevotella copri*, and *Clostridium ventriculi*, were over twice as prevalent in non-Westernized populations (Figure 2C; Data S1H). One of the two species not associated with the human gut is the halophilic archaeon *Halococcus morrhuae*, which survives on a high concentration of salt.²² It was observed in low abundance in the paleofeces sample 2612, the only sample that was not subjected to wet sieving. Therefore, we assume that the archaeon was introduced from the environment via the salt crystals (Figure S1B). In the paleofeces samples 2604 and 2611, we also identified *Clostridium perfringens*, a known intestinal foodborne pathogen,²³ that also occurs free living in the soil and other environments.²⁴ Since an infection with *C. perfringens* causes acute diarrhea and the paleofeces does not indicate any characteristics pointing to that disease, we assume that the presence of this bacterium is due to an environmental contaminant rather than a remnant of food spoilage in the miners' gut. When all paleofecal microbiome members were considered in population prevalence analysis, 100 out of 158 species were found in $\geq 5\%$ stool samples from modern healthy adult individuals and 65% of these species are overrepresented in non-Westernized populations in comparison with

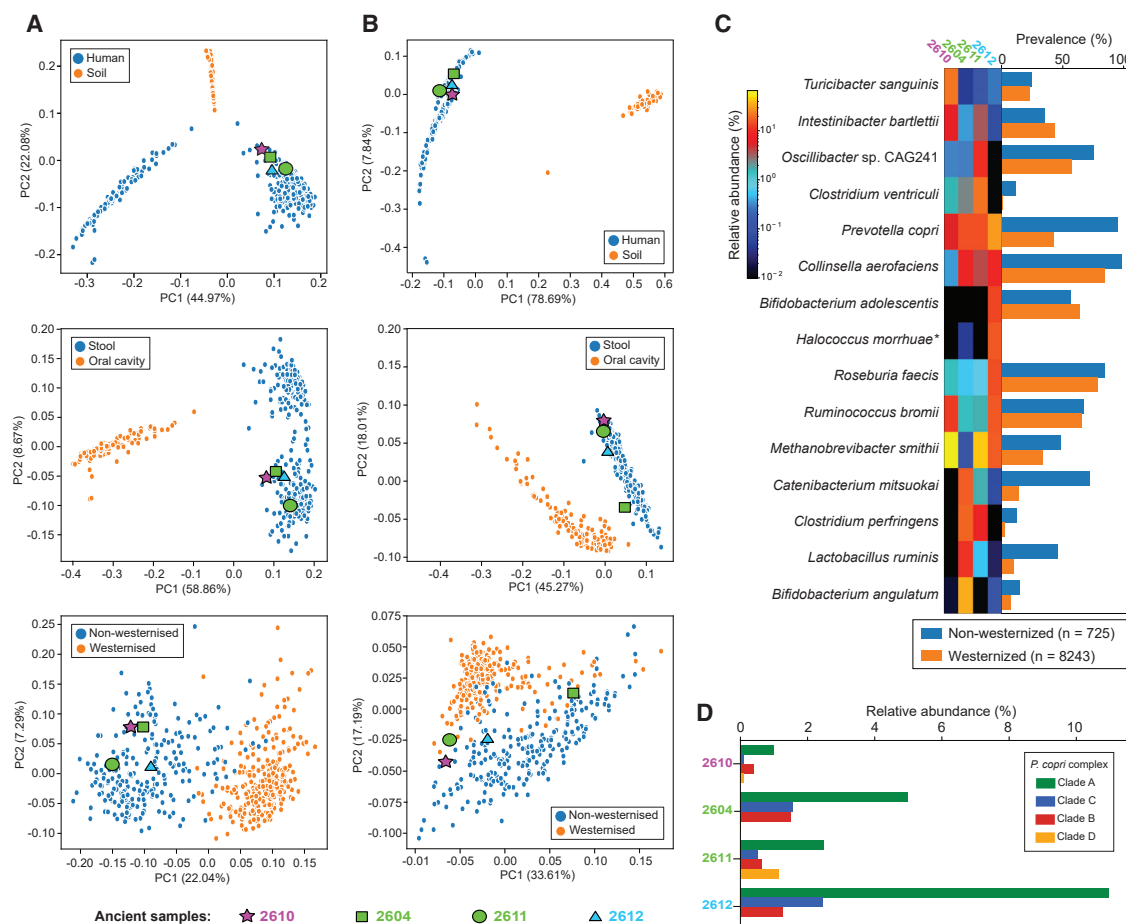


Figure 2. Overview of microbial composition and metabolic pathways of paleofeces samples in comparison to a large collection of contemporary metagenomes

(A) Principal coordinate analysis (PCoA) based on microbial abundance profiled using MetaPhlAn 3.0²¹ between four paleofeces samples and 823 contemporary samples characterized by sampling environment, body site, and non-Westernized lifestyle.

(B) Principal coordinate analysis (PCoA) based on metabolic pathway abundance profiled using HUMAnN 3.0²¹ between four paleofeces samples and the same contemporary samples used in (A).

(C) Prevalence of the top 15 most enriched species of four paleofeces samples in non-Westernized and Westernized datasets comprising 8,968 stool samples from healthy adult individuals. Asterisk indicates species that is likely from external contamination.

(D) Relative abundance of *P. copri* four clades estimated using MetaPhlAn 3.0²¹ in each paleofecal sample.

See also Figure S1 for additional microbial profiles in the DNA “wash-out” experiment. Data S1 contains additional information about the comparative datasets and the results obtained by the prevalence and abundance analysis.

the Westernized equivalent (Data S1J). A similar result was obtained when we decreased the prevalence threshold to 1% (Data S1J).

The *Prevotella copri* complex, which is highly prevalent in non-Westernized populations and prevalent in previously investigated ancient samples,⁹ was identified in all paleofecal samples, representing, on average, 7.3% (~1.6%–14.7%) of the relative abundance (Figure 2C; Data S1K). Consistent with previous findings,⁹ we found multiple clades of the complex to be present in each of the paleofeces samples (Figure 2D) with the exception of Clade D, which was barely detectable in sample 2604 and 2612. All other clades were detected with relative abundances > 0.01% in all samples (Figure 2D; Data S1L). Of note, in contrast to other samples, the sample 2604 displayed higher abundance of bacterial species such as *Lactobacillus brevis*,

Bifidobacterium merycicum, *Bifidobacterium angulatum*, and *Lactobacillus plantarum* (Data S1K) that are known to be of probiotic activities or involved in processing of dairy products.²⁵

Microscopic and molecular reconstruction of the Hallstatt miners' diet

Next, we aimed to reconstruct the dietary components in the paleofeces using both a microscopic and a molecular survey. The above-mentioned structural differences between the paleofeces became even more evident in the microscopic analyses. The Baroque period sample 2612 was much finer textured than all other samples from protohistory (Figure 1C). This was also reflected in the macro-remain composition of the paleofeces, showing that samples 2610, 2604, and 2611 contained a lot of seeds contrary to 2612, which consisted of frequent tissues of

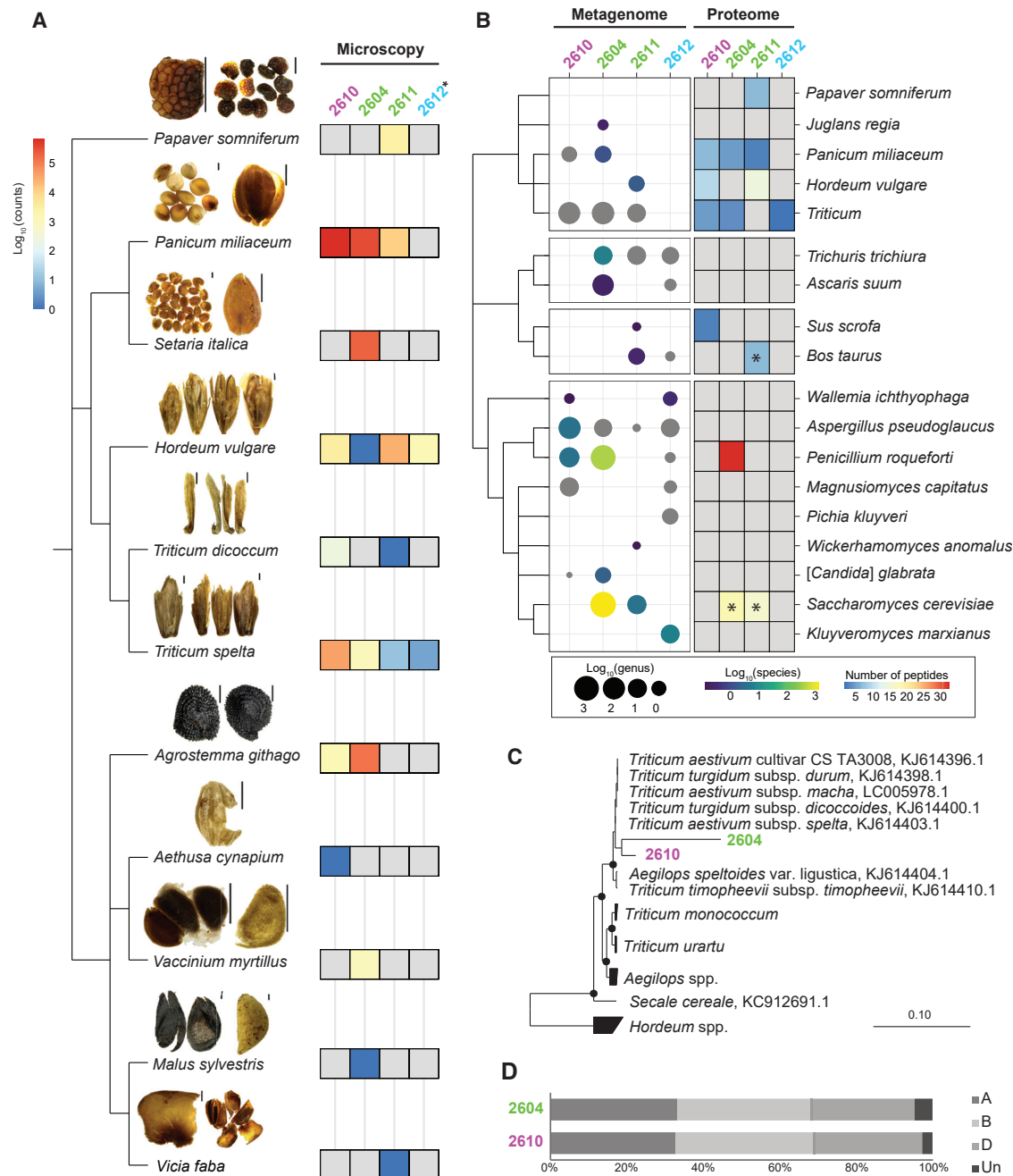


Figure 3. Microscopic and molecular dietary analysis of the Hallstatt paleofeces

(A) Plant macro-remains microscopically detected in the four paleofeces samples. The scale bar indicates 1 mm of length. The heatmap shows the log-scale macro-remain counts normalized to 3.7 g sample. The sample with asterisk was assessed in a semiquantitative manner. For further details, please refer to [Data S1](#).

(B) Most abundant taxa (plants, nematodes, animals, fungi) detected in the four paleofeces metagenomes and proteomes. The circle size and circle color correspond to log₁₀ “normalized” number of reads per million at genus and species levels, respectively. The asterisks in the proteome heatmap mean the peptides were assigned only to genus level.

(C) Phylogenetic assignment of two partial *Triticum* chloroplast genomes in the 2604 and 2610 metagenomes. The comparative dataset included complete chloroplast genomes of selected members of the *Triticeae* tribe (NCBI accession numbers are provided in the figure). The tree was calculated using the maximum-likelihood algorithm (PhyML) based on 136,160 informative positions. Black circles symbolize parsimony and neighbor joining bootstrap support (>90%) based on 100 and 1,000 iterations, respectively. The scale bar indicates 10% estimated sequence divergence.

(legend continued on next page)

fruit husks and seed coats (Figure 3A; Data S1M). Generally, all samples displayed a predominance of cereal remains.

Microscopic analysis revealed that the Bronze Age sample (2610) consisted more or less exclusively of cereal remains, which originated from barley (*Hordeum vulgare*), spelt (*Triticum spelta*), some emmer (*Triticum dicoccum*), proso millet (*Panicum miliaceum*), and a few weeds, e.g., corn cockle (*Agrostemma githago*) and poison parsley (*Aethusa cynapium*). The Iron Age samples (2604, 2611) were characterized by a predominance of cereal remains from barley (*Hordeum vulgare*), spelt (*Triticum spelta*), millets (*P. miliaceum*, *Setaria italica*), and a little emmer (*T. dicoccum*). Furthermore, in sample 2611, testa remains of broad beans (*Vicia faba*) and seeds of opium poppies (*Papaver somniferum*) were observed. Crab apples (*Malus sylvestris*) and bilberries/cranberries (*Vaccinium myrtillus/vitis-idea*) in sample 2604 document the consumption of gathered wild fruits. Striking in the Iron Age sample 2604 was the contamination with weeds, in particular corn cockle (*Agrostemma githago*). In the sample 2612 from the Baroque time, the microscopic pattern was notably different from the other samples. The plant material was finely ground, and entire fruits were missing apart from a digested mericarp of anise (*Pimpinella anisum*). The plant tissue belonged to bran (fragments of cereal testa, pericarp, hairs, hilum, endosperm) of wheat (*Triticum* sp.). The precise species was unidentifiable, but according to the rare occurrence of tube cells in the pericarp fragments, a member of the tetra- or hexaploid wheat group is suggested. Bran of barley (*H. vulgare*) was also observed in minor quantities. Furthermore, the consumption of legumes is documented by testa remains of garden bean (*Phaseolus vulgare*) in this sample.

In addition to the microscopic analysis, we subjected paleofeces biomolecules (DNA and proteins) to molecular dietary analyses. Both metagenomic and proteomic analyses included a homology search against different databases, followed by strict filtering steps of the obtained hits and a subsequent in-depth analysis of selected identified taxa (STAR Methods; Figure S3; Data S1N, S1O, S2B, S2D, S2F, and S2H). For the plant diet, we could confirm the presence of the most abundant domesticated plant macro-remains, including broomcorn millet (*P. miliaceum*), barley (*H. vulgare*), and wheat (*Triticum* spp.) (Figure 3B). In addition, we found DNA-based evidence of the presence of walnut (*Juglans regia*) in the sample 2604 and protein-based support for the occurrence of opium poppy seeds (*P. somniferum*) in the sample 2611. In addition to the foxtail millet (*S. italica*), which appeared with high grain number, all the low abundant wild plants unveiled by the microscopic investigation were not identified in our molecular survey, which has undergone strict filtering to minimize the false positives (STAR Methods). Further phylogenetic analysis assigned the *Triticum* spp. chloroplast genomes of the samples 2604 and 2610 closest to the chloroplasts of tetraploid (emmer, durum) and hexaploid (spelt wheat, bread wheat) wheat varieties, respectively (Figures 3C,

S3A, and S3B; Data S1P). Additional comparison with the bread wheat genome revealed an equal subgenome (A, B, and D) representation in the 2604 and 2610 metagenomic reads, which suggests, in combination with the microscopic identification of numerous characteristic grains, glumes, and spikelets, the presence of hexaploid spelt wheat (*T. spelta*) in these paleofeces (Figure 3D). Beside the plant diet, we obtained molecular evidence for the consumption of cattle (*Bos taurus*) and swine meats (*Sus scrofa*) throughout all investigated time periods (Figures 3B and S3D). Interestingly, the most abundant cattle proteins in sample 2611 (hemoglobin and coagulation proteins) indicate the plant diet was supplemented by blood-rich animal tissues (e.g., muscle, liver) (Data S2F). The molecular analyses revealed in addition, that individuals from both the Iron Age (2611) and the Baroque (2612) suffered from intestinal infections of whipworms (*Trichuris trichuria*) and roundworm (*Ascaris* spp.) (Figure 3B and S3D–S3G). Finally, all samples showed a continuous low background with fungal DNA mainly coming from different Ascomycota.

Molecular evidence for blue cheese and beer consumption during the Iron Age

In contrast to all other samples, the Iron Age sample 2604 displayed an exceptionally high abundance of *Penicillium roqueforti* and *Saccharomyces cerevisiae* proteins (Data S2D) and DNA (Data S1N), making up to 7%–22% of total eukaryotic reads. This was characteristic of this sample as compared with the other samples that did not show such prevalence—even the sample 2611, which was taken from a similar context and dated back to the same time point. To authenticate the data and gain further insights into their potential ecological significance, we mapped the high-quality reads of sample 2604 against the reference genomes of these two fungi (Figure S4A; Data S1O). With 11–13× coverage, we were able to reconstruct >92% of both genomes, displaying even coverage and SNP distribution. To confirm whether these two fungi are of ancient origin and not modern contaminants, we initially checked the ancient DNA damage pattern of the mapped reads. Both fungi displayed typical ancient DNA damage patterns, with levels comparable to the human endogenous DNA (Figures S4B and S4C). Hence, and considering their extraordinarily high abundance and exclusive incidence in this sample, we assumed their endogenous originality to the coprolite microbial community. Additionally, both fungi are commonly used nowadays in food processing: *P. roqueforti* is used for cheese fermentation, and *S. cerevisiae* is used for fermenting bread and alcoholic beverages including beer, mead, and wine. Therefore, we assume that they could have been involved in food processing at that time. To test this assumption, we used the reconstructed genomes for further comparative phylogeny and population genetic analyses to infer whether they had been truly involved in food processing or were just transient environmental microbes.

(D) Wheat subgenome (A, B, and D) representation in the 2604 and 2610 metagenomes (Data S1), aligned to the modern hexaploid bread wheat reference genome (accession number GCA_900519105). Both the wheat chloroplast and nuclear reads were highly fragmented and display aDNA-specific damage patterns (Figures S3H–S3K).

See also Figure S3 for details about the comparative analysis, phylogenetic assignment, and damage pattern of selected plant, animal, and parasite DNA. Data S1 provides further details of the macro-remains, comparative datasets, dietary DNA, and mapping statistics. Data S2 provides additional information about the comparative datasets and proteomics results.

First, we compared our putative *P. roqueforti* strain to 33 other sequenced modern *P. roqueforti* strains coming from different functional niches.²⁶ The comparative dataset included 18 cheese-fermenting and 15 non-cheese-fermenting strains (Data S1Q), in addition to *Penicillium psychrosexualis* and *Penicillium carneum* as an outgroup. After mapping the raw reads of all strains to the reference *P. roqueforti* genome FM164 and data filtering, we resolved 120,337 SNPs, which were used for inferring maximum likelihood (ML) phylogenetic relationships among the tested strains (STAR Methods). Consistent with the original publication of Dumas and colleagues,²⁶ the resulting phylogeny revealed four distinct clades: a Roquefort cheese clade, a non-Roquefort cheese clade containing blue cheeses others than Roquefort, a silage/food spoilage clade, and a wood/food spoilage clade (Figure 4A). Initially, the phylogenetic analysis separated the non-Roquefort cheese clade from the other, then the Roquefort cheese clade was diverged from the other food spoilage clades. The ancient *P. roqueforti* strain showed highest similarity to the non-Roquefort cheese strains, being clustered together with their corresponding clade as an earlier divergent. The reason behind such early divergence might be attributed to the recent acquisition of some genomic regions—most importantly, *CheesyTer* and *Wallaby*—by the non-Roquefort cheese strains. This gene acquisition most likely happened via repeated multiplication of selected spores of the best cheeses on bread used as a growth medium in the late 19th century and early 20th century before the advent of microbiological *in vitro* culturing techniques.²⁷ Thereby, modern non-Roquefort strains were exposed to extensive selection coupled to horizontal gene transfer events from other cheese-producing *Penicillium* spp. or even other genera.^{26,27–29} Importantly, our Iron Age strain did not contain any of those recently acquired fragments, which comes in congruence with the hypothesis that such domestication events occurred during the last two centuries.

We further used the tool ADMIXTURE to infer the degree of admixture among the strains. By assuming the presence of 3 ancestries ($K = 3$), we could clearly distinguish the non-Roquefort, Roquefort, and food spoilage strains (Figure 4B). Our putative strain displayed ~70% cheese-producing ancestry (60% of the non-Roquefort and 10% of Roquefort cheese) and ~30% food-spoiler ancestry. Both the phylogenetic placement and ADMIXTURE profile indicate that the ancient *P. roqueforti* has already been under positive selection toward the non-Roquefort cheese cluster, a selection process that most likely occurred during the process of cheese production. Some archeological findings excavated from the mines might have been used for that purpose (Figure 4C), as they showed some traces of fatty food products.

Next, we compared the ancient *S. cerevisiae* genome to 157 recent strains coming from different ecological niches, i.e., food, alcoholic beverages (e.g., beer, wine, sake, and spirits), biofuels, and laboratories, as well as wild strains (Data S1R). ML phylogenetic analysis, based on 375,629 SNP positions, distinguished 2 main clades. The first main clade splits into two subclades, with one containing most of the beer strains (beer 1 clade) that show a successive sub-clustering based on the origin of the strains (Figure 4D). The other subclade (henceforth referred to as “mixed” clade) included a mixture of bread, wine, beer, and spirit strains. The second main clade is composed of two subclades: a wine clade and another beer clade (beer 2). All other

wild, laboratory, and sake strains fall to the base of the whole phylogeny. The ancient *S. cerevisiae* strain clustered basal to the second main clade, which includes the wine and beer 2 strains. Further population structure analysis displayed high admixture in our putative strain, resembling primarily the wine ancestral population (47%), followed by 29.2% beer ancestries (Figure S4D) and only 19% wild strain ancestry. Therefore, and considering the ML phylogenetic assignment, we assume that the possibility of our strain to be of wild origin is unlikely. The results rather indicate higher similarity to wine and beer strains. Principal component analysis (PCA) provided further indication for the domestication of our strain in alcoholic beverage fermentation. Along the PC1 that explains 25.42% of the variation, our strain clustered closer to the strains of beer 2 than to the strains of the wine clade (Figure 4E). This was further supported with proteomic analysis (Data S2D) that unveiled that most of the peptides assigned to the genus *Saccharomyces* derived from proteins involved in alcohol fermentation pathways (e.g., glycolysis). To further narrow down the possible routes of domestication, we decided to differentiate the strains based on functional marker genes (Data S1S). According to recent literature,^{31,32,33} the genes *RTM1*, *BIO1/BIO6*, and the chromosomal regions A/B/C can be used to differentiate yeast strains based on their functional niches. The gene *RTM1* is a strong domestication marker responsible for conferring resistance against the toxicity of molasses and other rich-sugar substrates and is assumed to be positively selected in beer yeast strains.^{32,34} The genes *BIO1* and *BIO6*, which are involved in *de novo* biosynthesis of biotin, are highly selected in sake fermenting yeasts, due to lack of biotin in the fermentation substrates, such as rice.³⁵ The regions A, B, and C are horizontally transferred genomic regions from other yeast genera, e.g., *Kluyveromyces*, *Pichia*, and *Zygosaccharomyces*.³⁶ These regions contain 39 genes distributed over 3 different chromosomes and are assumed to play a role in wine fermentation.

Therefore, we searched the presence of these marker genes in our comparative dataset, including our ancient strain. In accordance with the literature, almost all beer strains—either of clade 1 or clade 2—were positive for *RTM1*, while the wine clade was mainly positive for the genomic regions A/B/C. The mixed clade contained both *RTM1* and the regions A/B/C. The sake clade exclusively contained the *BIO1/BIO6* genes and partially the *RTM1* gene. The wild strains, isolated from cacao in Africa, clustered in the basal clade and did not contain any of these marker genes (Figure 4D; Data S1T), contrary to our strain that contained the *RTM1* and lacked the *BIO1/BIO6* genes and the regions A/B/C.

Based on the previous findings—i.e., (1) the ancient DNA damage profile, (2) the high prevalence of *S. cerevisiae* reads, (3) the presence of fermentable cereal substrates such as wheat and barley, (4) the phylogenetic assignment of the ancient yeast strain, (5) the yeast admixture profile, and (6) the distribution of marker genes—we assume that this yeast is of ancient origin and has been involved in beer fermentation, although the mode of fermentation is unknown (i.e., bottom, top, or spontaneous fermentation).

DISCUSSION

Our interdisciplinary analyses of the samples have given detailed insight into the microbiome evolution and dietary habits and food

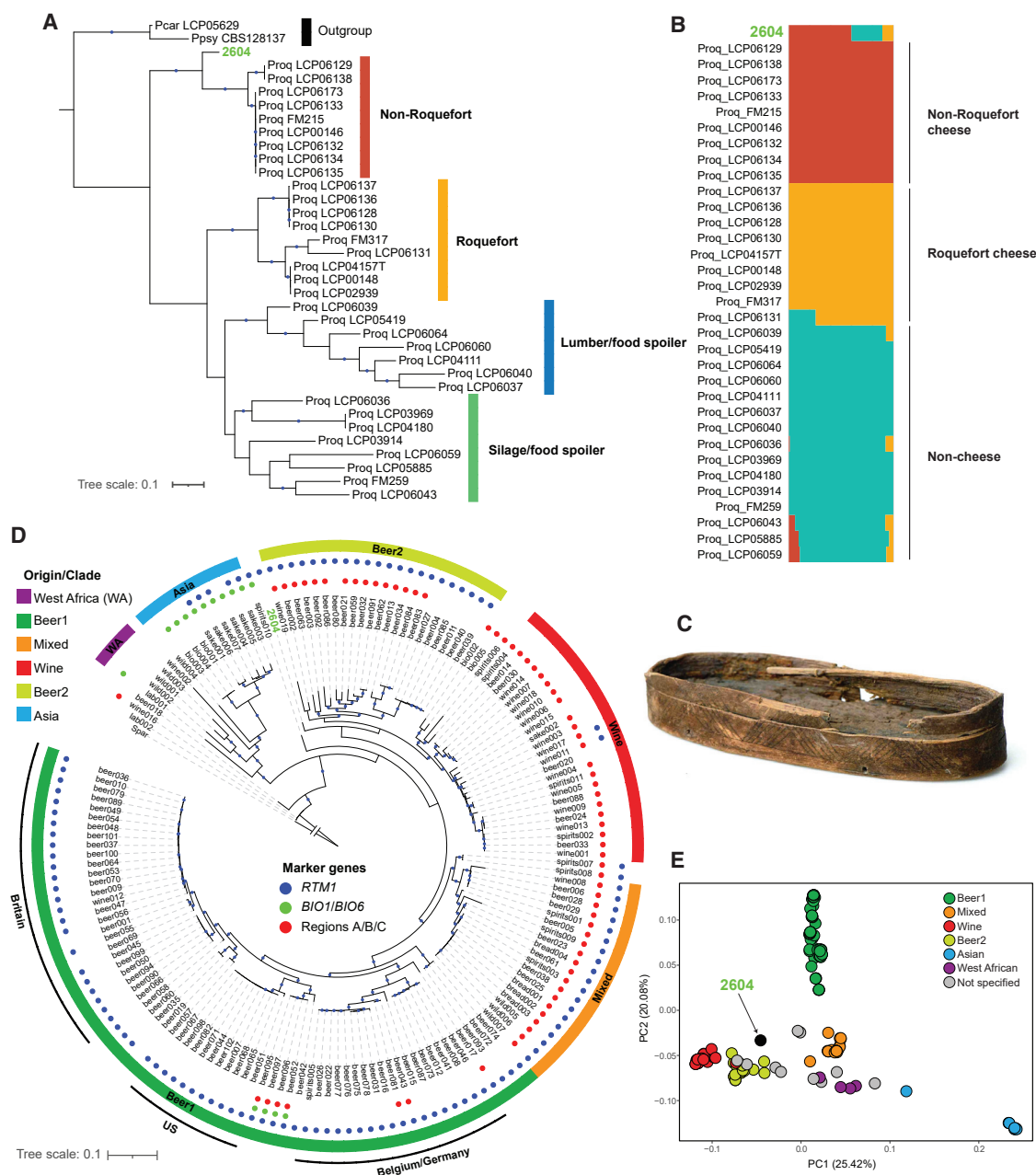


Figure 4. Genome-wide SNP analysis of ancient fungal “strains” versus modern industrial and wild/environmental strains

(A) Maximum likelihood (ML) phylogenetic analysis of the *Penicillium roqueforti* genome assembled from the sample 2604 in addition to other previously published *P. roqueforti* genomes.²⁶ A total number of 120,359 SNP positions were used for the analysis. *P. roqueforti* FM164 was used as a reference, while *P. carneum* and *P. psychrosexualis* were used as outgroups. The scale bar depicts 0.1 substitutions per residue. Colored strips indicate the *P. roqueforti* population as previously inferred.²⁶ For further information on the comparison dataset, please refer to [Data S1Q](#).

(B) Population structure analysis of *P. roqueforti* 2604 with the same previous dataset, considering 3 ancestries (K = 3 with lowest cross-validation error), based on 120,337 SNPs. The order of labels corresponds to the clustering in panel (A).

(C) Wooden containers that have been found among other archeological findings in the mines and assumed to be used as cheese strainers

(D) ML phylogenetic analysis of *Saccharomyces cerevisiae* genome assembled from the sample 2604 compared with other published *S. cerevisiae* genomes.³⁰ The dataset for the analysis included 375,629 SNPs. The *Saccharomyces paradoxus* CBS432 was used as an outgroup. The scale bar depicts 0.1 substitutions per residue. The colored strips indicate the clade/origin as reported previously.³⁰ The colored dots at the tree edges refer to the presence/absence of functional marker genes.³¹ Blue dots in (A) and (C) indicate bootstrap support >80% based on 1,000 bootstrap replicates.

(E) Principal component analysis based on 136,712 SNP, of the *S. cerevisiae* strains.

For additional information on the coverage and SNP density, DNA damage, and ADMIXTURE, please refer to [Figure S4](#). [Data S1](#) provides further details about the comparative datasets, mapping results, and functional marker analysis.

processing techniques of the Hallstatt miners over the past three millennia. Molecular and microscopic investigations revealed that the miner's diet was mainly composed of cereals, such as domesticated wheats (emmer and spelt), barley, common millets, and foxtail millets. This carbohydrate-rich diet was supplemented with proteins from broad beans and occasionally with fruits, nuts, or animal products. The food remains in the protohistoric sample with abundant entire fruits and seeds were less processed than those of the Baroque sample, which consisted of finely ground wheat. This suggests that the protohistoric miners consumed the cereals and legumes in a sort of gruel or porridge,³⁷ whereas miners in the 18th century AD ate their cereals in a more processed form, e.g., as a bread or biscuit.

In general, such carbohydrate-rich fibrous dietary components as observed in the Bronze Age and Iron Age samples are typical for traditional communities and are considered to be the main drivers of the non-Westernized microbiome structure.^{38,39} Consistent with this observation, our analysis showed that the Hallstatt paleofecal samples contain microbial features similar to gut microbiomes of modern non-Westernized populations (Figures 2A and 2B). Species identified in the samples, such as *Lactobacillus ruminis*, *Catenibacterium mitsuokai*, and *Prevotella copri*, were also found to be highly prevalent in present-day individuals with a more traditional lifestyle (Figure 2C). Furthermore, our paleofecal samples were rich in the *P. copri* complex (Figure 2D), including the four clades that are nearly ubiquitous and co-present in non-Westernized populations.⁹ Of particular interest, *P. copri* members have been shown to be associated with the digestion of complex carbohydrates,^{9,40–42} which are the major component of an unprocessed fibrous plant diet. Finding paleofeces highly resembling that of non-Westernized individuals, in terms of microbiome structure, supports previous observations.^{10,12} It also adds weight to the hypothesis that the modern industrialized human gut microbiome has diverged from an ancestral state, probably due to modern lifestyle, diet, or medical advances. Interestingly, this non-Westernized microbiome structure has been observed in all four paleofeces dating from the Bronze Age to the Baroque period, which would indicate quite a recent change in the gut community. However, to spot the critical time points when this shift in the human gut microbiome began requires more ancient samples spanning a wider time range; of particular interest would be samples from the past two or three centuries, when major dietary and medical changes occurred. Overall, our results support the theory that the shift from traditional to an industrial Westernized lifestyle might be the driving force for changing the human gut microbiome from its ancestral state.^{9,43–45}

In one of the Iron Age samples (2604), the molecular analyses indicated consumption of fermented food and beverages. The fungal analysis revealed a high prevalence of *Penicillium roqueforti* and *Saccharomyces cerevisiae*, which are nowadays involved in fermenting blue cheeses and alcoholic beverages, respectively, with clear signs of domestication. Following rapidly in the wake of ruminant animal domestication (mainly cattle, sheep, and goat), cheese production represents one of the oldest and widespread food preservation techniques developed by humans.⁴⁶ The oldest evidence for milk use dates back to the Neolithic in the Fertile Crescent yet provides only indirect evidence for fermentation.⁴⁷ The oldest reported chemical evidence for processing of milk into fermented products (i.e., kefir dairy) is

dated back to the Early Bronze Age in Western China.⁴⁸ Other indications, including actual preserved pieces of cheese, whey strainers, and recipes for cheese production, were found in Northern Europe, the Middle and Near East, and the Mediterranean basin.^{49–52} Here, we report evidence for the domestication of the fungus *Penicillium roqueforti* in the course of food processing in the 1st millennium BC that would likely produce a cheese resembling a blue cheese (non-Roquefort cheese clade, in Figure 4A). To our knowledge, this represents the earliest known evidence for directed cheese ripening and affinage in Europe, adding a crucial aspect to an emerging picture of highly sophisticated culinary traditions in European protohistory.³⁷ Importantly, the production of blue cheese today involves a surface application of dry salt; therefore, it is characterized by a high salt content of up to 7.5% (w/w).⁵³ The cheese curd could have been collected, desiccated, and inoculated with the fungi in wooden cheese containers like the ones excavated in the Hallstatt mines (Figure 4C). The presence of *P. roqueforti* indicates a major step in ruminant milk processing from fresh to ripened cheese, which could have offered, in addition to new flavors, several advantages to the Hallstatt miners including longer storage (i.e., months) and less lactose content in the fermented dairy product.⁵⁴ The reduced lactose content may have helped the ancient minors to better digest milk products, living in a time when lactose persistence frequencies only started to rise in Europe.⁵⁵ The presence of salt as well as the constant temperature (8°C) and humidity inside the Hallstatt mine workings represent ideal conditions for blue cheese production, following the current cheese production standards.⁵⁶ It is noteworthy that the early discovery of the Roquefort cheese was linked to Roquefort-sur-Soulzon caves in France, which maintain a temperature of 10°C and ~90% humidity over the year. With such conditions protecting the cheese from desiccation, these caves have been used exclusively for centuries for ripening and aging of the “Roquefort” cheese.^{57,58}

Indications for the production of fermented alcoholic beverages in protohistory are abundant, albeit frequently ambiguous,⁵⁹ and can be found in the Near East, Middle East, Far East, and Europe.^{60–67} Evidence for the production of grape wine in Europe and viniculture in the Near East dates back to the 6th and 7th millennium BC.⁶⁸ Such evidence was mainly based on chemical residue analysis or archaeobotanical analysis or indicated in ancient inscriptions. Recently scientists claimed that they were able to revive an ancient yeast strain from Egyptian potteries and used it to ferment beer.⁶⁹ Here we were able to reconstruct >90% of the *S. cerevisiae* genome from an Iron Age-dated paleofecal sample. We used different molecular analysis at the genome level to infer the possible routes of domestication for this yeast. Our results suggested it was used in beer fermentation. Together with the results of the dietary analysis that showed presence of different fermentable cereals, e.g., wheat, barley, and millets, we can envisage how the fermentation was carried out.

It might be assumed that the fermentation was carried out in a spontaneous manner—i.e., by adding water to wort and allowing the fermentation process to take place by the wild air-borne yeasts or the constitutional microbiota of the used cereals.⁷⁰ We do not see, however, indications for other yeasts species, such as *Brettanomyces bruxellensis*, that co-occur in spontaneously fermented beers.⁷¹ In addition, we see clear indications of

domestication and continuous supply of new admixture components to this yeast, which might suggest that fermentation vessels were repeatedly used for this purpose or the inoculation of the fermentation batches has been done by back-slopping (i.e., inoculation of new fermentation batches with portions of previous batches).⁷² Albeit varied evidence for beer production in protohistoric Europe exists,^{59,65–67} these beers could not be preserved for longer time periods and would have had to be consumed rapidly after production,⁶⁶ which also presupposes that the beer would have had to be produced either in Hallstatt itself or in the very near surroundings.

Considering the constant temperature of 8°C inside the Hallstatt mines, it might be expected that this yeast was used for production of lager-like beer, when fermentation is carried out at low temperatures (also known as bottom-fermentation) and results in a beer that can be stored for longer time periods.⁷³ Historically, however, the bottom-fermentation was most likely developed after the year 1553, when the Duke of Bavaria Albrecht V forbade brewing during summer months.⁷⁴ Additionally, Gonçalves and colleagues demonstrated that *Saccharomyces pastorianus* strains, which are hybrids of *S. cerevisiae* and another *Saccharomyces* species and are used for production of lager beers, belong to the main beer clade (Figure 4D).³¹ Therefore, we postulate that the beer produced at that time is similar to what would nowadays be known as pale beer, produced mainly by top-fermenting *S. cerevisiae* strains.

Paleofeces material displays an archaeological information source that provides insights into the diet and gut microbiome composition of ancestors. Here, we had access to four paleofeces samples from the Hallstatt salt mines dating from the Bronze Age to the Baroque period. The constant low annual temperature and high salt concentrations inside the mine preserved both plant macro-remains and biomolecules (DNA and protein) in the paleofeces. We demonstrate the indispensable complementarity of using microscopic and molecular approaches in resolving the paleofecal dietary residual components and to reconstruct the ancient gut microbiome. Furthermore, we extended our paleofeces microbiome analysis to focusing on key microbes that are involved in food processing, which opens new avenues in understanding fermentation history. In the future, additional samples from different time points will provide a more fine-scaled diachronic picture, which may help us to understand the role of dietary changes in shaping our gut microbiome and how much this was further influenced by modern lifestyles or medical advances recently introduced through industrialization and Westernization.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS

- Paleofeces samples, radiocarbon dating
- Microscopic analysis of the paleofeces
- Molecular analysis of the paleofeces
- Proteomic analysis

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2021.09.031>.

ACKNOWLEDGMENTS

We acknowledge the following funding sources: Programma Ricerca Budget prestazioni Eurac 2017 of the Province of Bolzano, Italy, and the South Tyrolean grant legge 14 (F.M., M.S.S., S.Z., and A.Z.). Additional support was provided by the European Regional Development Fund 2014-2020_CALL-FESR 2017 Research and Innovation_Autonomous Province of Bolzano South Tyrol_Project: FESR1078-MummyLabs. The authors thank the Department of Innovation, Research and University of the Autonomous Province of Bozen/Bolzano for covering the Open Access publication costs. We thank Dr. John Wilson (ProtiFi, USA) for helpful discussions regarding proteomics sample preparation. This work was in addition supported by the European Research Council grant ERC-STG Project MetaPG (N.S.); the US National Institutes of Health, National Institute for General Medical Sciences under grant no. GM087221 and the Office of the Director 1S10OD026936; and the US National Science Foundation award 1920268 (R.L.M.). We would like to thank Eva-Maria Geigl and the two anonymous reviewers for their insightful comments that helped to improve the manuscript.

AUTHOR CONTRIBUTIONS

F.M., K.O., N.S., A.Z., H.R., and K.K. conceived the investigation. F.M., A.S., R.L.M., K.O., N.S., A.Z., H.R., and K.K. designed experiments. K.O., A.S., F.M., A.Z., H.R., and K.K. were involved in the sampling campaign. F.M., A.S., U.K., M.R.H., and K.O. conducted the experiments. F.M., M.S.S., K.D.H., A.T., A.S., S.Z., A.B.-M., P.M., J.C.-K., W.R., U.K., S.R.M., M.R.H., O.R.-S., T.R., and K.O. performed analyses. F.M. and M.S.S. wrote the manuscript with contributions from K.D.H., A.T., A.B.-M., J.C.-K., M.R.H., O.R.-S., T.R., R.L.M., K.O., N.S., A.Z., H.R., and K.K.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 27, 2021

Revised: August 16, 2021

Accepted: September 14, 2021

Published: October 13, 2021

REFERENCES

1. Fry, G. (1985). Analysis of fecal material. In *The Analysis of Prehistoric Diets*, R.J. Gilbert, and J. Mielke, eds. (Academic Press), pp. 127–154.
2. Shillito, L.-M., Blong, J.C., Green, E.J., and van Asperen, E.N. (2020). The what, how and why of archaeological coprolite analysis. *Earth Sci. Rev.* 207, 103196.
3. Gilbert, M.T.P., Jenkins, D.L., Götherstrom, A., Naveran, N., Sanchez, J.J., Hofreiter, M., Thomsen, P.F., Binladen, J., Higham, T.F.G., Yohe, R.M., 2nd, et al. (2008). DNA from pre-Clovis human coprolites in Oregon, North America. *Science* 320, 786–789.
4. Maixner, F., Turaev, D., Cazenave-Gassiot, A., Janko, M., Krause-Kyora, B., Hoopmann, M.R., Kusebauch, U., Sartain, M., Guerriero, G., O'Sullivan, N., et al. (2018). The iceman's last meal consisted of fat, wild meat, and cereals. *Curr. Biol.* 28, 2348–2355.e9.

5. Poinar, H.N., Kuch, M., Sobolik, K.D., Barnes, I., Stankiewicz, A.B., Kuder, T., Spaulding, W.G., Bryant, V.M., Cooper, A., and Pääbo, S. (2001). A molecular analysis of dietary diversity for three archaic Native Americans. *Proc. Natl. Acad. Sci. USA* **98**, 4317–4322.
6. Mitchell, P.D. (2017). Human parasites in the Roman World: health consequences of conquering an empire. *Parasitology* **144**, 48–58.
7. Reinhard, K.J., Ferreira, L.F., Bouchet, F., Sianto, L., Dutra, J.M.F., Iniguez, A., Leles, D., Le Bailly, M., Fugassa, M., Pucu, E., and Araújo, A. (2013). Food, parasites, and epidemiological transitions: a broad perspective. *Int. J. Paleopathol.* **3**, 150–157.
8. Maixner, F., Krause-Kyora, B., Turaev, D., Herbig, A., Hoopmann, M.R., Hallows, J.L., Kusebauch, U., Vigl, E.E., Malferteiner, P., Megraud, F., et al. (2016). The 5300-year-old *Helicobacter pylori* genome of the Iceman. *Science* **351**, 162–165.
9. Tett, A., Huang, K.D., Asnicar, F., Fehner-Peach, H., Pasolli, E., Karcher, N., Armanini, F., Manghi, P., Bonham, K., Zolfo, M., et al. (2019). The *Prevotella copri* complex comprises four distinct clades underrepresented in Westernized populations. *Cell Host Microbe* **26**, 666–679.e7.
10. Tito, R.Y., Knights, D., Metcalf, J., Obregon-Tito, A.J., Cleeland, L., Najjar, F., Roe, B., Reinhard, K., Sobolik, K., Belknap, S., et al. (2012). Insights from characterizing extinct human gut microbiomes. *PLoS ONE* **7**, e51146.
11. Borry, M., Cordova, B., Perri, A., Wibowo, M., Prasad Honap, T., Ko, J., Yu, J., Britton, K., Girdland-Flink, L., Power, R.C., et al. (2020). CoproID predicts the source of coprolites and paleofeces using microbiome composition and host DNA content. *PeerJ* **8**, e9001.
12. Wibowo, M.C., Yang, Z., Borry, M., Hübner, A., Huang, K.D., Tierney, B.T., Zimmerman, S., Barajas-Olmos, F., Contreras-Cubas, C., Garcia-Ortiz, H., et al. (2021). Reconstruction of ancient microbial genomes from the human gut. *Nature* **594**, 234–239.
13. Harding, A. (2013). *Salt in Prehistoric Europe* (Sidestone Press).
14. Reschreiter, H., and Kowarik, K. (2019). Bronze Age mining in Hallstatt. A new picture of everyday life in the salt mines and beyond. *Archaeologia Austriaca* **103**, 99–136.
15. Grabner, M., Wächter, E., Nicolussi, K., Bolka, M., Sormaz, T., Steier, P., Wild, E.M., Barth, F.E., Kern, A., Rudorfer, J., et al. (2021). Prehistoric salt mining in Hallstatt, Austria. New chronologies out of small wooden fragments. *Dendrochronologia* **66**, 125814.
16. Kowarik, K. (2019). Mining and landscape: synthesis. In *Hallstätter Beziehungsgeschichten. Wirtschaftsstrukturen und Umfeldbeziehungen der bronze- und ältereisenzeitlichen Salzbergbaue von Hallstatt/OÖ.*, K. Kowarik, ed. (Oberösterreichisches Landesmuseum), pp. 245–266.
17. Festi, D., Brandner, D., Grabner, M., Knierzinger, W., Reschreiter, H., and Kowarik, K. (2021). 3500 years of environmental sustainability in the large-scale alpine mining district of Hallstatt, Austria. *Journal of Archaeological Science: Reports* **35**, 102670.
18. Wochenberichte (1723). Hofschreiberamt Hallstatt. <https://www.landesarchiv-ooe.at/>.
19. Brewster, R., Tamburini, F.B., Asiimwe, E., Oduaran, O., Hazelhurst, S., and Bhatt, A.S. (2019). Surveying gut microbiome research in Africans: toward improved diversity and representation. *Trends Microbiol.* **27**, 824–835.
20. Key, F.M., Posth, C., Krause, J., Herbig, A., and Bos, K.I. (2017). Mining metagenomic data sets for ancient DNA: recommended protocols for authentication. *Trends Genet.* **33**, 508–520.
21. Beghini, F., McIver, L.J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A.M., et al. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* **10**, e65088.
22. Grant, W.D. (2015). *Halococcus*. In *Bergey's Manual of Systematics of Archaea and Bacteria*, M.E. Trujillo, S. Dedysh, P. DeVos, B. Hedlund, P. Kämpfer, F.A. Rainey, and W.B. Whitman, eds. (John Wiley and Sons).
23. García, S., Vidal, J.E., Heredia, N., and Juneja, V.K. (2019). *Clostridium perfringens*. In *Food Microbiology: Fundamentals and Frontiers*, 5th Edition, M.P. Doyle, F. Diez-Gonzalez, and C. Hill, eds. (American Society of Microbiology).
24. Voidarou, C., Bezirtzoglou, E., Alexopoulos, A., Plessas, S., Stefanis, C., Papadopoulos, I., Vavias, S., Stavropoulou, E., Fotou, K., Tzora, A., and Skoufos, I. (2011). Occurrence of *Clostridium perfringens* from different cultivated soils. *Anaerobe* **17**, 320–324.
25. Pasolli, E., De Filippis, F., Mauriello, I.E., Cumbo, F., Walsh, A.M., Leech, J., Cotter, P.D., Segata, N., and Ercolini, D. (2020). Large-scale genome-wide analysis links lactic acid bacteria from food with the gut microbiome. *Nat. Commun.* **11**, 2610.
26. Dumas, E., Feurtey, A., Rodríguez de la Vega, R.C., Le Prieur, S., Snirc, A., Coton, M., Thierry, A., Coton, E., Le Piver, M., Roueyre, D., et al. (2020). Independent domestication events in the blue-cheese fungus *Penicillium roqueforti*. *Mol. Ecol.* **29**, 2639–2660.
27. Ropars, J., Rodríguez de la Vega, R.C., López-Villavicencio, M., Gouzy, J., Sallet, E., Dumas, É., Lacoste, S., Debuchy, R., Dupont, J., Branca, A., and Giraud, T. (2015). Adaptive horizontal gene transfers between multiple cheese-associated fungi. *Curr. Biol.* **25**, 2562–2569.
28. Cheeseman, K., Ropars, J., Renault, P., Dupont, J., Gouzy, J., Branca, A., Abraham, A.-L., Ceppi, M., Conseiller, E., Debuchy, R., et al. (2014). Multiple recent horizontal transfers of a large genomic region in cheese making fungi. *Nat. Commun.* **5**, 2876.
29. Ropars, J., Didot, E., Rodríguez de la Vega, R.C., Bennetot, B., Coton, M., Poirier, E., Coton, E., Snirc, A., Le Prieur, S., and Giraud, T. (2020). Domestication of the emblematic white cheese-making fungus *Penicillium camemberti* and its diversification into two varieties. *Curr. Biol.* **30**, 4441–4453.e4.
30. Gallone, B., Steensels, J., Prah, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., et al. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* **166**, 1397–1410.e16.
31. Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., Hutzler, M., Gonçalves, P., and Sampaio, J.P. (2016). Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts. *Curr. Biol.* **26**, 2750–2761.
32. Pontes, A., Hutzler, M., Brito, P.H., and Sampaio, J.P. (2020). Revisiting the taxonomic synonyms and populations of *Saccharomyces cerevisiae*—phylogeny, phenotypes, ecology and domestication. *Microorganisms* **8**, 903.
33. Marsit, S., Leducq, J.-B., Durand, É., Marchant, A., Filteau, M., and Landry, C.R. (2017). Evolutionary biology through the lens of budding yeast comparative genomics. *Nat. Rev. Genet.* **18**, 581–598.
34. Ness, F., and Aigle, M. (1995). RTM1: a member of a new family of telomeric repeated genes in yeast. *Genetics* **140**, 945–956.
35. Wu, H., Ito, K., and Shimoi, H. (2005). Identification and characterization of a novel biotin biosynthesis gene in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **71**, 6845–6855.
36. Novo, M., Bigey, F., Beyne, E., Galeote, V., Gavory, F., Mallet, S., Cambon, B., Legras, J.-L., Wincker, P., Casaregola, S., and Dequin, S. (2009). Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc. Natl. Acad. Sci. USA* **106**, 16333–16338.
37. Heiss, A.G., Jakobitsch, T., Wiesinger, S., and Trebsche, P. (2021). Dig out, dig in! Plant-based diet at the Late Bronze Age copper production site of Prigglitz-Gasteil (Lower Austria) and the relevance of processed foodstuffs for the supply of Alpine Bronze Age miners. *PLoS ONE* **16**, e0248287.
38. Makki, K., Deehan, E.C., Walter, J., and Bäckhed, F. (2018). The impact of dietary fiber on gut microbiota in host health and disease. *Cell Host Microbe* **23**, 705–715.
39. Crittenden, A.N., and Schnorr, S.L. (2017). Current views on hunter-gatherer nutrition and the evolution of the human diet. *Am. J. Phys. Anthropol.* **162** (Suppl 63), 84–109.

40. De Filippis, F., Pasolli, E., Tett, A., Tarallo, S., Naccarati, A., De Angelis, M., Neviani, E., Cocolin, L., Gobbetti, M., Segata, N., and Ercolini, D. (2019). Distinct genetic and functional traits of human intestinal *Prevotella copri* strains are associated with different habitual diets. *Cell Host Microbe* 25, 444–453.e3.
41. Gálvez, E.J.C., Iljazovic, A., Amend, L., Lesker, T.R., Renault, T., Thiemann, S., Hao, L., Roy, U., Gronow, A., Charpentier, E., and Strowig, T. (2020). Distinct polysaccharide utilization determines interspecies competition between intestinal *Prevotella* spp. *Cell Host Microbe* 28, 838–852.e6.
42. Fehlnner-Peach, H., Magnabosco, C., Raghavan, V., Scher, J.U., Tett, A., Cox, L.M., Gottsegen, C., Watters, A., Wiltshire-Gordon, J.D., Segata, N., et al. (2019). Distinct polysaccharide utilization profiles of human intestinal *Prevotella copri* isolates. *Cell Host Microbe* 26, 680–690.e5.
43. Blaser, M.J. (2017). The theory of disappearing microbiota and the epidemics of chronic diseases. *Nat. Rev. Immunol.* 17, 461–463.
44. Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., et al. (2019). Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* 176, 649–662.e20.
45. Sonnenburg, E.D., and Sonnenburg, J.L. (2019). The ancestral and industrialized gut microbiota and implications for human health. *Nat. Rev. Microbiol.* 17, 383–390.
46. Fox, P.F., Guinee, T.P., Cogan, T.M., and McSweeney, P.L. (2017). Cheese: historical aspects. *Fundamentals of Cheese Science* (Springer), pp. 1–10.
47. Evershed, R.P., Payne, S., Sherratt, A.G., Copley, M.S., Coolidge, J., Urem-Kotsu, D., Kotsakis, K., Ozdoğan, M., Ozdoğan, A.E., Nieuwenhuys, O., et al. (2008). Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature* 455, 528–531.
48. Yang, Y., Shevchenko, A., Knaust, A., Abuduresule, I., Li, W., Hu, X., Wang, C., and Shevchenko, A. (2014). Proteomics evidence for kefir dairy in Early Bronze Age China. *J. Archaeol. Sci.* 45, 178–186.
49. Bottéro, J. (1985). The cuisine of ancient Mesopotamia. *The Biblical Archaeologist* 48, 36–47.
50. Greco, E., El-Aguizy, O., Ali, M.F., Foti, S., Cunsolo, V., Saletti, R., and Ciliberto, E. (2018). Proteomic analyses on an ancient Egyptian cheese and biomolecular evidence of brucellosis. *Anal. Chem.* 90, 9673–9676.
51. McClure, S.B., Magill, C., Podrug, E., Moore, A.M.T., Harper, T.K., Culleton, B.J., Kennett, D.J., and Freeman, K.H. (2018). Fatty acid specific $\delta^{13}\text{C}$ values reveal earliest Mediterranean cheese production 7,200 years ago. *PLoS ONE* 13, e0202807.
52. Salque, M., Bogucki, P.J., Pyzel, J., Sobkowiak-Tabaka, I., Grygiel, R., Szmyt, M., and Evershed, R.P. (2013). Earliest evidence for cheese making in the sixth millennium BC in northern Europe. *Nature* 493, 522–525.
53. Cakmakci, S., Gundogdu, E., Hayaloglu, A.A., Dagdemir, E., Gurses, M., Cetin, B., and Tahmas-Kahyaoglu, D. (2012). Chemical and microbiological status and volatile profiles of mouldy C ivil cheese, a T urkish mould-ripened variety. *Int. J. Food Sci. Technol.* 47, 2405–2412.
54. Monti, L., Pelizzola, V., Povolito, M., Fontana, S., and Contarini, G. (2019). Study on the sugar content of blue-veined “Gorgonzola” PDO cheese. *Int. Dairy J.* 95, 1–5.
55. Burger, J., Link, V., Blöcher, J., Schulz, A., Sell, C., Pochon, Z., Diekmann, Y., Žegarac, A., Hofmanová, Z., Winkelbach, L., et al. (2020). Low prevalence of lactase persistence in Bronze Age Europe indicates ongoing strong selection over the last 3,000 years. *Curr. Biol.* 30, 4307–4315.e13.
56. Cantor, M.D., van den Tempel, T., Hansen, T.K., and Ardö, Y. (2017). Blue cheese. In *Cheese*, P.L.H. McSweeney, P.F. Fox, P.D. Cotter, and D.W. Everett, eds. (Elsevier), pp. 929–954.
57. Capozzi, V., and Spano, G. (2011). Food microbial biodiversity and “microbes of protected origin”. *Front. Microbiol.* 2, 237.
58. Desmasures, N. (2014). Mold-ripened varieties. In *Encyclopedia of Food Microbiology*, C. Batt, ed. (Elsevier), pp. 409–416.
59. Heiss, A.G., Azorín, M.B., Antolín, F., Kubiak-Martens, L., Marinova, E., Arendt, E.K., Biliaderis, C.G., Kretschmer, H., Lazaridou, A., Stika, H.-P., et al. (2020). Mashies to mashies, crust to crust. Presenting a novel microstructural marker for malting in the archaeological record. *PLoS ONE* 15, e0231696.
60. Zarnkow, M., Otto, A., and Einwag, B. (2011). Interdisciplinary investigations into the brewing technology of the ancient Near East and the potential of the cold mashing process. In *Liquid Bread: Beer and Brewing in Cross-cultural Perspective*, W. Schiefenhövel, and H. Macbeth, eds. (Berghahn Books), pp. 47–54.
61. Liu, L., Wang, J., Rosenberg, D., Zhao, H., Lengyel, G., and Nadel, D. (2018). Fermented beverage and food storage in 13,000 y-old stone mortars at Raqefet Cave, Israel: Investigating Natufian ritual feasting. *Journal of Archaeological Science: Reports* 21, 783–793.
62. Wang, J., Liu, L., Ball, T., Yu, L., Li, Y., and Xing, F. (2016). Revealing a 5,000-y-old beer recipe in China. *Proc. Natl. Acad. Sci. USA* 113, 6444–6448.
63. Farag, M.A., Elmassry, M.M., Baba, M., and Friedman, R. (2019). Revealing the constituents of Egypt’s oldest beer using infrared and mass spectrometry. *Sci. Rep.* 9, 16199.
64. Damerow, P. (2012). Sumerian beer: the origins of brewing technology in ancient Mesopotamia. *Cuneiform Digital Library Journal* 2, 1–20.
65. Stika, H.P. (2011). Beer in prehistoric Europe. In *Liquid Bread: Beer and Brewing in Cross-Cultural Perspective*, W. Schiefenhövel, and H. Macbeth, eds. (Berghahn Books), pp. 55–62.
66. Rosenstock, E., and Scheibner, A. (2017). Fermentierter Brei und vergorenes Malz: Bier in der Vorgeschichte Südwestasiens und Europas. *Mitteilungen der Anthropologischen Gesellschaft Wien* 147, 31–62.
67. Guerra-Doce, E. (2015). The origins of inebriation: archaeological evidence of the consumption of fermented beverages and drugs in prehistoric Eurasia. *J. Archaeol. Method Theory* 22, 751–782.
68. McGovern, P., Jalabadze, M., Batiuk, S., Callahan, M.P., Smith, K.E., Hall, G.R., Kvavadze, E., Maghradze, D., Rusishvili, N., Bouby, L., et al. (2017). Early neolithic wine of Georgia in the South Caucasus. *Proc. Natl. Acad. Sci. USA* 114, E10309–E10318.
69. Aouizerat, T., Gutman, I., Paz, Y., Maeir, A.M., Gadot, Y., Gelman, D., Sztienberg, A., Drori, E., Pinkus, A., Schoemann, M., et al. (2019). Isolation and characterization of live yeast cells from ancient vessels as a tool in bio-archaeology. *MBio* 10, e00388, e19.
70. Basso, R.F., Alcarde, A.R., and Portugal, C.B. (2016). Could non-*Saccharomyces* yeasts contribute on innovative brewing fermentations? *Food Res. Int.* 86, 112–120.
71. Spitaels, F., Wieme, A.D., Janssens, M., Aerts, M., Daniel, H.-M., Van Landschoot, A., De Vuyst, L., and Vandamme, P. (2014). The microbial diversity of traditional spontaneously fermented lambic beer. *PLoS ONE* 9, e95384.
72. Spitaels, F., Wieme, A.D., Snauwaert, I., De Vuyst, L., and Vandamme, P. (2017). Microbial ecology of traditional beer fermentations. In *Brewing Microbiology: Current Research, Omics and Microbial Ecology*, N.A. Bokulich, and C.W. Bamforth, eds. (Caister Academic Press), pp. 179–196.
73. Baker, E.P., Peris, D., Moriarty, R.V., Li, X.C., Fay, J.C., and Hittinger, C.T. (2019). Mitochondrial DNA and temperature tolerance in lager yeasts. *Sci. Adv.* 5, v1869.
74. Dornbusch, H.D. (1998). *Prost!: The Story of German Beer* (Brewers Publications).
75. NCBI Resource Coordinators (2017). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 45 (D1), D12–D17.
76. Banchi, E., Ametrano, C.G., Greco, S., Stanković, D., Muggia, L., and Pallavicini, A. (2020). PLANITS: a curated sequence reference dataset for plant ITS DNA metabarcoding. *Database (Oxford)* 2020, baz155.

77. Nilsson, R.H., Larsson, K.-H., Taylor, A.F.S., Bengtsson-Palme, J., Jeppesen, T.S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F.O., Tedersoo, L., et al. (2019). The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* 47 (D1), D259–D264.
78. Ratnasingham, S., and Hebert, P.D. (2007). bold: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Mol. Ecol. Notes* 7, 355–364.
79. Buchfink, B., Reuter, K., and Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* 18, 366–368.
80. Huson, D.H., Beier, S., Flade, I., Górski, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.J., and Tappu, R. (2016). MEGAN Community Edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* 12, e1004957.
81. Ondov, B.D., Bergman, N.H., and Phillippy, A.M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12, 385.
82. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595.
83. Neukamm, J., Peltzer, A., and Nieselt, K. (2021). DamageProfiler: fast damage pattern calculation for ancient DNA. *Bioinformatics* btab190. <https://doi.org/10.1093/bioinformatics/btab190>.
84. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* 16, 224.
85. Skoglund, P., Storå, J., Götherström, A., and Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* 40, 4477–4482.
86. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
87. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* 44 (W1), W58–63.
88. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
89. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* 15, 356.
90. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066.
91. Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., et al. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371.
92. Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.
93. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
94. Cook, D.E., and Andersen, E.C. (2017). VCF-kit: assorted utilities for the variant call format. *Bioinformatics* 33, 1581–1582.
95. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
96. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49 (W1), W293–W296.
97. Alexander, D.H., and Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* 12, 246.
98. Kessner, D., Chambers, M., Burke, R., Agus, D., and Mallick, P. (2008). ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* 24, 2534–2536.
99. Deutsch, E.W., Mendoza, L., Shteynberg, D., Slagel, J., Sun, Z., and Moritz, R.L. (2015). Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics Clin. Appl.* 9, 745–754.
100. Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74, 5383–5392.
101. Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 75, 4646–4658.
102. Hoopmann, M.R., Winget, J.M., Mendoza, L., and Moritz, R.L. (2018). StPeter: seamless label-free quantification with the trans-proteomic pipeline. *J. Proteome Res.* 17, 1314–1320.
103. Eng, J.K., Jahan, T.A., and Hoopmann, M.R. (2013). Comet: an open-source MS/MS sequence database search tool. *Proteomics* 13, 22–24.
104. Gurdeep Singh, R., Tanca, A., Palomba, A., Van der Jeugt, F., Verschaffelt, P., Uzzau, S., Martens, L., Dawyndt, P., and Mesuere, B. (2019). Unipept 4.0: functional analysis of metaproteome data. *J. Proteome Res.* 18, 606–615.
105. Stuiver, M., and Polach, H.A. (1977). Discussion reporting of ¹⁴C data. *Radiocarbon* 19, 355–363.
106. Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R.L., Friedrich, M., et al. (2020). The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* 62, 725–757.
107. Pearsall, D.M. (2015). *Paleoethnobotany: A Handbook of Procedures* (Left Coast Press).
108. Cappers, R.T., Bekker, R.M., and Jans, J.E. (2006). *Digitale zadenatlas van Nederland* (Barkhuis Publishing).
109. Jacomet, S. (2006). *Bestimmung von Getreidefunden aus archäologischen Ausgrabungen* (Universität Basel).
110. Neef, R., Cappers, R.T., and Bekker, R.M. (2012). *Digital Atlas of Economic Plants in Archaeology* (Barkhuis).
111. Hohmann, B., Deutschmann, F., and Gassner, G. (1989). *Mikroskopische Untersuchungen pflanzlicher Lebensmittel: mit einem Kapitel über die mikroskopische Untersuchung der wichtigsten als Futtermittel verwendeten tierischen Produkte* (Gustav Fischer Verlag).
112. Tang, J.N., Zeng, Z.G., Wang, H.N., Yang, T., Zhang, P.J., Li, Y.L., Zhang, A.Y., Fan, W.Q., Zhang, Y., Yang, X., et al. (2008). An effective method for isolation of DNA from pig faeces and comparison of five different methods. *J. Microbiol. Methods* 75, 432–436.
113. Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 40, e3.
114. Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338, 222–226.
115. Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2015). The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* 43, D670–D681.
116. Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 23, 147.
117. Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R., and Kelley, S.T. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* 8, 761–763.
118. Marchione, D.M., Ilieva, I., Devins, K., Sharpe, D., Pappin, D.J., Garcia, B.A., Wilson, J.P., and Wojcik, J.B. (2020). HYPERsol: high-quality data

- from archival FFPE tissue for clinical proteomics. *J. Proteome Res.* **19**, 973–983.
119. Martens, L., Chambers, M., Sturm, M., Kessner, D., Levander, F., Shofstahl, J., Tang, W.H., Römpp, A., Neumann, S., Pizarro, A.D., et al. (2011). mzML—a community standard for mass spectrometry data. *Mol. Cell. Proteomics* **10**, 000133.
 120. Moosa, J.M., Guan, S., Moran, M.F., and Ma, B. (2020). Repeat-preserving decoy database for false discovery rate estimation in peptide identification. *J. Proteome Res.* **19**, 1029–1036.
 121. Shteynberg, D., Deutsch, E.W., Lam, H., Eng, J.K., Sun, Z., Tasman, N., Mendoza, L., Moritz, R.L., Aebersold, R., and Nesvizhskii, A.I. (2011). iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol. Cell. Proteomics* **10**, 007690.
 122. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47** (D1), D442–D450.

STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|--|--------------------|---|
| Biological samples | | |
| Paleofeces sample from the Bronze Age | This study | 2610 |
| Paleofeces sample from the Iron Age | This study | 2604 |
| Paleofeces sample from the Iron Age | This study | 2611 |
| Paleofeces sample from the Baroque time | This study | 2612 |
| Chemicals, peptides, and recombinant proteins | | |
| Sodium phosphate | Sigma-Aldrich | Cat #342483 |
| Trypsin gold | Promega | Cat # V528A |
| Critical commercial assays | | |
| S-Trap | ProtiFi | Cat #K02-mini-10 |
| Deposited data | | |
| Paleofeces metagenomic shotgun datasets | This study | ENA: PRJEB44507 |
| NCBI-nr database | 75 | https://www.ncbi.nlm.nih.gov/protein/ |
| NCBI-nt database | 75 | https://www.ncbi.nlm.nih.gov/nucleotide/ |
| PlantITS database | 76 | https://github.com/apallavicini/PLANIITS |
| Chloroplast genome database | N/A | https://www.ncbi.nlm.nih.gov/genome/organelle |
| Fungal ITS database | 77 | https://unite.ut.ee/ |
| BOLD system databases | 78 | https://v3.boldsystems.org/ |
| Software and algorithms | | |
| MetaPhlAn | 21 | https://github.com/biobakery/MetaPhlAn/wiki/MetaPhlAn-3.0 |
| DIAMOND | 79 | https://github.com/bbuchfink/diamond |
| MEGAN6 | 80 | https://www.wsi.uni-tuebingen.de/lehrstuehle/algorithms-in-bioinformatics/software/megan6 |
| Krona tool | 81 | https://github.com/marbl/Krona/wiki |
| BWA | 82 | http://bio-bwa.sourceforge.net/ |
| DeDup tool | N/A | https://github.com/apeltzer/DeDup |
| DamageProfiler | 83 | https://damageprofiler.readthedocs.io/en/latest/index.html |
| Schmutzi | 84 | https://github.com/grenaud/schmutzi |
| Molecular sex determination | 85 | https://github.com/pontussk/ry_compute |
| SAMtools | 86 | http://samtools.github.io/ |
| HaploGrep | 87 | https://haplogrep.i-med.ac.at/ |
| HUMAnN | 21 | https://github.com/biobakery/humann |
| python package scikit-bio | N/A | http://scikit-bio.org/ |
| bowtie2 | 88 | http://bowtie-bio.sourceforge.net/bowtie2/ |
| FastQ Screen | N/A | https://github.com/StevenWingett/FastQ-Screen |
| ANGSD tool | 89 | https://github.com/ANGSD/angsd |
| MAFFT | 90 | https://mafft.cbrc.jp/alignment/software/ |
| ARB software package | 91 | http://www.arb-home.de/ |
| PhyML | 92 | https://github.com/stephaneguindon/phyml |
| BLASTn | 93 | N/A |
| Picard tools | N/A | https://broadinstitute.github.io/picard/ |
| GATK4 | N/A | https://gatk.broadinstitute.org/hc |

(Continued on next page)

Continued

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------------------|--------|---|
| vcf-kit | 94 | https://vcf-kit.readthedocs.io/en/latest/ |
| RAxML | 95 | https://github.com/stamatak/standard-RAxML |
| Interactive Tree of Life (iTOL) | 96 | https://itol.embl.de/ |
| ADMIXTURE | 97 | http://dalexander.github.io/admixture/download.html |
| R-Studio | N/A | https://www.rstudio.com/ |
| ProteoWizard | 98 | https://proteowizard.sourceforge.io/ |
| Trans-Proteomic Pipeline | 99 | http://tools.proteomecenter.org/software.php |
| PeptideProphet | 100 | http://peptideprophet.sourceforge.net/ |
| ProteinProphet | 101 | http://proteinprophet.sourceforge.net/ |
| StPeter | 102 | http://tools.proteomecenter.org/wiki/index.php?title=Software:StPeter |
| Comet | 103 | http://comet-ms.sourceforge.net/ |
| Unipept 4.0 | 104 | https://unipept.ugent.be/ |

RESOURCE AVAILABILITY

Lead contact

Further information on materials, datasets, and protocols should be directed to and will be fulfilled by the lead contact, Frank Maixner (frank.maixner@eurac.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Sequencing data are available at the European Nucleotide Archive (ENA) under ENA: PRJEB44507. Mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository, PRIDE: PXD027613. All code used in this study and other previously published genomic data is available at the sources referenced in the [key resources table](#).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

In this study, we subjected four paleofeces samples to in-depth microscopic and molecular analysis. The paleofeces material stems from Bronze Age and Iron Age mine workings in the Hallstatt salt mines in Upper Austria ([Figures 1A–1C](#)). See [Data S1A](#) for additional details to the samples used in this study.

METHOD DETAILS

Paleofeces samples, radiocarbon dating

Three paleofeces (2610, 2604, 2611) were recovered in 1983, 1989 (2604) and 2003 (2610) through wet-sieving of larger blocks of protohistoric production debris excavated in the mine workings. One additional paleofeces (2612) was sampled in 2019 *in situ* at the site Edlersbergwerk-oben with sterile sampling tools. This sample was not subjected to wet-sieving. All four samples (using approx. 500 mg each) were subjected to radiocarbon dating at the Curt-Engelhorn-Centre for Archaeometry in Mannheim, Germany ([Figure 1C](#); [Data S1B](#)). The ^{14}C content was measured with an AMS system of the MICADAS type. The isotopic ratios $^{14}\text{C}/^{12}\text{C}$ and $^{13}\text{C}/^{12}\text{C}$ of the samples, the calibration standards (oxalic acid II), the blanks and the control standards were measured simultaneously in the AMS system. The determined ^{14}C ages are based on $^{13}\text{C} = -25\text{‰}$ ¹⁰⁵ and were calibrated to calendar ages with the dataset INTCAL20 ¹⁰⁶ and the software SwissCal (L.Wacker, ETH-Zurich). The remaining material has been used for microscopic and molecular analyses.

Microscopic analysis of the paleofeces

Before chemical treatment of the paleofeces the surface of each sample was stripped off to avoid contamination. Then rehydration in a 0.5% solution of trisodium phosphate for 72 h. ¹⁰⁷ After olfactory and visual testing, the liquid was screened through 500, 250, 125, and 63 μm steel meshes, and the outwash was kept for further microfossil studies. Plant macro-remains were picked out of the fractionated residues in the steel meshes and then identified and quantified under a stereo microscope with magnifications up x63. The

identification of the plant remains was conducted with identification keys^{108–111} and the reference collection of the Department of Botany, Innsbruck University, was consulted for comparative purposes.

Molecular analysis of the paleofeces

DNA extraction, library preparation and sequencing

The molecular analysis of the paleofeces samples was conducted at the ancient DNA laboratory of the EURAC Institute for Mummy Studies in Bolzano, Italy. Sample documentation, sample preparation and DNA extraction were performed in a dedicated pre-PCR area following the strict procedures required for studies of ancient DNA: use of protective clothing, UV-light exposure of the equipment and bleach sterilization of surfaces, use of PCR workstations and filtered pipette tips. The DNA was extracted from the paleofeces samples (200 mg) using a chloroform-based DNA extraction method according to the protocol of Tang and colleagues with minor modifications.¹¹² To test whether the wet-sieving step during the classical archaeological excavation results in an “wash-out” effect of DNA from the paleofeces, we re-hydrated additional 200 mg material of two samples (2611, 2612) in 1 ml of 0.5N Na₃PO₄ (trisodium phosphate)-buffer for 18 h at RT.¹ After re-hydration, the samples were centrifuged (5000 g, 5 min) and both the supernatant (2611RW, 2612RW) and the pelleted paleofeces (2611R, 2612R) were subjected to DNA extraction as described above. From all DNA extracts double-indexed libraries were generated for the sequencing runs with a modified protocol for Illumina multiplex sequencing.^{113,114} Libraries were first shallow sequenced on an Illumina MiSeq platform and then deep sequenced on Illumina HiSeq2500 and HiSeqX10 platforms using 101–base pair (HiSeq2500) and 151–base pair (MiSeq, HiSeqX) paired-end sequencing kits. For details to the metagenomic shotgun datasets please refer to [Data S1C](#). Data are available from the European Nucleotide Archive under accession number PRJEB44507.

Pre-processing of the sequence data, general taxonomic overview, human DNA analysis

The paired Illumina reads were first quality-checked and processed (adaptor removal and read merging) as previously described in Maixner et al.⁴ Initially we tested for the above-described possible DNA wash-out effect by assigning taxonomically the microbial reads using MetaPhlAn 3.0²¹ in the shallow sequenced MiSeq datasets of the untreated samples (2611, 2612), the washed and pelleted paleofeces (2611R, 2612R), and the supernatant (2611RW, 2612RW). For all subsequent analyses, we used the combined sequencing data available for each sample ([Data S1C](#)). First, we assessed a general taxonomic profile of the sequencing reads using DIAMOND (v2.0.7) blastx search⁷⁹ against the NCBI nr database (Release 237, April 2020). The DIAMOND tables were converted to rma6 (blast2rma tool) format (–minPercentIdentity 97), imported into MEGAN6 software,⁸⁰ and subsequently visualized using the Krona tool.⁸¹ Next, we assessed the endogenous human DNA content in the paleofeces samples by aligning the sequence reads against the human genome (build hg19)¹¹⁵ and the human mtDNA reference genome (rCRS)¹¹⁶ using BWA⁸² with a seed length of 1,000. The minimum mapping and base quality were both 30. To deduplicate the mapped reads, we used the DeDup tool (<https://github.com/apeltzer/DeDup>). For details to the mapping results please refer to [Data S1D](#). The resulting bam files were used to check for characteristic aDNA nucleotide misincorporation frequency patterns and for the fragment length distribution using the DamageProfiler tool.⁸³ Mitochondrial human contamination rates were assessed using Schmutzi.⁸⁴ The sex of the mapped human reads was assigned using a Maximum likelihood method, based on the karyotype frequency of X and Y chromosomal reads.⁸⁵ Variants in the mitochondrial genome were called using SAMtools mpileup and bcftools⁸⁶ with stringent filtering options (quality > 30). The haplogroup was identified by submitting the variant calling file to the HaploGrep website.⁸⁷

Comparing paleofeces microbiome structure to contemporary metagenomic datasets

To compare the paleofeces microbiome structure with contemporary individuals', we downloaded 9,368 publicly available shotgun metagenomes representing modern-day human populations (n = 9,207), and, as a control, environmental soil sample (n = 161). Each metagenome was characterized by source (human or soil), and in the case of human samples, by body-site, and whether from a Westernized or non-Westernized population. The term “non-Westernized” describes a population that follows a traditional non-urbanized lifestyle encompassing factors such as diet, hygiene, and with no or limited access to medical healthcare and pharmaceuticals (e.g., antibiotics) as previously described.^{9,19,44} Afterward, we performed profiling of the microbial composition of each metagenomic sample including the four paleofeces samples from this study using MetaPhlAn 3.0²¹ using default settings. To survey paleofeces microbiome members in different populations, microbial abundance profiles of four paleofeces were first merged using `merge_metaphlan_tables.py`, a utility script in MetaPhlAn 3.0. The merged abundance table was then visualized in the form of a hierarchically-clustered heatmap using `hclust2` (<https://github.com/SegataLab/hclust2>) with parameters `–ftop 15–f_dist_f braycurtis–s_dist_f braycurtis`, with a result of displaying only 15 most abundant microbial species in the four paleofeces samples. We further analyzed the prevalence of these top 15 ancient-sample enriched species in the context of global populations using a subset (n = 8,968) of shotgun metagenomes that are stool samples from healthy adult individuals characterized by non-Westernization ([Data S1H](#)). A species was counted as presence in a sample if the relative abundance was above 0.01%. We repeated the prevalence analysis for all species detected in the four paleofeces samples to take low-abundance members into account as well ([Data S1I](#) and [S1J](#)). Next, we randomly selected 662 metagenomes (132 oral cavity and 530 stool samples) from all publicly available human metagenomes, along with 161 soil metagenomes, for profiling of the metabolic pathways using HUMAnN 3.0²¹ with default settings. The Python script `humann_renorm_table.py` as part of the pipeline was used to normalize the metabolic pathways to relative abundance, followed by converting individual profiles into a merged abundance table using the utility script `humann_join_tables.py`. For the same selected metagenomic samples, the individual profiles of microbial composition generated by MetaPhlAn 3.0 as described above were similarly merged into a single abundance table using a python script `merge_metaphlan_tables.py`. Subsequently, the merged abundance tables were used for calculating Bray-Curtis's distance with logarithmic transformation based on which a Principal

Coordinates Analysis was performed using python package scikit-bio (version 0.5.6; <http://scikit-bio.org/>). Lastly, to predict the source of microbial communities in the paleofeces samples SourceTracker2¹¹⁷ was used as described in a recent study,¹² comparing the paleofeces samples' abundance profiles with those of different sources (a subset of samples randomly selected from the aforementioned Principal Coordinates Analysis: 16 non-Westernized oral samples, 20 non-Westernized stool samples, 26 soil samples, 10 Westernized oral samples and 30 Westernized stool samples). For details to the comparative datasets please refer to [Data S1E](#) and [S1G](#).

Molecular characterization of ancient diet

The merged reads were searched for homology, using BLAST search, against different specific DNA barcoding databases ([Data S1N](#)). Specifically, the databases of plant Internal transcribed spacers,⁷⁶ RuBisCo-large subunit and Maturase K genes (rbcL/matK, <https://v3.boldsystems.org/>), and chloroplast genomes (<https://www.ncbi.nlm.nih.gov/genome/organelle>) were used to identify the plant dietary components. The BOLD database of the cytochrome c oxidase subunit I gene (COI, <https://v3.boldsystems.org/>) was used to specifically target the animal diet and intestinal parasites. The UNITE database of the fungal ITS gene database (<https://unite.ut.ee/>) was used to identify potential food processing fungi. Finally, to target all the eukaryotic dietary and diet-related components, the reads were aligned against the currently available full mitochondrial and chloroplast genomes of the NCBI database⁷⁵ using BWA⁸² with default parameters. Subsequently, to obtain a taxonomic overview of the aligned reads we performed a sequence similarity search using BLASTn⁹³ with default parameters against the complete NCBI-nt database.⁷⁵

Both the resulting BLAST and the previously created DIAMOND tables were converted to rma6 (blast2rma tool) format and imported into MEGAN6 software. The read counts at genus- and species-level were exported as csv files and imported into R-Studio software for further analysis.

To distinguish the true hits from the false hits, we applied two filters as follows: i) For the BLAST and DIAMOND tables, we only considered hits of $\geq 90\%$ identity and $\geq 90\%$ coverage of each query read; then ii) We considered the sample positive for a certain component only if it shows incidence in majority of databases of that component, e.g., the sample 2604 was considered positive for *Bos taurus*, because it contained true hits for mitochondrial-, COI-, and nr-databases; and finally iii) for the plant dietary components, we used the plantITS database as a primary filter, since it is highly curated and recently updated. So, we initially filtered out all the plantITS negative taxa then considered the same majority rule.

Finally, we used transformed sums of counts of the genera and species that passes the aforementioned filters into normalized log base 2 (\log_{10} reads million⁻¹) and plotted them as dot-plots proportional to their abundances. The dendrogram to the left of the dot-plot represent the phylogeny based on NCBI taxonomy (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>).

Selected identified taxa were further subjected to in-depth analysis by mapping the quality-filtered sequence reads against organellar (mitochondrial and chloroplast) and nuclear genomes using bowtie2 (v1.2.1.1) and the parameter “end-to-end”⁸⁸ (for details to the references used please refer to the [Data S1O](#)). In two cases where the previously described molecular dietary analysis resulted in an assignment down to the genus level (*Triticum* spp., *Ascaris* spp.) we assessed an appropriate reference sequence by mapping the sequence reads against selected chloroplast and mitochondrial genomes using BWA⁸² (with default parameters) implemented in the program FastQ Screen (<https://github.com/StevenWingett/FastQ-Screen>), and selected as reference the species that belongs to one of these two genera and has the most specific hits. After the bowtie2 alignment against the reference sequences the mapped reads were deduplicated and checked for damage patterns as described above for the human DNA. For details to the mapping results please refer to [Data S1O](#) and [S1P](#). Organellar sequences that showed higher than 70% sequence coverage were further subjected to phylogenetic analysis as previously detailed.⁴ In brief, a consensus FASTA sequence was generated using the ANGSD tool⁸⁹ and together with other comparative datasets subjected to a multiple alignment using the MAFFT multiple sequence alignment program.⁹⁰ Phylogenetic analysis was performed by applying distance-matrix, maximum-parsimony, and maximum-likelihood methods implemented in the ARB software package:⁹¹ neighbor-joining (using the Jukes-Cantor algorithm for nucleic acid correction with 1000 bootstrap iterations), DNA parsimony (PHYLIP version 3.66 with 100 bootstrap iterations), and DNA maximum-likelihood [PhyML⁹² with the HKY substitution model]. The number of informative nucleotide positions used for the phylogenetic analysis and the bootstrap support is indicated in the respective figure captions.

Genome-level analysis of ancient fungi - Variant calling

The sample 2604 showed high prevalence of *Penicillium roqueforti* and *Saccharomyces cerevisiae* DNA allowing further genome-level comparative analysis with modern datasets ([Data S1Q](#) and [S1R](#)). As comparative datasets, we included the recently published genomic data of 34 *P. roqueforti*²⁶ as well as 157 *S. cerevisiae*.³⁰

We mapped the quality-filtered reads against the reference genomes of *Penicillium roqueforti* FM164 and *Saccharomyces cerevisiae* S288c, independently, using bowtie2 (v1.2.1.1) and the parameter “end-to-end.”⁸⁸ Using samtools,⁸⁶ the mapping qualities of <30 were filtered out and the bam files were sorted and indexed. The read duplicates were marked and removed using the DeDup tool (<https://github.com/apeltzer/DeDup/>) and new read-groups were assigned using the “AddOrReplaceReadGroups” of Picard tools (<https://broadinstitute.github.io/picard/>). Then, we used the Genome Analysis Toolkit 4 (GATK4) (<https://gatk.broadinstitute.org/hc>) to call genome variants, following the best practices workflow as follows: i) Variants within each of the single bam files were called using the tool “HaplotypeCaller”; ii) The resulting VCF files were combined into a single file using the tool “CombineGVCFs”; iii) The combined VCF file was then genotyped using the tool “GenotypeGVCFs”; iv) InDels were filtered out and only single nucleotide polymorphisms (SNPs) were kept using the tool “SelectVariants”; and finally v) the VCF file was filtered to include SNPs with minor allele frequency (MAF) of ≥ 0.05 -0.1 and $\leq 10\%$ missing data. A total of 120359 and 375629 SNPs were resolved for *P. roqueforti* and

S. cerevisiae, respectively. The total number of positions with alternative alleles to the reference genome accounted for 0.97% in case of *P. roqueforti* and 10.05% in case of *S. cerevisiae*. More than 99% of the positions with alternative alleles have a major allele frequency of > 90%, which minimize the likelihood of having chimeric alignments. In addition, we applied majority rule selection for the positions with multiple alleles, thereby analyzing only the major allele type.

Phylogenetic- and population structure analysis

To analyze the phylogenetic relationship of the strains, we used the vcf-kit to convert the SNP datasets (vcf files) to fastA alignment format.⁹⁴ Then we used the Fasta2phylip.pl script to create PHYLIP alignment files (<https://mullinslab.microbiol.washington.edu/>), which were then subjected to maximum likelihood (ML) phylogenetic analysis using RAXML (v8.2.12), following generalized time reversible substitution model (GTR) and GAMMA model of rate heterogeneity.⁹⁵ Best-scoring ML trees were searched after 1000 bootstrap replicates and visualized and annotated using the interactive tree of life, iTOL tool (<https://itol.embl.de>).

To infer the degree of admixture and the number of populations in the analyzed datasets (120,337 SNPs for *P. roqueforti* and 136,712 *S. cerevisiae*), we carried out unsupervised population structure analysis using ADMIXTURE (v1.3.0), testing K in the range of 3 to 22.⁹⁷ The best K values were determined based on the lowest cross-validation error (cv) and maximum likelihood.

Functional marker genes

To infer the potential functions of the modern and ancient *Saccharomyces cerevisiae* yeast strains based on their genome sequence, we searched for the presence of functional marker genes (*RTM1*, *BIO1/BIO6*, and Regions A/B/C) that were recently described by Cheeseman et al.²⁸ The *RTM1* gene is responsible for conferring resistance to the toxicity of molasses, and therefore it is assumed to be positively selected in beer yeast strains.³⁴ While the genes *BIO1* and *BIO6*, which are involved in *de novo* synthesis of biotin, are highly selected in sake fermenting yeasts.³⁵ The regions A, B, and C are horizontally transferred genomic regions from other yeast genera, e.g., *Kluyveromyces*, *Pichia*, and *Zygosaccharomyces*.³⁶ These regions contain 39 genes distributed over 3 different chromosomes (Data S1S) and are assumed to play a role in wine fermentation. To search for these marker genes, we used bowtie2 to map the quality filtered reads, of our sample as well as the recently published *S. cerevisiae* comparative genomic data,³⁰ against the individual genes. Then, we considered a gene as present if it is covered $\geq 90\%$ with 3x depth (Data S1T). For phylogenetic tree annotation (Figure 4D), the regions A/B/C were considered present if any of the genes was positive, while *BIO1/BIO6* were considered positive if both genes were positive.

Proteomic analysis

Sample preparation

Paleofeces samples (ID 2610, 2604, 2611, 2612) and a blank control sample to highlight any contamination were resuspended in 800 μ L 5% (w/v) sodium dodecyl sulfate (SDS), 50 mM triethylammonium bicarbonate (TEAB, pH 8.5) and disrupted at 4°C using three 2.8 mm ceramic beads (QIAGEN, USA) and a Precellys 24 homogenizer (Bertin Corp, USA) at 6500 rpm for 30 s. Samples were transferred to Eppendorf tubes and centrifuged at 4,000 rpm for 7 min, the supernatant removed and centrifuged twice at 13,000 g for 10 min. The supernatant was subjected to the S-Trap mini spin column digestion protocol (ProtiFi, USA).¹¹⁸ Briefly, samples were acidified with 12% phosphoric acid in water (pH ≤ 1). Binding/wash buffer (100 mM TEAB (final) in 90% methanol) was added to each sample at 6.4 x the volume of each sample. Samples were vortexed and applied to the S-Trap column in ≤ 500 μ L aliquots until the entire sample was loaded (J. Wilson, personal communication). After each sample load the S-Trap column was centrifuged at 4,000 g for 30 s. Samples were washed to remove any residual SDS by adding 3x 400 μ L binding/wash buffer and centrifugation at 4,000 g for 30 s after each wash, and 1 min after the last wash. Proteins were digested on the S-Trap at 37°C for 13 h by adding 125 μ L digestion buffer containing trypsin gold (Promega, USA) in 125 mM NH_4HCO_3 . The S-Trap was rehydrated with 100 μ L 125 mM NH_4HCO_3 and peptides eluted with 80 μ L 125 mM NH_4HCO_3 , 80 μ L 0.2% (v/v) formic acid in water, and 80 μ L 50% (v/v) acetonitrile in water. Peptide eluates were pooled and dried under centrifugal evaporation (Savant, Thermo-Fisher Scientific, USA).

Liquid chromatography - mass spectrometry analysis (LC-MS/MS)

Samples were analyzed by high-resolution nano LC-MS/MS on an Orbitrap-Eclipse Tribrid mass spectrometer (Thermo-Fisher Scientific, USA) equipped with an Easy-nLC 1000 (Thermo-Fisher Scientific, USA). Peptides were resolubilized in 50 μ L 0.1% (v/v) formic acid in water. 5 μ L of each sample were loaded onto a 2 cm Acclaim PepMap 100 trap column (75 μ m ID, C18 3 μ m (dp), Thermo-Fisher Scientific, USA) and separated using a 50 cm C18 2 μ m (dp) Easy Spray column (ES903, Thermo-Fisher Scientific, USA) using a 120-min gradient. Mobile phase A consisted of 0.1% (v/v) formic acid in water, and mobile phase B consisted of 0.1% (v/v) formic acid in acetonitrile. The gradient used 5% to 22% mobile phase B over 105 min, followed by separation from 22% to 32% mobile phase B over the remaining 15 min. The column was washed and equilibrated at 95% and 5% mobile phase B, respectively, over 40 min following each run. The flow rate was set at 300 nL/min and the column was heated at 45°C. Mass spectra were acquired on the Orbitrap-Eclipse mass spectrometer using data dependent acquisition (DDA) with dynamic exclusion. The precursor scan range was from 375–1500 *m/z* at 120,000 resolution and 100% normalized target AGC (4e5) with a 50 ms maximum injection time. The duty cycle was set to acquire as many MS/MS spectra as possible over a three second period following each precursor scan. A 1.2 *m/z* selection window was used to acquire MS/MS spectra at 30,000 resolution, a normalized AGC target of 100% (5e4), 54 ms maximum injection time, and fragmented using HCD with a normalized collision energy of 30. Dynamic exclusion was set to 60 s, with peptide match set to preferred and isotope exclusion turned on. Charge exclusion was set to 1 and greater than 7.

Protein data analysis

The raw mass spectra files were converted to mzML¹¹⁹ format using msConvert from ProteoWizard⁹⁸ and analyzed using the Trans-Proteomic Pipeline (TPP v6.0.0-rc16 Noctilucent).⁹⁹ The analysis pipeline consisted of database searching with Comet (version 2018.01 rev. 4)¹⁰³ against a 17-species UniProt FASTA database (Data S2A) and shuffled decoy sequences.¹²⁰ The 17 species were selected based on the microscopic and metagenomics analyses described above. An equal number of decoy sequences were appended to the FASTA database for downstream statistical analysis and validation of any observed peptide sequences, as described below. For details to the comparative datasets used in this study and to the protein analysis results please refer to Data S2. Comet parameters included variable modifications of +15.994915 Da (Met, Cys, His, Tyr, Trp, and Phe), and +0.984016 (Asn and Gln), as these types of protein degradations are commonly found in aged samples. A precursor tolerance of 20 ppm was set, a fragment bin tolerance of 0.2 and fragment bin offset of 0. Semi-enzymatic cleavage with up to 3 missed cleavages was allowed, and an isotope error tolerance of 3. Peptide-spectrum matches (PSMs) were validated using PeptideProphet¹⁰⁰ and iProphet.¹²¹ Protein inference was performed using ProteinProphet.¹⁰¹ Relative quantitation of protein groups was performed using StPeter.¹⁰² Observed protein groups for each sample were derived from the set of protein groups validated with ProteinProphet at an estimated 1% false discovery rate (FDR). Those protein groups were then curated by removing any group that intersected with the set of protein groups identified in the blank control analysis. A list of observed peptides was created from the peptide sequences that were used to infer the final set of protein groups. The peptide list was then filtered to contain only those peptides with a probability below a 1% FDR threshold as determined from the iProphet analysis. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE¹²² partner repository with the dataset identifier PXD027613.

The identified peptides, evaluated by statistical analysis to 1% false discovery rate, were further taxonomically screened for homologous peptides and subsequently filtered for peptide sequences that can be unambiguously assigned to certain plants, animals, and fungi. Therefore, the peptide sequences of each sample were imported into Unipept 4.0 Desktop version,¹⁰⁴ and subjected to LCA assignment, and filtered out the following: i) peptides assigned to bacteria; ii) peptides assigned to multiple species; iii) single peptide incidences that uniquely infer a species presence (Data S2). Finally, functional annotation of the samples was carried out on different taxonomic levels to infer the molecular function of the obtained peptides.

QUANTIFICATION AND STATISTICAL ANALYSIS

Phylogenetic analyses were carried out based on the maximum-likelihood method using the tool Randomized Accelerated Maximum Likelihood (RAxML) with 1000 bootstrap replicates.⁹⁵ For the dietary analysis, majority rule-based selection was implemented using in-house R-script in R-Studio (<https://www.rstudio.com/>).