

# Pathogens and host immunity in the ancient human oral cavity

Christina Warinner<sup>1,2</sup>, João F Matias Rodrigues<sup>3,4</sup>, Rounak Vyas<sup>3,4</sup>, Christian Trachsel<sup>5</sup>, Natallia Shved<sup>1</sup>, Jonas Grossmann<sup>5</sup>, Anita Radini<sup>6,7</sup>, Y Hancock<sup>8</sup>, Raul Y Tito<sup>2</sup>, Sarah Fiddyment<sup>6</sup>, Camilla Speller<sup>6</sup>, Jessica Hendy<sup>6</sup>, Sophy Charlton<sup>6</sup>, Hans Ulrich Luder<sup>9</sup>, Domingo C Salazar-García<sup>10–12</sup>, Elisabeth Eppler<sup>13,14</sup>, Roger Seiler<sup>1</sup>, Lars H Hansen<sup>15,16</sup>, José Alfredo Samaniego Castruita<sup>17</sup>, Simon Barkow-Oesterreicher<sup>5</sup>, Kai Yik Teoh<sup>6</sup>, Christian D Kelstrup<sup>18</sup>, Jesper V Olsen<sup>18</sup>, Paolo Nanni<sup>5</sup>, Toshihisa Kawai<sup>19,20</sup>, Eske Willerslev<sup>17</sup>, Christian von Mering<sup>3,4</sup>, Cecil M Lewis Jr<sup>2</sup>, Matthew J Collins<sup>6</sup>, M Thomas P Gilbert<sup>17,21</sup>, Frank Rühli<sup>1,22</sup> & Enrico Cappellini<sup>17,22</sup>

**Calcified dental plaque (dental calculus) preserves for millennia and entraps biomolecules from all domains of life and viruses. We report the first, to our knowledge, high-resolution taxonomic and protein functional characterization of the ancient oral microbiome and demonstrate that the oral cavity has long served as a reservoir for bacteria implicated in both local and systemic disease. We characterize (i) the ancient oral microbiome in a diseased state, (ii) 40 opportunistic pathogens, (iii) ancient human-associated putative antibiotic resistance genes, (iv) a genome reconstruction of the periodontal pathogen *Tannerella forsythia*, (v) 239 bacterial and 43 human proteins, allowing confirmation of a long-term association between host immune factors, ‘red complex’ pathogens and periodontal disease, and (vi) DNA sequences matching dietary sources. Directly datable and nearly ubiquitous, dental calculus permits the simultaneous investigation of pathogen activity, host immunity and diet, thereby extending direct investigation of common diseases into the human evolutionary past.**

Unlike other human microbiomes, the oral microbiome will cause disease in a majority of people during their lifetime, suggesting that it is currently in a state of dysbiosis rather than symbiosis<sup>1,2</sup>. The human oral microbiome comprises more than 2,000 bacterial taxa, including a large number of opportunistic pathogens involved in periodontal, respiratory, cardiovascular and systemic diseases<sup>3–7</sup>. Dental calculus, a complex, calcified bacterial biofilm formed from dental plaque, saliva and gingival crevicular fluid<sup>8</sup>, is emerging as a potential substrate for the direct investigation of the evolution of the oral microbiome and associated measures of oral health and diet<sup>9,10</sup>. Recently, a DNA-based 16S rRNA phylotyping study identified the major bacterial phyla in dental calculus and argued for shifts in microbial diversity associated with the origins of agriculture and industrialization<sup>11</sup>, and, so far, five common oral bacteria have been identified in historic and prehistoric

dental calculus using targeted PCR<sup>12</sup>, quantitative PCR (qPCR)<sup>11</sup> and immunohistochemistry<sup>13</sup>. However, phylum-level community analysis and single-species targeted amplification are insufficient to characterize oral health and disease states, as this requires a deeper taxonomic and functional understanding of microbiome ecology<sup>14</sup>.

We present the first detailed analysis to our knowledge of ancient oral microbiome ecology and function at the genus and species levels, leading to a deeper understanding of recent evolution of the human oral microbiome. Focusing on the dental tissues of four adult human skeletons (G12, B17, B61 and B78) with evidence of mild to severe periodontal disease from the medieval monastic site of Dalheim, Germany (c. 950–1200 CE) (**Supplementary Fig. 1**), as well as modern dental calculus from nine individuals with known dental histories, we demonstrate for the first time, to our knowledge, that the human oral microbiome has long served as a

<sup>1</sup>Centre for Evolutionary Medicine, Institute of Anatomy, University of Zürich, Zürich, Switzerland. <sup>2</sup>Department of Anthropology, University of Oklahoma, Norman, Oklahoma, USA. <sup>3</sup>Institute of Molecular Life Sciences, University of Zürich, Zürich, Switzerland. <sup>4</sup>Swiss Institute of Bioinformatics, Lausanne, Switzerland. <sup>5</sup>Functional Genomics Center Zürich, University of Zürich/Swiss Federal Institute of Technology (ETH) Zürich, Zürich, Switzerland. <sup>6</sup>BioArCh, Department of Archaeology, University of York, York, UK. <sup>7</sup>University of Leicester Archaeological Services (ULAS), School of Archaeology and Ancient History, University of Leicester, Leicester, UK. <sup>8</sup>Department of Physics, University of York, York, UK. <sup>9</sup>Centre of Dental Medicine, Institute of Oral Biology, University of Zürich, Zürich, Switzerland. <sup>10</sup>Research Group on Plant Foods in Hominin Dietary Ecology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. <sup>11</sup>Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. <sup>12</sup>Department of Prehistory and Archaeology, University of Valencia, Valencia, Spain. <sup>13</sup>Research Group Neuro-Endocrine-Immune Interactions, Institute of Anatomy, University of Zürich, Zürich, Switzerland. <sup>14</sup>Zürich Center for Integrative Human Physiology, University of Zürich, Zürich, Switzerland. <sup>15</sup>Department of Biology, Microbiology, University of Copenhagen, Copenhagen, Denmark. <sup>16</sup>Department of Environmental Science, Aarhus Universitet, Roskilde, Denmark. <sup>17</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark. <sup>18</sup>Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. <sup>19</sup>Department of Immunology and Infectious Diseases, Forsyth Institute, Cambridge, Massachusetts, USA. <sup>20</sup>Department of Oral Medicine, Infection and Immunity, Harvard School of Dental Medicine, Harvard University, Boston, Massachusetts, USA. <sup>21</sup>Ancient DNA Laboratory, Murdoch University, Perth, Western Australia, Australia. <sup>22</sup>These authors jointly directed this work. Correspondence should be addressed to C.W. (twarinner@gmail.com) or E.C. (ecappellini@gmail.com).

Received 31 May 2013; accepted 3 February 2014; published online 23 February 2014; doi:10.1038/ng.2906

**Figure 1** Taxonomic and phylogenetic characterization of ancient dental calculus. **(a)** Relative proportions of bacterial, archaeal, eukaryotic and viral DNA in ancient calculus estimated from assembled whole-metagenome shotgun sequences of two individuals. **(b)** Phylogenetic tree of the 100 most abundant OTUs in ancient dental calculus samples from 4 pooled individuals. Evidence for the presence and abundance of each microbial OTU is represented by colored, size-scaled circles for each targeted 16S rRNA hypervariable region (V3, V5, V6), shotgun 16S rRNA sequences (S) and other genes (G) and proteins (P) assigned to that OTU. OTUs for which no reference genome exists or for which insufficient proteome data have been validated for inclusion in the protein search databases are marked with gray squares, as no hits could be matched to those OTUs. The tree scale bar indicates nucleotide substitutions per site. Relative phylum abundance (normalized mean of all genetic data generated from 16S rRNA sequences) is represented by a column chart showing the phyla represented in the top 100 OTUs (colored), remaining phyla (dark gray) and unidentified OTUs (light gray).

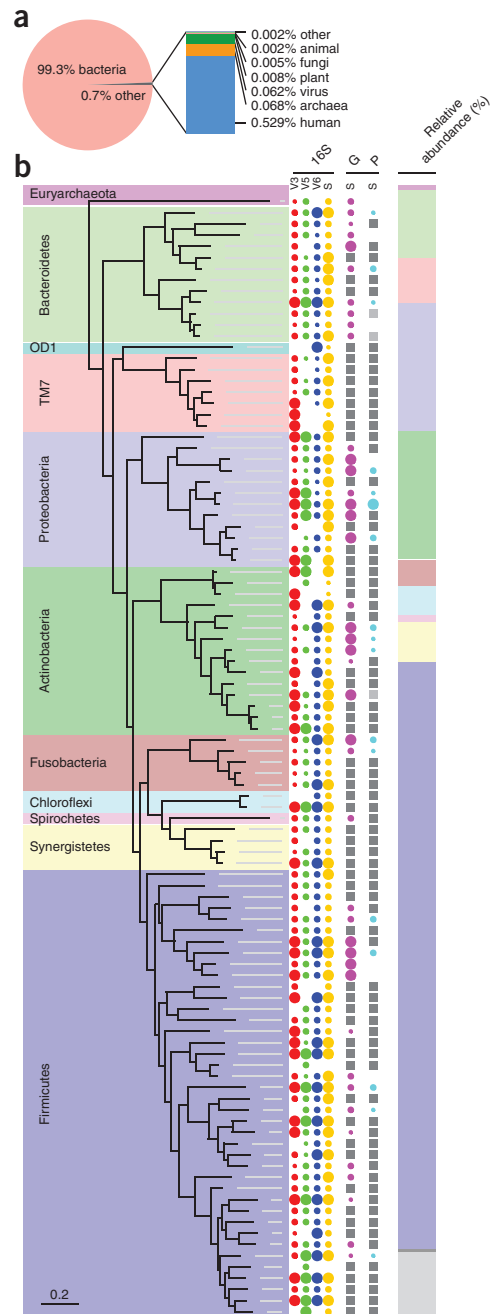
reservoir for a broad suite of opportunistic pathogens implicated in both local and systemic disease and harbored a diverse range of putative antibiotic resistance genes. We confirm the long-term role of host immune activity and red complex pathogen virulence in periodontal pathogenesis, despite major changes in lifestyle, hygiene and diet over the past millennium. We reconstruct the genome of a major periodontal pathogen and present the first evidence, to our knowledge, of dietary biomolecules to be recovered from ancient dental calculus. Finally, we further validate our findings by applying multiple microscopic, genetic and proteomic analyses in parallel, providing a systematic biomolecular evaluation of ancient dental calculus preservation, taphonomy and contamination.

## RESULTS

### The ancient oral microbiome

Applying shotgun DNA sequencing to dental calculus for the first time, we found that it was strongly dominated by bacterial DNA, with minor contributions from human, viral, dietary and fungal sources (Fig. 1a). Using both targeted and shotgun 16S rRNA sequences ( $n = 509,067$ ), we identified a total of 2,699 microbial operational taxonomic units (OTUs) in the ancient dental calculus, with the 100 most abundant taxa accounting for 86.6% of the total reads (Fig. 1b and Supplementary Fig. 2). One archaeal and nine bacterial phyla were dominant in ancient dental calculus (Supplementary Table 1): Firmicutes ( $49.5 \pm 10.6\%$ ), Actinobacteria ( $12.0 \pm 6.1\%$ ), Proteobacteria ( $11.5 \pm 8.6\%$ ), Bacteroidetes ( $6.6 \pm 3.6\%$ ), TM7 ( $4.6 \pm 4.0\%$ ), Synergistetes ( $3.3 \pm 2.6\%$ ), Chloroflexi ( $2.7 \pm 1.5\%$ ), Fusobacteria ( $2.1 \pm 1.8\%$ ), Spirochetes ( $0.6 \pm 0.3\%$ ) and Euryarchaeota ( $0.4 \pm 0.6\%$ ), all of which are also dominant in the human oral microbiome today<sup>4</sup>. Notably rare in ancient dental calculus was Acidobacteria, a ubiquitous and abundant bacterial phylum in soil<sup>15</sup>.

To address biases resulting from the sequencing approach and the 16S rRNA gene hypervariable region (V3, V5, V6) primer choice (Supplementary Fig. 3), we visualized evidence for each OTU separately for each targeted and shotgun 16S rRNA detection method, as well as for shotgun metagenomic and metaproteomic data (Fig. 1b). Most OTUs were detected using multiple methods. OTUs detected from targeted V3 and shotgun data generally showed good agreement, whereas V5 and V6 primers showed clear evidence of primer bias and OTU dropout. Shotgun metagenomic data showed excellent agreement with consensus 16S rRNA OTUs when reference genomes were available. Shotgun metaproteomic data also showed good agreement with the OTUs identified on the basis of genetic data, and agreement is expected to improve as protein databases grow to include more predicted proteins and epigenetic variants. Because ancient DNA and proteins undergo different taphonomic processes and have different contamination risks, the high degree of



phylogenetic consensus observed for data generated from independent extractions, using different methods and targeting different biomolecular types, demonstrates that an endogenous oral microbiome can be robustly and reliably recovered from ancient dental calculus.

### Carriage of specific pathogens

The normal human oral flora includes a large number of endogenous cariogenic, periodontal and other opportunistic pathogens. Although these taxa generally do not cause extraoral disease in healthy subjects, they nevertheless pose a serious risk for the elderly and immunocompromised<sup>16,17</sup> and are known to be involved in the etiology of chronic systemic diseases, including cardiovascular disease<sup>18</sup>. As the detection of particular species from metagenomic sequence data is an open area of research, we applied a conservative contig assembly and BLAST strategy and screened our results against the Pathosystems Resource

Integration Center (PATRIC) database<sup>19</sup> to identify 40 putative opportunistic pathogens in ancient dental calculus (Table 1), of which only 5 had been previously reported in ancient samples<sup>11–13</sup>. We also identified phage DNA sequences specific to particular bacteria (Table 1), including *Streptococcus mitis* phage SM1, which has been previously shown to mediate *S. mitis* attachment to platelets and to increase bacterial virulence in the endocardium<sup>20</sup>.

Both DNA and proteins from the periodontal pathogens *T. forsythia*, *Porphyromonas gingivalis* and *Treponema denticola* were particularly abundant in our ancient dental calculus samples, demonstrating that these so-called red complex bacteria<sup>21</sup> were strongly associated with periodontal disease during the medieval period, just as they are today, despite substantial changes in oral hygiene, diet and lifestyle. Additionally, all three of these pathogens were found at substantially higher frequency in our ancient dental calculus samples than in the Human Microbiome Project (HMP)<sup>3</sup> healthy cohort (Supplementary Fig. 4a–c and Supplementary Table 2), consistent with expectations for periodontal disease. We also identified several oral taxa (for example, *Aggregatibacter actinomycetemcomitans*, *Streptococcus mutans* and *S. mitis*) that have been shown to cause bacteremia and infective endocarditis<sup>7,18</sup>, demonstrating that the human oral microbiome has long harbored pathogens that contribute to risk of cardiovascular disease. Additional pathogens included those implicated in acute dental infections (for example, *Actinomyces odontolyticus*), caries (*S. mutans*) and opportunistic upper and lower respiratory illness (for example, *Streptococcus pneumoniae*, *Streptococcus pyogenes* and *Haemophilus influenzae*). Of interest, all ancient dental calculus samples were also found to contain disordered carbon (microcharcoal), a respiratory irritant. Two obligate human taxa, *Neisseria meningitidis* and *Neisseria gonorrhoeae*, causative agents of bacterial meningitis and gonorrhoea, respectively, were also observed. *N. meningitidis* and *N. gonorrhoeae* form a recently diverged pathogenic clade of *Neisseria*, a genus comprising many commensal species inhabiting the mucosa and dental surfaces of animals<sup>22</sup>, and both are prevalent members of the human oral microbiome. Genital *N. gonorrhoeae* strains can infect the pharynx and engage in genetic exchange with other *Neisseria* species;<sup>23</sup> however, oral strains are not known to cause genital infection. Oral *N. meningitidis* is a leading cause of bacterial meningitis, although disease susceptibility is determined by a combination of host genetics and strain virulence<sup>24</sup>. Finally, we observed two additional oral taxa present at substantially higher frequency in at least one ancient dental calculus sample compared to the HMP healthy cohort: *Filifactor alocis* and *Olsenella uli*

**Table 1 Putative pathogens identified from assembled metagenomic and metaproteomic sequences in ancient dental calculus**

Pathogens <sup>a</sup>	Genes (contigs)	Proteins (peptides)	Virulence	Drug resistance <sup>b</sup>	Plasmid	CTn or phage
<i>Actinomyces odontolyticus</i> <sup>c</sup>	3 (4)	3 (34)				
<i>Aggregatibacter actinomycetemcomitans</i>	50 (68)	0			+	+
<i>Campylobacter concisus</i>	10 (20)	0			+	
<i>Campylobacter curvus</i>	12 (11)	0				
<i>Campylobacter rectus</i> <sup>c</sup>	3 (9)	3 (15)	++			
<i>Campylobacter showae</i> <sup>c</sup>	3 (13)	1 (2)				
<i>Capnocytophaga gingivalis</i> <sup>c</sup>	2 (11)	3 (7)	+			
<i>Capnocytophaga ochracea</i>	938 (4,909)	0	+	+	+	+
<i>Capnocytophaga sputigena</i> <sup>c</sup>	2 (2)	0		+		
<i>Clostridium difficile</i> <sup>d,e</sup>	30 (76)	0		+		+
<i>Corynebacterium matruchotii</i> <sup>c</sup>	2 (15)	12 (89)				
<i>Eikenella corrodens</i> <sup>c</sup>	11 (38)	2 (11)			+	
<i>Fusobacterium nucleatum</i>	656 (1,525)	4 (21)	++	+	+	+
<i>Fusobacterium periodonticum</i> <sup>c</sup>	3 (6)	3 (8)	+			
<i>Gemella morbillorum</i> <sup>c</sup>	9 (38)	0				
<i>Gordonibacter pamela</i> <sup>d</sup>	3 (30)	0				
<i>Haemophilus influenzae</i>	19 (43)	1 (4)				+
<i>Histophilus somni</i> <sup>d,f</sup>	9 (18)	0				+
<i>Leptotrichia buccalis</i>	492 (1,104)	0	+	+	+	+
<i>Neisseria gonorrhoeae</i>	127 (250)	1 (2)	+		+	
<i>Neisseria meningitidis</i>	336 (821)	1 (2)	+	+		+
<i>Neisseria sicca</i> <sup>c</sup>	3 (8)	4 (35)				
<i>Neisseria subflava</i> <sup>c</sup>	4 (12)	0				
<i>Porphyromonas gingivalis</i>	802 (2,588)	7 (72)	++	+		+
<i>Rothia mucilaginosa</i>	24 (17)	1 (2)		+		
<i>Streptobacillus moniliformis</i> <sup>d,f</sup>	8 (23)	0				
<i>Streptococcus agalactiae</i>	7 (27)	0			+	+
<i>Streptococcus dysgalactiae</i> <sup>d</sup>	2 (8)	0				+
<i>Streptococcus equi</i> <sup>d,f</sup>	29 (101)	0	+			+
<i>Streptococcus gallolyticus</i> <sup>d,f</sup>	8 (11)	0				+
<i>Streptococcus gordonii</i>	882 (3,397)	1 (8)	+	+	+	+
<i>Streptococcus mitis</i>	88 (161)	1 (37)	++	+		+
<i>Streptococcus mutans</i>	21 (67)	0				
<i>Streptococcus pneumoniae</i>	144 (339)	1 (8)	+	+		+
<i>Streptococcus pyogenes</i>	14 (32)	1 (8)		+		+
<i>Streptococcus sanguinis</i>	850 (3,272)	1 (4)	+	+		
<i>Streptococcus suis</i> <sup>d,f</sup>	2 (3)	0	+			
<i>Tannerella forsythia</i>	1,099 (11,279)	10 (137)	++	+		+
<i>Treponema denticola</i>	917 (6,106)	3 (15)	++	+	+	+
<i>Veillonella parvula</i>	96 (109)	0		+		

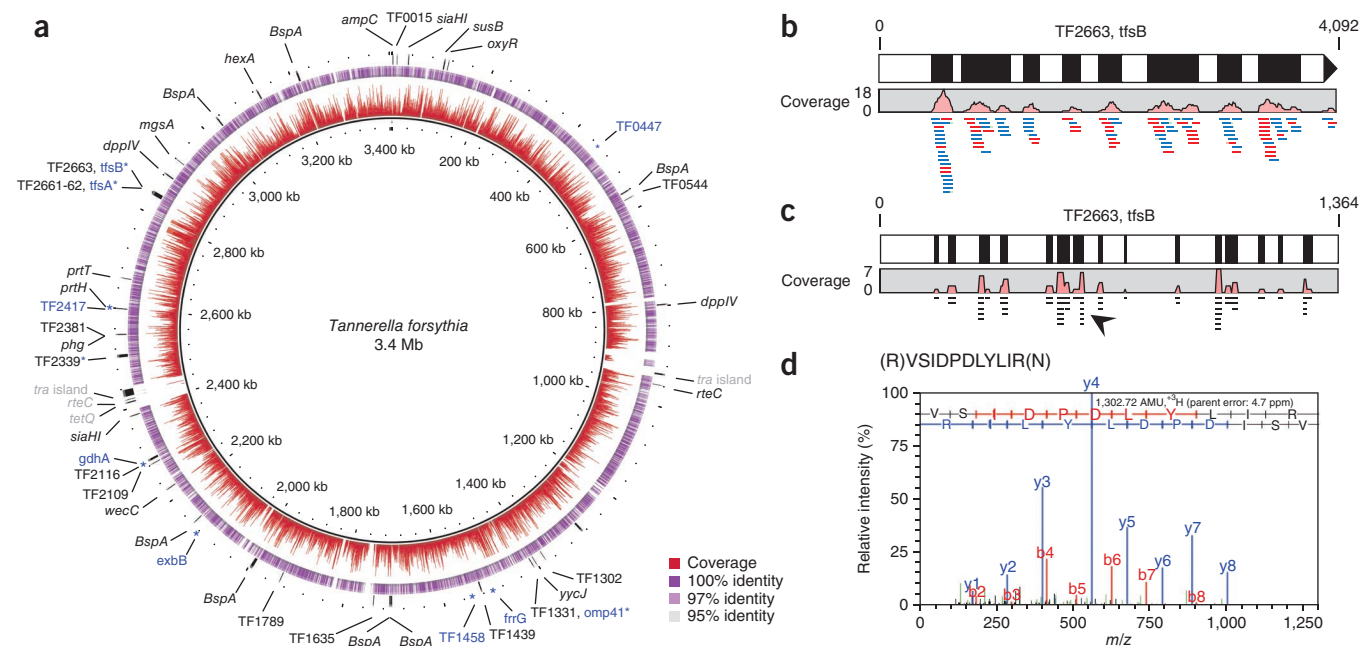
Metagenomic data from G12 and B61 and proteomic data from G12, B17, B61 and B78. +, gene(s) detected; ++, gene(s) and protein(s) detected.

<sup>a</sup>Includes only pathogens with valid entries in the PATRIC database. Only taxa represented by more than one DNA contig are shown. All pathogens are known inhabitants of the human oral cavity, as confirmed by cross-referencing with HMP data for supragingival dental plaque. <sup>b</sup>Putative function based on gene homology and NCBI annotation; functionality was not independently validated. <sup>c</sup>Reference genome sequencing and annotation incomplete. <sup>d</sup>Not a prevalent inhabitant of the oral cavity; not present in the Human Oral Microbiome Database (HOMD). <sup>e</sup>Tentative identification; sequences correspond almost exclusively to mobile genetic elements. <sup>f</sup>Putative zoonosis.

(Supplementary Fig. 4e,f). Although not classified as pathogens in the PATRIC database, these bacteria have recently been associated with periodontitis and endodontic infections, respectively<sup>25,26</sup>.

### Virulence

To further characterize the pathogens detected in ancient human dental calculus, we compared information on the functional features of putative genes and proteins associated with virulence, drug resistance, plasmids, transposons and phages to the information available in NCBI records. Although not exhaustive, a preliminary list of well-supported virulence-associated genes and proteins was compiled



**Figure 2** Genomic coverage plot for the periodontal pathogen *T. forsythia*, with details of gene and protein coverage of the virulence factor TF2663, tfsB, from medieval human dental calculus (G12). **(a)** Plot of the *T. forsythia* genome with depth of coverage (0- to 30-fold) shown in red and identity shown in purple. Gene locations of the major virulence factors present (black) or absent (gray) in the assembly are indicated in the outer ring. Notably absent are two transposon-related *tra* pathogenicity islands containing putative tetracycline resistance genes in the *T. forsythia* ATCC 43037 reference strain. Proteins identified by MS/MS are indicated in blue by an asterisk at the corresponding gene locus. **(b)** Enlarged view of forward (blue) and reverse (red) DNA reads mapped to gene TF2663, which encodes the surface layer B protein tfsB. **(c)** Enlarged view of peptide coverage for tfsB, a species-diagnostic virulence factor involved in hemagglutination, adherence and host tissue invasion. **(d)** Example of a tandem mass spectrum indicating the measured mass-to-charge ratios ( $m/z$ ) of the fragments supporting the identification of the tfsB peptide VSIDPDLYLIR (arrowhead in c).

using this method (Table 1 and Supplementary Table 3), showing a wide range of virulence factors associated with adhesion and aggregation (for example, adhesins and lectins) and parasitism (for example, phospholipases, hemagglutinins and hemolysins), as well as extensive machinery for horizontal gene transfer (for example, pilin, CTn and phage sequences). In several cases, we detected both the virulence-associated gene and its protein product, for example, *Msp* and major sheath protein in *T. denticola* and *Rgp* and Arg-gingipain in *P. gingivalis*. Arg-gingipain and Lys-gingipain, another extracellular cysteine proteinase identified by proteomic evidence, are highly antigenic and extremely abundant in *P. gingivalis*, accounting for 10% of the total protein produced by the organism<sup>27</sup>. Notably, we also detected type IV fimbriin, an outer membrane protein variant associated with virulent *P. gingivalis* strains<sup>28</sup>.

### Antibiotic resistance

The human microbiome is an important site of horizontal gene transfer and a potential reservoir of antimicrobial resistance<sup>29</sup>. Metagenomic studies of modern dental plaque have found a wide range of predicted genes related to resistance to diverse antibiotics and toxic compounds<sup>30</sup>. The antiquity of bacterial antibiotic resistance genes has recently been tested in permafrost soils dating to the Pleistocene<sup>31</sup>, but, until now, the antiquity of antibiotic resistance in human microbiota before the use of therapeutic antibiotics had not been investigated.

Using both automated and manual search strategies, we identified within ancient dental calculus numerous DNA sequences with homology to antibiotic resistance genes found in oral and pathogenic bacteria, including genes for multidrug efflux pumps and native resistance genes to aminoglycosides,  $\beta$ -lactams, bacitracin, bacteriocins

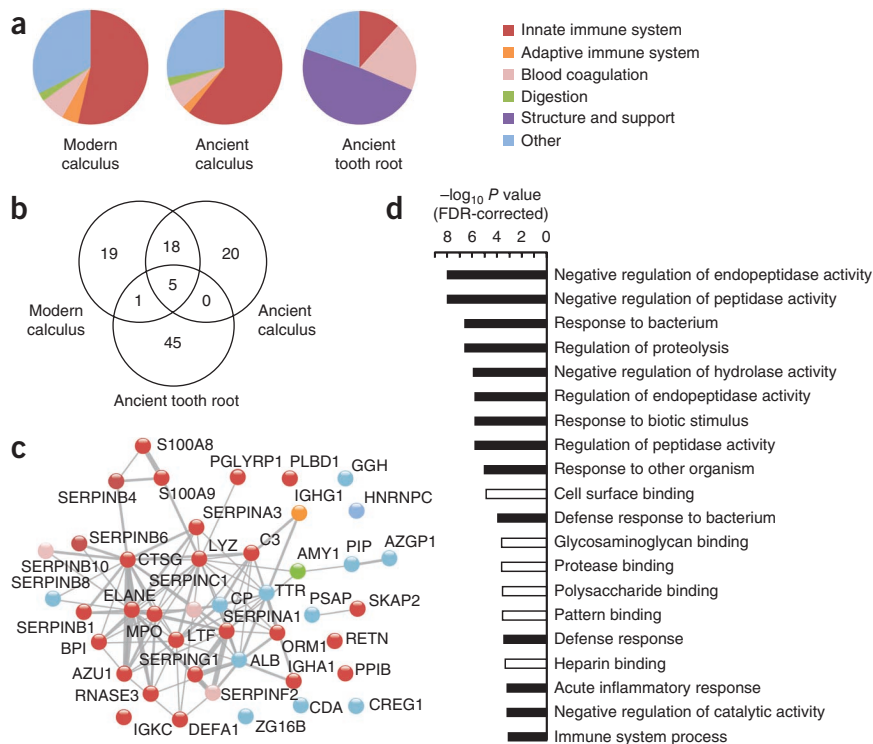
and macrolides, among others, as well as a near-complete plasmid-encoded conjugative transposon carrying efflux pump genes with high homology to CTn5 of *Clostridium difficile* (Supplementary Table 4). Although the exact function of these genes in our samples is unclear, their presence nevertheless demonstrates that the biomolecular machinery for broad-spectrum, low-level antibiotic resistance has long been present in the human microbiome, illustrating how the oral microbiome functions as both a source and a reservoir of new antibiotic resistance<sup>23</sup>.

### Pathogen genome reconstruction: *T. forsythia*

*T. forsythia* (formerly *Bacteroides forsythus* and *Tannerella forsythensis*) is an anaerobic, Gram-negative member of the phylum Bacteroidetes and is a known inhabitant of supragingival and subgingival plaque<sup>32</sup>. It is associated with advanced forms of periodontal disease and has been reported in atherosclerotic lesions<sup>7</sup>. On the basis of 16S rRNA gene data, *T. forsythia* was observed to be at moderate abundance (0.09–0.84%) in the dental calculus of one individual (G12) and, as a pathogen of interest, was selected for genome reconstruction.

Using a conservative mapping strategy, a total of 10,991 contigs were recruited to the ancient *T. forsythia* genome reconstruction, at a mean nucleotide depth of coverage of 5.7 (Fig. 2a). Ninety-one percent of *T. forsythia* genes ( $n = 2,799$ ) were mapped by at least one contig, and unmapped genes included 94 transposases, transfer factors and other mobilization genes that may be specific to the *T. forsythia* ATCC 43037 reference genome strain used for alignment. The largest gap in our genome reconstruction, which spanned ~48,000 bp and 53 genes, corresponded to a complete conjugative transposon carrying putative tetracycline resistance genes that was absent in our reconstructed ancient *T. forsythia* genome. In addition to genetic sequences, tandem

**Figure 3** Metaproteomic comparison of human proteins in modern and ancient dental samples. **(a)** Functional characterization of human proteins identified in modern dental calculus (two individuals), ancient dental calculus (four individuals) and ancient tooth roots (four individuals). **(b)** Venn diagram of shared human proteins by sample type. **(c)** STRING network representation of human proteins identified in ancient dental calculus; nodes are labeled by protein name and colored in accordance with functional categories and connections (gray) to predicted functional partners. The network is set to medium confidence (0.4) for all active prediction methods. **(d)** Gene ontology categories with significant enrichment ( $P < 0.001$ , FDR-corrected) in ancient dental calculus for biological processes (black) and molecular functions (white) related to proteinase regulation, substrate binding and innate immune function. Enrichment is calculated relative to the human genome using the STRING-embedded AmiGO term enrichment tool.



mass spectrometry (MS/MS) identified 118 peptides belonging to 10 *T. forsythia* proteins (Fig. 2a). Of these proteins, nine were outer membrane or S-layer proteins, seven had a known function and four were antigenic: *T. forsythia* surface protein A (TF2661-2, tfsA), *T. forsythia* surface protein B (TF2663, tfsB), outer membrane protein 41 (TF1331, omp41) and one hypothetical protein (TF2339)<sup>33</sup>.

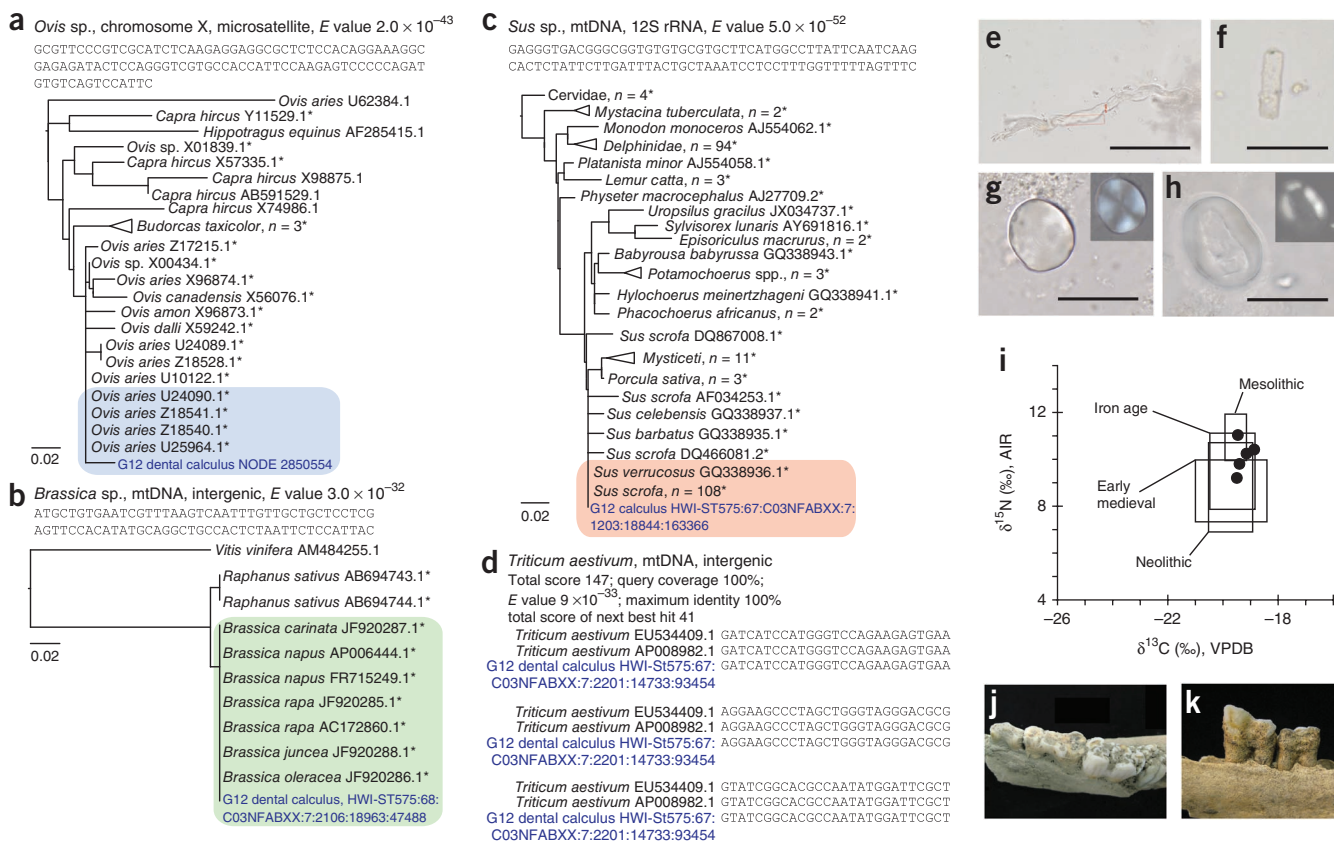
Several virulence factors and antigenic proteins have been identified in *T. forsythia* (Fig. 2a), including Bacteroides surface protein A (BspA), dipeptidyl peptidase-4 (dppIV), tfsA and tfsB, among others<sup>27,33</sup>. The genes encoding each of these virulence factors were present in our reconstruction. The glycosylated *T. forsythia* S-layer proteins tfsA and tfsB are directly involved in hemagglutination, adhesion and tissue invasion<sup>34</sup>. They are also unique and are species diagnostic, as they have no homology to other known S-layer proteins or glycoproteins<sup>35</sup>. DNA and protein coverage of tfsA and tfsB was high in our data set; for example, 10 contigs comprising 116 reads mapped to the TF2663 (*tfsB*) gene (Fig. 2b), and we identified 65 spectra belonging to 27 unique tfsB peptides (Fig. 2c,d). Given that functional *T. forsythia* S-layer protein is essential for host immune evasion and biofilm coaggregation<sup>34</sup>, the discovery of abundant, well-preserved S-layer gene and protein sequences makes ancient dental calculus an excellent candidate for investigating the evolution of periodontal pathogenesis in humans.

### MS/MS analysis of host immunity and disease pathogenesis

Despite dense microbial colonization and the regular introduction of foreign substances, the oral cavity is effective at preventing most infections. At least 45 antimicrobial gene products acting as early responders of the innate immune system have been identified in saliva and gingival crevicular fluid<sup>36</sup>. We identified 43 human proteins within ancient dental calculus, of which 25 are involved in the innate immune system (Fig. 3a). Eight of these proteins have demonstrated antimicrobial properties and include cationic peptides ( $\alpha$ -defensin and azurocidin), metal ion chelators (calgranulin A, calgranulin B and lactoferrin), protease inhibitors (myeloperoxidase) and bactericidal proteins (bactericidal permeability-increasing protein, lysozyme C and peptidoglycan recognition protein 1). Expression of many of these proteins is specific to a particular cell type and even a particular

subcellular component (for example, azurocidin is specific to neutrophil lysosomal azurophilic granules), allowing highly resolved characterization of the immune system response. Approximately one-third of the identified human proteins were shared by ancient and modern calculus (Fig. 3b), and functional profiles were highly similar (Fig. 3a). In contrast, ancient tooth roots were distinct, both in protein composition and function and being dominated by collagens and other proteins involved in mineralized tissue (biglycan, periostin) and vascular (prothrombin) development and maintenance.

The STRING resource<sup>37</sup> was used to investigate functional interaction networks among the human proteins in ancient dental calculus. A large number of functional interactions were predicted (Fig. 3c), and 79% of proteins ( $n = 34$ ) were functionally connected to at least one other protein in the network. Immunoglobulin heavy chain (IgA and IgG) and light chain (kappa) peptides were detected in ancient calculus, as was  $\alpha$ -amylase, a salivary enzyme that breaks down dietary starch; however, the majority of proteins were related to the innate immune system. Human proteins in ancient dental calculus were strongly enriched in extracellular ( $P$  value of  $3.2 \times 10^{-12}$ , false discovery rate (FDR)-corrected) and secretory ( $P$  value of  $4.3 \times 10^{-9}$ , FDR-corrected) proteins, mostly of neutrophilic origin. Extravasated neutrophils are recruited to sites of injury by IgG and have a life span of less than 24 h<sup>38</sup>; thus, neutrophil proteins are only released into calcifying dental plaques during active infection and inflammation. Relatively few human cellular proteins were found, suggesting that immune cells do not invade the calcifying plaque but rather release antimicrobial substances from the junctional and pocket epithelia, a process that is consistent with neutrophil 'frustrated phagocytosis' (ref. 39) and NETosis<sup>40</sup>. Human proteins in ancient calculus were significantly enriched in biological processes related to inflammation, innate immunity and host defense, as well as in molecular functions such as cell surface, protease and glycosaminoglycan binding (Fig. 3d). The observation of an abundance of inflammatory (myeloperoxidase, azurocidin, lysozyme, calprotectin and elastase) and anti-inflammatory ( $\alpha$ -1-antitrypsin and  $\alpha$ -1-antichymotrypsin) innate immune system



**Figure 4** Genetic, microfossil, zooarchaeological and stable isotopic evidence for the medieval human diet at Dalheim, Germany. (**a–c**) Neighboring trees for GenBank sequences aligning to putative dietary sheep (**a**), crucifer (**b**) and pig or boar (**c**) sequences. Trees include accessions with alignment scores of  $>45$ , except for **c**, which is limited to the top 250 alignments; highly significant alignments ( $E$  value  $< 1 \times 10^{-30}$ ) are indicated with an asterisk. BLAST top hits for each dietary sequence are highlighted. The maximum fraction of mismatched bases is 0.75 for tree generation, and distance was calculated using a Jukes-Cantor substitution model. Tree scale bars indicate nucleotide substitutions per site. mtDNA, mitochondrial DNA. (**d**) One sequence aligned to two accessions of bread wheat only. (**e–h**) The microfossils recovered from ancient human dental calculus yielded morphological matches to animal collagen fibers (**e**), a smooth long-cell phytolith (**f**) and starch granules of the grass tribe Triticeae (**g**) and the legume family Fabaceae (**h**). Characteristic starch granule birefringence is shown under polarized light in the insets in **g,h**. Scale bars, 50  $\mu\text{m}$  in **e**, 25  $\mu\text{m}$  in **f–h**. (**i**) C and N stable isotopic values for human bone collagen (black circles) fall within 2 s.d. (boxes) of those measured for other Central European populations and are consistent with a diet of mixed  $\text{C}_3$  terrestrial plant and animal resources. Isotopic values are reported in delta notation:  $\delta^{15}\text{N} = (^{15}\text{N}/^{14}\text{N})_{\text{sample}} / (^{15}\text{N}/^{14}\text{N})_{\text{AIR}} - 1$ , with the air (AIR) standard;  $\delta^{13}\text{C} = (^{13}\text{C}/^{12}\text{C})_{\text{sample}} / (^{13}\text{C}/^{12}\text{C})_{\text{VPDB}} - 1$ , with the Vienna Pee Dee Belemnite (VPDB) standard. (**j,k**) Recovered food waste includes skeletal material from *Sus* species (**j**) and Caprinae (**k**).

proteins in ancient dental calculus, coupled with morphological evidence of attachment loss and alveolar recession, is strongly supportive of active periodontal inflammation and disease.

In addition to this host immunological data, we identified oral pathogens and bacterial virulence proteins in ancient and modern dental calculus known to provoke strong immunological reaction and to contribute to periodontal pathogenesis (**Supplementary Table 3**), most notably *P. gingivalis* (gingipains), *T. forsythia* (S-layer proteins) and *T. denticola* (major sheath protein). *P. gingivalis* has recently been shown to stimulate neutrophils to release resistin, a protein implicated in acquired insulin resistance<sup>41</sup>. Resistin may exacerbate the progression of type 2 diabetes<sup>42</sup>, and, interestingly, we identified resistin on the basis of reasonably abundant evidence (36 spectra, 9 unique peptides) in ancient dental calculus. Resistin was also identified in modern calculus (seven spectra, five unique peptides) but not in ancient tooth roots.

### Ancient dietary reconstruction

Given current challenges in nutritional health and obesity<sup>43</sup>, a growing interest in dietary aspects of the hygiene hypothesis<sup>44</sup> and a recent study suggesting shifts in the ancient oral microbiome associated with periods

of agricultural transition<sup>11</sup>, there is great interest in better understanding the evolutionary history of the human diet. However, paleodietary reconstruction is made difficult by the generally poor preservation of plants and small animals in the archaeological record. Stable isotope analysis of human bone and dental calculus-based plant microfossil research have broadened our knowledge of past dietary practices, but these tools are insufficient to characterize many major dietary components at high taxonomic resolution. Ancient DNA-based approaches offer great advantages and have been used to identify dietary components from archaeological feces (coprolites), as well as to investigate plant remains directly<sup>45</sup>. However, as coprolites and preserved plant remains are relatively rare, we sought to characterize dietary information from dental calculus using both biomolecular and conventional methods.

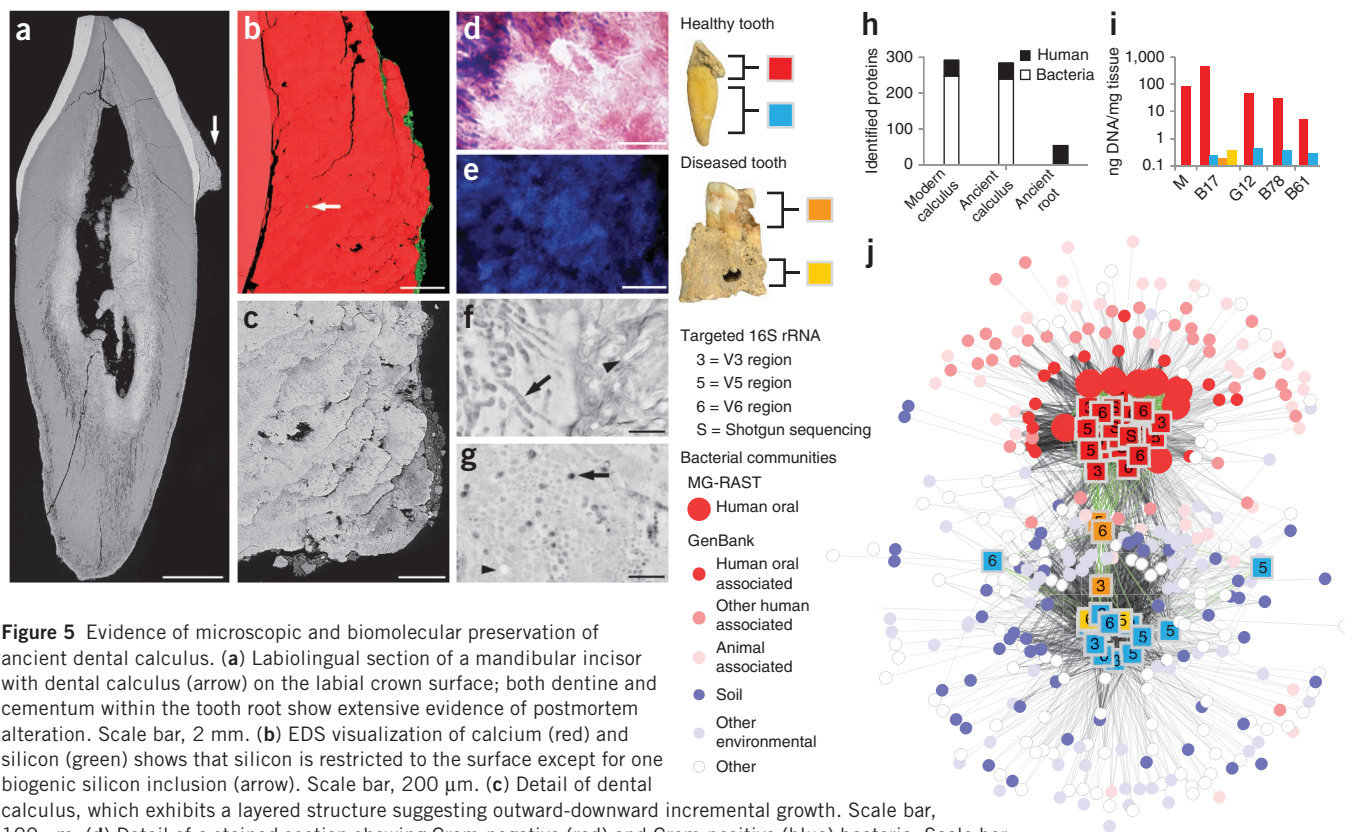
From our metagenomic sequence reads, a total of 487 reads (0.0003%) were confidently identified as eukaryotic organelle sequences; of these, 266 were assigned to the kingdom Viridiplantae, and 21 were assigned to the kingdom Animalia. Within these kingdoms, most of the organelle reads mapped ambiguously to multiple organisms or genera, leaving only 20 reads that could be positively identified at a subfamily level. Of these 20 reads, 17 were of host origin, and the remaining 3 reads

matched diagnostic mitochondrial sequences for pig/boar (*Sus* species), crucifer (*Brassica* species) and bread wheat (*Triticum aestivum*). Analysis of assembled contigs additionally identified one putative sheep (*Ovis* species) and several human ( $n = 326$ ) nuclear genomic sequences (Fig. 4a–d). Although previous studies have reported trace animal domestic DNA contamination (from cattle, pig and chicken) in some PCR reagents<sup>46</sup>, we found no evidence of such contamination, and additionally wheat, crucifers and sheep are not part of this supply chain. The discovery of preserved dietary biomolecules is consistent with previous observations of intact dietary microfossils, such as starch granules, in archaeological dental calculus<sup>10</sup> and with reports of wheat and cassava (tapioca) chloroplast DNA in the dental plaque of living subjects<sup>4</sup>. Turning to proteins, we identified one putative dietary plant protein, chloroplast glyceraldehyde 3-phosphate dehydrogenase (GAPDH), in ancient calculus, but disambiguation below the phylum Viridiplantae was not possible. Faunal proteins were not confidently identified within ancient dental calculus, but we did identify bovine  $\beta$ -lactoglobulin, a milk protein, in modern dental calculus, demonstrating that recovery of dietary animal proteins from dental calculus is possible.

Because our discovery of dietary biomolecules in dental calculus is new, we sought to validate our results using independent paleodietary

methods. Microfossil analysis of ancient dental calculus yielded morphological matches to animal connective tissue fragments ( $n = 2$ ; Fig. 4e), an unidentified monocot phytolith (Fig. 4f), plant bast fibers ( $n = 3$ ) and starch granules consistent with the cereal tribe Triticeae ( $n = 27$ ; Fig. 4g) and the legume family Fabaceae ( $n = 1$ ; Fig. 4h), among other debris (Supplementary Fig. 5). Stable isotope analysis of human bone collagen (Fig. 4i) from the four ancient human individuals indicated a mixed diet of  $C_3$  terrestrial plant and animal resources typical of Central European populations from the late Mesolithic through the medieval period<sup>47–50</sup>. Zooarchaeological analysis of food waste at the site confirmed the presence of pig or boar (*Sus* species; Fig. 4j) and sheep or goat (Caprinae; Fig. 4k), as well as cattle (*Bos* species) and equids (*Equus* species).

Biomolecular analysis of dental calculus thus yields complementary dietary information compared to conventional methods, as well as new findings. The high taxonomic precision of genetic approaches allows closely related taxa (for example, Caprinae) to be distinguished in the absence of diagnostic skeletal elements, and under-represented plant taxa, such as *Brassica*, can be identified without the biological and taphonomic biases that compromise macro- and microfossil preservation of leafy greens and vegetables.



**Figure 5** Evidence of microscopic and biomolecular preservation of ancient dental calculus. (a) Labiolingual section of a mandibular incisor with dental calculus (arrow) on the labial crown surface; both dentine and cementum within the tooth root show extensive evidence of postmortem alteration. Scale bar, 2 mm. (b) EDS visualization of calcium (red) and silicon (green) shows that silicon is restricted to the surface except for one biogenic silicon inclusion (arrow). Scale bar, 200  $\mu$ m. (c) Detail of dental calculus, which exhibits a layered structure suggesting outward-downward incremental growth. Scale bar, 100  $\mu$ m. (d) Detail of a stained section showing Gram-negative (red) and Gram-positive (blue) bacteria. Scale bar, 10  $\mu$ m. (e) Detail of a Hoehchst-stained section showing abundant *in situ* double-stranded DNA. Scale bar, 20  $\mu$ m. (f) Detail of the calculus matrix containing numerous lacunae of filamentous microorganisms (arrow), some of which are mineralized (arrowhead). (g) In another area, the matrix contains both unmineralized (arrow) and mineralized (arrowhead) lacunae of microorganisms resembling cocci. Scale bars, 5  $\mu$ m. (h) Proportions of human and bacterial proteins identified in ancient and modern samples. (i) DNA extraction yields from modern (M) and ancient dental calculus (red), dentine (blue and orange) and alveolar bone samples (yellow). (j) Comparison of microbial communities in ancient dental samples (squares) to those in diverse publicly available samples (circles). 16S rRNA data were generated by shotgun sequencing (S) and targeted amplification and sequencing of hypervariable regions (V3, V5, V6), followed by OTU clustering. Ancient metagenomes ( $n = 38$ ) were plotted within a network where distance scales with OTU community similarity. Modern metagenomes with >20% shared community are shown connected by black lines, and ancient metagenomes with >20% shared community are connected by green lines. A total of 315 GenBank studies were recruited to the network. Ancient metagenomes segregate into two distinct groups: ancient dental calculus samples cluster tightly together, are connected by thick lines and show similarity to modern metagenomes of primarily human and oral origin, and ancient dentine and abscessed bone tissue samples form a more diffuse cluster and recruit primarily soil and environmental metagenomes. Carious dentine forms an intermediate cluster that shares OTUs with both human-associated and environmental sources.

### Taphonomy and contamination

Postmortem taphonomy and contamination pose challenges in ancient biomolecular research. To address these potential problems in our data set, we employed multiple protocols for authenticating our data, including scanning electron microscopy (SEM), energy-dispersive X-ray spectroscopy (EDS), optical microscopy, Raman spectroscopy, protein damage analysis, genetic network analysis and probabilistic genetic source tracking.

After death, environmental microbes are known to infiltrate the dentition, causing substantial tissue degradation, loss of organic matter and altered mineralization patterns in dentine and cementum (Supplementary Fig. 6)<sup>51</sup>. We observed, however, little evidence of postmortem alteration in ancient dental calculus samples (Fig. 5a). EDS imaging showed a thin deposit of silicon-rich soil matrix only on the dental calculus surface (Fig. 5b), and no evidence of altered mineralization was observed within ancient dental calculus, a finding that we confirmed by Raman spectroscopic comparison with modern controls (Supplementary Fig. 7). During life, growth of dental calculus is appositional<sup>8,52</sup>, resulting in a laminar cross-sectional structure characterized by alternating bands of Gram-positive and Gram-negative bacteria (Supplementary Fig. 8), a pattern we also observed in ancient calculus (Fig. 5c,d). DNA fluorescent dye showed a similar distribution of double-stranded DNA in ancient and modern calculus, in many cases resolving to individual cells (Fig. 5e and Supplementary Fig. 9) corresponding to a diverse range of *in situ* bacteria embedded within undisturbed dental calculus matrix (Fig. 5f,g).

Ancient dental calculus yielded microbial ( $n = 239$ ) and human ( $n = 43$ ) proteins in the same relative proportion and with similar functions as in modern controls (Figs. 3a and 5h, and Supplementary Fig. 10), whereas only human proteins ( $n = 53$ ) were confidently identified from tooth roots and bone. Damage analysis of dental calculus proteins showed a higher proportion of spontaneous, non-enzymatic post-translational modifications in ancient samples compared to modern controls; however, both modern and ancient dental calculus peptides exhibited relatively high proportions of non-tryptic cleavage (>10%), an observation consistent with *in vivo* exposure to bacterial and immune system proteases (Supplementary Fig. 11).

Total DNA recovery from ancient dental calculus (5–437 ng DNA/mg calculus) was comparable to that for modern calculus and one to three orders of magnitude greater than from paired dentine (0.3–0.5 ng/mg), carious dentine (0.2 ng/mg) and abscessed bone (0.4 ng/mg) (Fig. 5i). Analysis of 16S rRNA phylotypes using a new network analysis tool developed for this study showed that the bacterial communities within ancient dental calculus closely resembled published human oral microbiomes and were distinct from the communities observed in ancient dentine and bone, which clustered primarily with published soil samples, indicating environmental contamination after death (Fig. 5j and Supplementary Fig. 12). This pattern was found to be robust to extraction method, decontamination method, primer selection, sequencing method and interindividual variation. Reanalysis of our data using the methods employed by HMP<sup>3</sup> yielded equivalent results (Supplementary Figs. 13–15) that were also confirmed using the Bayesian tool SourceTracker<sup>53</sup> (Supplementary Fig. 16). Ancient dental calculus is thus shown to be a remarkably well-preserved biological material that allows direct and detailed investigations of the ancient oral microbiome.

### DISCUSSION

Dental calculus is among the richest biomolecular sources yet identified in the archaeological record. Given the exceptional preservation of DNA within dental calculus (5–437 ng/mg), next-generation

shotgun sequencing libraries can be built from milligrams of material, thereby reducing typical sample requirements for ancient DNA analysis by two orders of magnitude. We demonstrate that the red complex pathogens *T. forsythia*, *P. gingivalis* and *T. denticola* have long been associated with periodontal disease, despite changes in lifestyle, hygiene and diet since the medieval period. We confirm the long-term carriage of opportunistic pathogens in the human oral cavity, including the causative agents of oral and respiratory diseases, as well as bacteria implicated in the progression of cardiovascular disease and the formation of arterial plaques. We find genetic evidence that the human oral cavity has long harbored genes with homology to putative antibiotic resistance genes, the first such demonstration, to our knowledge, in an ancient human-associated sample. We reconstruct the genome of the periodontal pathogen *T. forsythia* without previous enrichment and identify the absence of a complete conjugative transposon carrying putative tetracycline resistance genes found in the reference strain. We report for the first time, to our knowledge, the presence of well-preserved proteins within ancient dental calculus and show that, although the dental calculus metagenome is dominated by bacterial DNA (>99%), the dental calculus metaproteome contains high proportions of both host and microbial proteins of clinical significance. Because the growth of calculus is appositional without remodeling, it may offer a potential solution to the ‘osteological paradox’ in studies of ancient disease<sup>54</sup>, and, given that proteins are known to survive longer in the archaeological record than DNA, dental calculus may allow the recovery of valuable proteomic data from deep time periods that are out of reach using genomic technologies. Finally, we report the first plant and animal DNA sequences recovered from ancient dental calculus; these sequences allow greater taxonomic precision than is currently possible using microfossil or stable isotope paleodietary techniques. Dental calculus is a robust, long-term biomolecular reservoir of ancient disease and dietary information, and it has important implications for the fields of medicine, microbiome research, archaeology and human evolutionary studies.

**URLs.** Graphviz package, <http://www.graphviz.org/>; GitHub, <https://github.com/jfmod/metagenome-sample-network-generator>.

### METHODS

Methods and any associated references are available in the [online version of the paper](#).

**Accession codes.** Illumina and 454 genetic data have been deposited in the NCBI Short Read Archive (SRA) under the project accession SRP029257 and sample accessions SRS473742–SRS473771 and SRS480529–SRS480539 and to MG-RAST<sup>55</sup> under project accession 365 and accessions 4486524.3, 4486533.3, 4486537.3, 4486539.3, 4486540.3, 4486544.3, 4486613.3, 4486614.3, 4486617.3, 4487224.3–4487231.3, 4487233.3–4487235.3, 4487237.3–4487248.3, 4488534.3–4488536.3, 4488542.3, 4517539.3, 4530391.3, 4530438.3, 4530439.3 and 4530473.3–4530475.3. Proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository<sup>56</sup> with the data set identifier PXD000412 and accessions 34605–34628. Computer source code for the network analysis in Figure 5j has been deposited to GitHub (see URLs).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

### ACKNOWLEDGMENTS

We thank the Kantonale Ethik-Kommission Zürich, the Functional Genomics Center Zürich, the Center for Microscopy and Image Analysis, and the Institute of Oral Biology at the University of Zürich; the PRIDE Team; G. Akgül, K. Alt,



D. Ashford, P. Ashton, H. Barton, A. Bouwman, C. Burger, D. Coulthard, J. Hublin, V. Meskenaitė, F. Najar, M. Richards, K. Sankaranarayanan, R. Schlapbach, L. Shillito, T. Stöllner, O. Ullrich and H. Zbinden for assistance with data collection, analysis and management; and M. Carver, F. Dewhirst, A. Tanner, K. Hardy and A. Henry for helpful comments on early drafts and data analyses. This work was supported by the Mäxi Foundation Zürich, the Swiss Foundation for Nutritional Research, Danish Research Foundation grant 29396, Danish Council for Independent Research grant 10-081390, Lundbeck Foundation grants R52-A5062 and R44-A4384, US National Institutes of Health grants R01-HG005172, R01-GM089886, R01-DE018499 and R21-DE018310, European Research Council grant UMICIS/242870, Marie Curie grants EUROTAST FP7-PEOPLE-2010 MC ITN, PALIMPSEST FP7-PEOPLE-2011-IEF 299101 and ORCA FP7-PEOPLE-2011-IOF 299075, a C2D2 Research Priming Fund grant partly funded by Wellcome Trust 097829, Swiss National Science Foundation grant 31003A-135688, the Novartis Foundation, the Novo Nordisk Foundation, the Max Planck Society and the University of York.

#### AUTHOR CONTRIBUTIONS

C.W. conceived the project, with input from M.J.C. R.S. and F.R. contributed samples. C.W., E.C., M.J.C., M.T.P.G., C.v.M., A.R. and Y.H. designed the experiments. C.W., E.C., N.S., C.T., A.R., Y.H., D.C.S.-G., S.C., S.F., H.U.L., P.N., C.D.K., J.V.O., K.Y.T. and E.E. performed the experiments. J.F.M.R., R.V., C.W., C.v.M., J.G., A.R., Y.H., R.Y.T., S.F., C.S., S.C., D.C.S.-G., J.H., J.A.S.C., L.H.H. and T.K. analyzed the data. S.B.-O., Y.H., E.W., C.M.L., M.T.P.G., M.J.C. and F.R. contributed material support to the project. Y.H. wrote the supplementary Raman section. C.W. wrote the manuscript, with critical input from C.M.L., M.T.P.G., M.J.C., C.v.M., E.W., E.C. and the remaining authors.

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Marsh, P.D. Are dental diseases examples of ecological catastrophes? *Microbiology* **149**, 279–294 (2003).
- Pihlstrom, B.L., Michalowicz, B.S. & Johnson, N.W. Periodontal diseases. *Lancet* **366**, 1809–1820 (2005).
- Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
- Dewhirst, F.E. *et al.* The human oral microbiome. *J. Bacteriol.* **192**, 5002–5017 (2010).
- Hujoel, P. Dietary carbohydrates and dental-systemic diseases. *J. Dent. Res.* **88**, 490–502 (2009).
- Kuo, L.C., Polson, A.M. & Kang, T. Associations between periodontal diseases and systemic diseases: a review of the inter-relationships and interactions with diabetes, respiratory diseases, cardiovascular diseases and osteoporosis. *Public Health* **122**, 417–433 (2008).
- Leishman, S.J., Do, H.L. & Ford, P.J. Cardiovascular disease and the role of oral bacteria. *J. Oral Microbiol.* **2**, 5781–5793 (2010).
- Jin, Y. & Yip, H.K. Supragingival calculus: formation and control. *Crit. Rev. Oral Biol. Med.* **13**, 426–441 (2002).
- Hardy, K. *et al.* Starch granules, dental calculus and new perspectives on ancient diet. *J. Archaeol. Sci.* **36**, 248–255 (2009).
- Henry, A.G., Brooks, A.S. & Piperno, D.R. Microfossils in calculus demonstrate consumption of plants and cooked foods in Neanderthal diets (Shanidar III, Iraq; Spy I and II, Belgium). *Proc. Natl. Acad. Sci. USA* **108**, 486–491 (2011).
- Adler, C.J. *et al.* Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat. Genet.* **45**, 450–455 (2013).
- De La Fuente, C.P., Flores, S.V. & Moraga, M.L. DNA from human ancient bacteria: a novel source of genetic evidence from archaeological dental calculus. *Archaeometry* **55**, 767–778 (2013).
- Linossier, A., Gajardo, M. & Olavarria, J. Paleomicrobiological study in dental calculus: *Streptococcus mutans*. *Scanning Microsc.* **10**, 1005–1013; discussion 1014 (1996).
- Wang, J. *et al.* Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. *Sci. Rep.* **3**, 1843 (2013).
- Fierer, N., Bradford, M.A. & Jackson, R.B. Toward an ecological classification of soil bacteria. *Ecology* **88**, 1354–1364 (2007).
- Munro, C.L. & Grap, M.J. Oral health and care in the intensive care unit: state of the science. *Am. J. Crit. Care* **13**, 25–34 (2004).
- Shay, K. Infectious complications of dental and periodontal diseases in the elderly population. *Clin. Infect. Dis.* **34**, 1215–1223 (2002).
- Nakano, K. *et al.* Detection of oral bacteria in cardiovascular specimens. *Oral Microbiol. Immunol.* **24**, 64–68 (2009).
- Gillespie, J.J. *et al.* PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infect. Immun.* **79**, 4286–4298 (2011).
- Willner, D. *et al.* Metagenomic detection of phage-encoded platelet-binding factors in the human oral cavity. *Proc. Natl. Acad. Sci. USA* **108** (suppl. 1), 4547–4553 (2011).
- Socransky, S.S. & Haffajee, A.D. Periodontal microbial ecology. *Periodontol.* **2000** **38**, 135–187 (2005).
- Marri, P.R. *et al.* Genome sequencing reveals widespread virulence gene exchange among human *Neisseria* species. *PLoS ONE* **5**, e11835 (2010).
- Deguchi, T., Yasuda, M. & Ito, S. Management of pharyngeal gonorrhoea is crucial to prevent the emergence and spread of antibiotic-resistant *Neisseria gonorrhoeae*. *Antimicrob. Agents Chemother.* **56**, 4039–4040 (2012).
- Emonts, M., Hazelzet, J.A., de Groot, R. & Hermans, P.W. Host genetic determinants of *Neisseria meningitidis* infections. *Lancet Infect. Dis.* **3**, 565–577 (2003).
- Goker, M. *et al.* Complete genome sequence of *Olsenella uli* type strain (VPI D76D-27C<sup>T</sup>). *Stand. Genomic Sci.* **3**, 76–84 (2010).
- Palmer, R.J. Composition and development of oral bacterial communities. *Periodontol.* **2000** **64**, 20–39 (2014).
- O'Brien-Simpson, N.M., Veith, P.D., Dashper, S.G. & Reynolds, E.C. Antigens of bacteria associated with periodontitis. *Periodontol.* **2000** **35**, 101–134 (2004).
- Amano, A., Nakagawa, I., Okahashi, N. & Hamada, N. Variations of *Porphyromonas gingivalis fimbriae* in relation to microbial pathogenesis. *J. Periodontol. Res.* **39**, 136–142 (2004).
- Sommer, M.O., Dantas, G. & Church, G.M. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128–1131 (2009).
- Xie, G. *et al.* Community and gene composition of a human dental plaque microbiota obtained by metagenomic sequencing. *Mol. Oral Microbiol.* **25**, 391–405 (2010).
- D'Costa, V.M. *et al.* Antibiotic resistance is ancient. *Nature* **477**, 457–461 (2011).
- Tanner, A.C. & Izard, J. *Tannerella forsythia*, a periodontal pathogen entering the genomic era. *Periodontol.* **2000** **42**, 88–113 (2006).
- Sharma, A. Virulence mechanisms of *Tannerella forsythia*. *Periodontol.* **2000** **54**, 106–116 (2010).
- Shimotahira, N. *et al.* The S-layer of *Tannerella forsythia* contributes to serum resistance and oral bacterial co-aggregation. *Infect. Immun.* **81**, 1198–1206 (2013).
- Lee, S.W. *et al.* Identification and characterization of the genes encoding a unique surface (S-) layer of *Tannerella forsythia*. *Gene* **371**, 102–111 (2006).
- Gorr, S.U. Antimicrobial peptides of the oral cavity. *Periodontol.* **2000** **51**, 152–180 (2009).
- Szklarczyk, D. *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, D561–D568 (2011).
- Coxon, A., Tang, T. & Mayadas, T.N. Cytokine-activated endothelial cells delay neutrophil apoptosis *in vitro* and *in vivo*. A role for granulocyte/macrophage colony-stimulating factor. *J. Exp. Med.* **190**, 923–934 (1999).
- Ryder, M.I. Comparison of neutrophil functions in aggressive and chronic periodontitis. *Periodontol.* **2000** **53**, 124–137 (2010).
- Brinkmann, V. & Zychlinsky, A. Neutrophil extracellular traps: is immunity the second function of chromatin? *J. Cell Biol.* **198**, 773–783 (2012).
- Furugen, R., Hayashida, H. & Saito, T. *Porphyromonas gingivalis* and *Escherichia coli* lipopolysaccharide causes resistin release from neutrophils. *Oral Dis.* **19**, 479–483 (2013).
- Kusminski, C.M., McTernan, P.G. & Kumar, S. Role of resistin in obesity, insulin resistance and Type II diabetes. *Clin. Sci. (Lond.)* **109**, 243–256 (2005).
- Gracia-Arnaiz, M. Fat bodies and thin bodies. Cultural, biomedical and market discourses on obesity. *Appetite* **55**, 219–225 (2010).
- Frei, R., Lauener, R.P., Cramer, R. & O'Mahony, L. Microbiota and dietary interactions: an update to the hygiene hypothesis? *Allergy* **67**, 451–461 (2012).
- Palmer, S.A., Smith, O. & Allaby, R.G. The blossoming of plant archaeogenetics. *Ann. Anat.* **194**, 146–156 (2012).
- Leonard, J.A. *et al.* Animal DNA in PCR reagents plagues ancient DNA research. *J. Archaeol. Sci.* **34**, 1361–1366 (2007).
- Bocherens, H., Grupe, G., Mariotti, A. & Turban-Just, S. Molecular preservation and isotopy of Mesolithic human finds from the Ofnet cave (Bavaria, Germany). *Anthropol. Anz.* **55**, 121–129 (1997).
- Oelze, V.M. *et al.* Multi-isotopic analysis reveals individual mobility and diet at the early iron age monumental tumulus of Magdalenenberg, Germany. *Am. J. Phys. Anthropol.* **148**, 406–421 (2012).
- Oelze, V.M. *et al.* Early Neolithic diet and animal husbandry: stable isotope evidence from three Linearbandkeramik (LBK) sites in Central Germany. *J. Archaeol. Sci.* **38**, 270–279 (2011).
- Schutkowski, H., Herrmann, B., Wiedemann, F., Bocherens, H. & Grupe, G. Diet, status and decomposition at Weingarten: trace element and isotope analyses on early mediaeval skeletal material. *J. Archaeol. Sci.* **26**, 675–685 (1999).
- Turner-Walker, G. in *Advances in Human Paleopathology* (ed. Pinhasi, R. & Mays, S.) Ch. 1, 29 (John Wiley & Sons, New York, 2008).
- Zijge, V. *et al.* Oral biofilm architecture on natural teeth. *PLoS One* **5**, e9321 (2010).
- Knights, D. *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* **8**, 761–763 (2011).
- Wood, J.W., Milner, G.R., Harpending, H.C. & Weiss, K.M. The osteological paradox—problems of inferring prehistoric health from skeletal samples. *Curr. Anthropol.* **33**, 343–370 (1992).
- Meyer, F. *et al.* The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**, 386 (2008).
- Vizcaino, J.A. *et al.* The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069 (2013).

## ONLINE METHODS

**Study design and samples.** Narrative and graphical overviews of the study design are provided in the **Supplementary Note** and **Supplementary Figure 17**. Archaeological material was obtained from the medieval St. Petri church and convent complex in Dalheim, Germany (**Supplementary Fig. 18**), and radiocarbon dated to c. 950–1200 CE (**Supplementary Table 5**). The assemblage was evaluated for pathologies (**Supplementary Table 6**), and dental tissues from four well-preserved adult skeletons (G12, B17, B61, B78) and two fauna (F1, F5) were selected for further analysis (**Supplementary Figs. 1, 19 and 20**). Additionally, dental tissues from nine modern controls (P1–P5, P7, P8, P10, P13) with known dental health histories (**Supplementary Table 7**) were obtained under informed consent, and protocols were approved by the Zürich Ethics Commission (KEK ZH-Nr. 2012-0119).

**Microscopy and spectroscopy.** A mandibular incisor from B78 was sectioned longitudinally and examined according to standard protocols with a Tescan VEGA SEM using backscattered electron (BSE) imaging and EDS with a Si(Li) detector. Dental calculus deposits from B78 and P3 were fixed, decalcified and prepared into serial thin sections using modified standard protocols, followed by Gram and Hoechst staining and visualization using a Zeiss Axio Imager M2 and a Leica DMI6000 B microscope. Microfossils were obtained from dental calculus (G12, B17, B61, B78) and dental calculus/crown cementum (F5) deposits (**Supplementary Table 8**) using an incremental HCl decalcification protocol (**Supplementary Note**) and visualized using a Zeiss compound microscope under white and polarized light to identify pollen, phytoliths, starch granules and other debris (**Supplementary Table 9**) by comparison to reference collections. To evaluate mineralogical composition, Raman spectroscopy was applied to six calculus (G12, B17, B61, B78, P3, P13), nine dentine (G12, B17, B61, B78, P4, P5, P7, P8, P10) and five soil matrix (M1–M5) specimens using a HORIBA XploRA instrument (100× magnification and 532-nm laser wavelength) and analyzed for the main  $\text{PO}_4^{3-}$  peak position and peak area, as well as the peak intensity ratios of C-H ( $\sim 2,940 \text{ cm}^{-1}$ )  $I(\text{CH})$  and main phosphate peak  $I(\text{P})$  (**Supplementary Table 10**).

**Isotope ratio mass spectrometry.** Rib specimens from G12, B17, B61 and B78 were cleaned by abrasion, and collagen was extracted according to the method of Richards and Hedges<sup>57</sup> with an additional ultrafiltration step. Carbon and nitrogen isotopic values were measured in duplicate using a Thermo-Finnigan Delta XP continuous-flow isotope-ratio mass spectrometer following combustion in an elemental analyzer FLASH EA 2112 (**Supplementary Table 11**).

**DNA extraction.** Ancient samples were extracted in a dedicated ancient DNA laboratory at the University of Zürich Centre for Evolutionary Medicine in accordance with established contamination control precautions and workflows. DNA was extracted from dental calculus (G12, B17, B61, B78, P2), dentine (G12, B17, B61, B78), carious dentine (B17), abscessed alveolar bone (B78) and burial matrix (M1–M5) by phenol-chloroform extraction followed by Qiagen MinElute column purification (**Supplementary Tables 12–17**). Burial matrix and NaOCl-decontaminated dentine were tested for the presence of endogenous human DNA using targeted PCR and qPCR (**Supplementary Tables 18 and 19**). To optimize DNA extraction from dental calculus, five extraction buffers (A–E) and three decontamination methods were tested and compared. Two extraction buffers (A, 0.45 M EDTA, 10% proteinase K; B, 0.1 M EDTA, 10% proteinase K, 10 mM Tris-HCl, 10 mM NaCl, 2% SDS, 5 mM  $\text{CaCl}_2$ , 40 mM DTT) and three decontamination methods (2% NaOCl, 0.5 M EDTA wash, none) were selected for further analysis and were used in combination to produce nine DNA extracts from B61 and G12 dental calculus.

**DNA library construction and sequencing.** DNA extracts from the optimization experiment were built into nine shotgun libraries using a NEBNext Quick DNA Library Prep Master Mix Set (e6090) with DNA oligonucleotides containing a sample-specific multiplex index sequence (**Supplementary Fig. 21 and Supplementary Table 20**). Libraries were amplified with Phusion HS II enzyme and sequenced on one lane of an Illumina HiSeq 2000 using single-end 1 × 100-bp chemistry, resulting in 93,677,545 reads after the removal of low-quality sequences (Illumina CASAVA 1.8.0, default settings, sequences <25 bp and/or with Phred scores <35 removed; **Supplementary Table 21**).

Separately, 30 16S rRNA gene amplicon libraries were generated from dental calculus (G12, B17, B61, B78), dentine (G12, B17, B61, B78), carious dentine (B17) and alveolar bone abscess (B17) ancient DNA extracts generated using extraction buffer A and without previous decontamination. Universal primers targeting variable regions V3, V5 and V6 of the 16S rRNA gene were developed and tested *in silico* (**Supplementary Tables 22 and 23**) and *in vitro* (**Supplementary Fig. 22**). Each library was generated from a minimum of three amplifications (30–35 cycles) using Phusion HS II enzyme and 454 amplicon Fusion primers with multiplex identifiers (MIDs), and pooled 454 libraries were sequenced with a Roche GS Junior, resulting in 170,807 reads after the removal of low-quality sequences (Roche GS RunProcessor, default settings; **Supplementary Table 24**).

**16S rRNA taxonomic classification.** A reference data set containing full-length 16S ribosomal RNA sequences was constructed from the NCBI GenBank database, whereby all publicly available 16S ribosomal gene sequences found in the NCBI GenBank database were downloaded, screened for chimeras using uchime<sup>58</sup>, aligned using the INFERNAL aligner<sup>59</sup> v1.0.2, trimmed and clustered at a sequence identity cutoff of 98% with a hierarchical clustering algorithm using sequence identity as the measure of distance and single linkage as the cluster metric. This data set has high overlap with both the Greengenes<sup>60</sup> (90%) and RDP<sup>61</sup> (92%) databases and was constructed to standardize filtering and alignment methods, as well as to streamline GenBank data retrieval for network analysis. Amplicon and shotgun sample reads were aligned to the reference OTU data set, and reads with a bit score of <40 or negative structure score were discarded. Sample reads were mapped to the reference OTUs by assigning the OTU ID of the most similar reference sequences. Conflicting OTU IDs were discarded. OTUs containing 16S rRNA gene sequences belonging to a reference genome or culture collection were assigned the consensus taxonomy of all such sequences in the OTU. In the case of OTUs that contained no reliable source of taxonomy, the taxonomy of the OTU was inferred by decreasing the clustering threshold until the point at which the OTU was merged with another in which sequences with reliable taxonomy existed.

**Network analysis.** Network analysis of community similarity was performed to compare the microbial communities of ancient dental samples to each other and to environmental samples deposited in GenBank and MG-RAST (project 128). Only environmental samples with at least 20 OTUs were considered (1,818 of 37,689), and only samples with at least 20% similarity to one of the ancient samples are shown in the network (315 of 1,818). The similarity between a pair of samples was calculated as the number of shared OTUs divided by the total number of different OTUs found in both samples. The network was rendered using the neato program from the Graphviz package (see URLs).

**Phylogenetic tree.** Ancient dental calculus, amplicon and shotgun OTU tables were merged, and a full-length 16S rRNA sequence representative for each OTU was chosen. Phylogenetic relationships were inferred with FastTree<sup>62</sup> v2.1.3 (generalized time-reversible model).

**Validation of results using the RDP and QIIME pipelines.** To confirm that the taxonomic characterization of ancient dental samples was robust to database choice and clustering parameters, the 16S rRNA amplicon data were reanalyzed using the Greengenes database (v.4Feb2011) and the RDP Pyrosequencing<sup>61</sup> and QIIME<sup>63</sup> pipelines. Only reads of  $\geq 70$  bp with 100% identity to both forward and reverse primers were analyzed. OTUs were clustered at 97% identity, and singleton OTUs were discarded (**Supplementary Table 25**). The OTU table was rarefied to 1,265 sequences/sample and analyzed at the L2, L5 and L6 levels. Alpha and beta diversity were calculated using QIIME default parameters. The BIOM file for these data is available as **Supplementary Data Set 1**. This OTU table was merged with an OTU table generated from the HMP data set using the same parameters, and the two data sets were compared using Principal Coordinates Analysis; the BIOM file for these data is available as **Supplementary Data Set 2**.

**Source tracking.** To test for contamination in the ancient dental samples, Bayesian microbial source tracking<sup>53</sup> was performed (1,000 'burn-in' iterations

using Gibbs sampling with 25 random restarts) on the merged OTU file using HMP plaque, HMP skin, HMP gut and ancient tooth root (environmental proxy) as sources.

**Dietary DNA analysis.** Shotgun reads  $\geq 75$  bp in length were searched against a complete collection of full mitochondrial and chloroplast genome sequences published as of July 2012 ( $>6,000$  organelle genomes) using BLASTN. Results were accepted only if they exhibited 100% query coverage and 100% sequence identity, were not hits to 16S or 23S rRNA genes and did not match more than one genus perfectly, and any secondary hits outside the genus of the first hit had to show at least two diagnostic point mutations relative to the perfect hit.

**Total taxonomic characterization of dental calculus.** Library reads were pooled by individual (B61, S1–S4; G12, S5–S8) and *de novo* assembled into 2,005,273 contigs using Velvet<sup>64</sup> v.1.0.2.3 (*k*-mer length of 29 bp, minimum of 100-bp contig length) (Supplementary Table 26). Contigs were searched against the NCBI nr and gss databases available as of July 2012 using Megablast, filtered for highly unique, high-scoring top hits ( $>95$ -bp alignment,  $>97\%$  identity, *E* value of  $<1 \times 10^{-14}$ ). A total of 61,584 contigs passing these filters were assigned taxonomy.

**Pathogen analysis.** Contigs were further filtered to remove contigs with second hits of comparable quality and  $>90\%$  identity to other taxa, resulting in 53,924 highly unique contigs that can be reasonably assigned to a single species. Species-level assignments were then cross-referenced against the PATRIC database<sup>19</sup>, resulting in 40 putative pathogen identifications. To determine whether these species assignments were reasonable for the oral cavity, we applied the same BLAST and conditional filter approach to shotgun metagenomic contigs reported for 109 HMP supragingival dental plaque samples and compared the results. Feature information for each ancient contig was retrieved from the top-hit BLAST results and manually screened for putative genes associated with virulence, drug resistance, plasmids, transposons and phages with annotations in PubMed records.

**Antibiotic resistance analysis.** Sequences for all identified taxa were screened for putative antibiotic resistance elements using three methods: (i) BLASTX search against the Antibiotic Resistance Database (ARDB)<sup>65</sup>, (ii) BLASTX search against the NCBI nr database followed by keyword search of translated gene function and (iii) manual search of gene annotations assigned to pathogens.

**Genome reconstruction.** All G12 contigs of  $\geq 100$  bp were searched against the NCBI nt and gss databases using Megablast and filtered for contigs aligning to *T. forsythia* strain ATCC 43037 with an *E* value of  $\leq 1 \times 10^{-6}$  within the top 100 hits. Filtered contigs were pooled and submitted to the BLAST Ring Image Generator (BRIG)<sup>66</sup> tool for mapping. Using BRIG, contigs were aligned to *T. forsythia* strain ATCC 43037 using the Megablast search option and a sequence identity cutoff of  $\geq 95\%$ . In cases where a contig aligned to the *T. forsythia* genome more than once, the alignment with the highest bit score was mapped. In cases where multiple alignments with identical top bit scores were observed, the contig was mapped to all top bit score loci, but the depth of coverage for each locus was divided by the number of loci. Genes not mapped in the assembly and large gaps (Supplementary Fig. 23) were analyzed for function.

**Protein analysis.** Total proteins were extracted from dental calculus (G12, B17, B61, B78, P1, P2), dentine (G12, B17, B61, B78), carious dentine (B17), abscessed alveolar bone (B78) and dental calculus/crown cementum (F1, F5) and from four negative extraction controls using a modified filter-aided sample

preparation (FASP)<sup>67</sup> protocol. A total of 290,466 MS/MS spectra were generated using 3 instruments (LTQ-Orbitrap Velos, Q-Exactive Hybrid Quadrupole Orbitrap and MaXis UHR-Qq-TOF) (Supplementary Table 27). Tandem mass spectra were converted to Mascot generic format using ProteoWizard v.2.2.3101 with vendor peak picking option for MS level 2 and deisotoped and deconvoluted using the H-Score script<sup>68</sup>. ProteinPilot v.4 was used to analyze protein modification and damage patterns (Supplementary Table 28). MS/MS peak lists were searched using Mascot v.2.3.02 against all proteins in UniProtKB/SwissProt as of 31 October 2012 and two custom protein databases built from the Human Oral Microbiome Database (HOMD)<sup>69</sup> as of 11 October 2012 and all complete soil bacterial genomes in GenBank as of 22 February 2012. The results were further validated using Scaffold v.4.0.5, resulting in 12,609 unique peptide identifications resolving to 589 proteins identified with  $>99\%$  confidence and  $\geq 2$  unique peptides. Contaminants were identified and removed (Supplementary Table 29). Metadata for human proteins was retrieved using the GeneCards v.3 GeneALaCart tool<sup>70</sup> and used to manually classify each protein into six categories: innate immune system, adaptive immune system, blood coagulation, digestion, structure and support, and other. Protein interaction and gene ontology (GO) information was obtained using STRING 9.0 (ref. 37) in protein mode. Bacterial proteins were binned by length (group 1,  $<15$  residues; group 2,  $>15$  residues) and searched against the NCBI database using BLASTP (group 1, expect value 20,000, PAM30 Score Matrix; group 2, expect value 1,000, BLOSUM62 Score Matrix). Resulting BLASTP files were then parsed using MEGAN<sup>71</sup> and analyzed for protein function using SEED hierarchy<sup>72</sup>.

57. Richards, M.P. & Hedges, R.E.M. Stable isotope evidence for similarities in the types of marine foods used by late mesolithic humans at sites along the Atlantic coast of Europe. *J. Archaeol. Sci.* **26**, 717–722 (1999).
58. Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
59. Nawrocki, E.P., Kolbe, D.L. & Eddy, S.R. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**, 1335–1337 (2009).
60. McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618 (2012).
61. Cole, J.R. *et al.* The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **37**, D141–D145 (2009).
62. Price, M.N., Dehal, P.S. & Arkin, A.P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
63. Kuczynski, J. *et al.* Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Curr. Protoc. Bioinformatics* Chapter 10, Unit 10.17 (2011).
64. Zerbino, D.R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
65. Liu, B. & Pop, M. ARDB antibiotic resistance genes database. *Nucleic Acids Res.* **37**, D443–D447 (2009).
66. Alikhan, N.F., Petty, N.K., Ben Zakour, N.L. & Beatson, S.A. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12**, 402 (2011).
67. Cappellini, E. *et al.* Resolution of the type material of the Asian elephant, *Elephas maximus* Linnaeus, 1758 (Proboscidea, Elephantidae). *Zool. J. Linn. Soc.* **170**, 222–232 (2014).
68. Savitski, M.M., Mathieson, T., Becher, I. & Bantscheff, M. H-score, a mass accuracy driven rescoring approach for improved peptide identification in modification rich samples. *J. Proteome Res.* **9**, 5511–5516 (2010).
69. Chen, T. *et al.* The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database (Oxford)* **2010**, baq013 (2010).
70. Stelzer, G. *et al.* In-silico human genomics with GeneCards. *Hum. Genomics* **5**, 709–717 (2011).
71. Huson, D.H., Auch, A.F., Qi, J. & Schuster, S.C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
72. Mitra, S. *et al.* Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG. *BMC Bioinformatics* **12** (suppl. 1), S21 (2011).