PLOS ONE

Metagenomic Analysis Reveals Presence of *Treponema denticola* in a Tissue Biopsy of the Iceman

Frank Maixner[1]*[◕], Anton Thomma[2][◕], Giovanna Cipollini[1], Stefanie Widder[2], Thomas Rattei[2][¶], Albert Zink[2][¶]

1 Institute for Mummies and the Iceman, EURAC Research, Bolzano, Italy, 2 CUBE - Division of Computational Systems Biology, Department of Microbiology and Ecosystem Science, University of Vienna, Vienna, Austria

Abstract

Ancient hominoid genome studies can be regarded by definition as metagenomic analyses since they represent a mixture of both hominoid and microbial sequences in an environment. Here, we report the molecular detection of the oral spirochete *Treponema denticola* in ancient human tissue biopsies of the Iceman, a 5,300-year-old Copper Age natural ice mummy. Initially, the metagenomic data of the Iceman's genomic survey was screened for bacterial ribosomal RNA (rRNA) specific reads. Through ranking the reads by abundance a relatively high number of rRNA reads most similar to *T. denticola* was detected. Mapping of the metagenome sequences against the *T. denticola* genome revealed additional reads most similar to this opportunistic pathogen. The DNA damage pattern of specifically mapped reads suggests an ancient origin of these sequences. The haematogenous spread of bacteria of the oral microbiome often reported in the recent literature could already explain the presence of metagenomic reads specific for *T. denticola* in the Iceman's bone biopsy. We extended, however, our survey to an Iceman gingival tissue sample and a mouth swab sample and could thereby detect *T. denticola* and *Porphyrimonas gingivalis*, another important member of the human commensal oral microflora. Taken together, this study clearly underlines the opportunity to detect disease-associated microorganisms when applying metagenomics- enabled approaches on datasets of ancient human remains.

Introduction

Next generation sequencing (NGS) provides a unique opportunity to address fundamental research questions in various scientific fields [1–3]. Numerous studies have applied NGS to reconstruct the genomes of different lifeforms [4–6] and have used this high-throughput methodology to analyse the genetic anomalies in human cancers [7] or to resolve the structure and function of the human microbiome [8], to name two examples.

In addition to this, the field of ancient DNA research has also benefitted from advancements in sequencing technology. Highly degraded ancient DNA [9,10], with a majority of fragments having a size of 50 bp or smaller, cannot be fully targeted by conventional polymerase chain reaction (PCR) and/or Sanger-based sequencing. However, NGS technologies that produce read lengths between 35 and 400 bp [11] cover this small size sequence range. Thus, highly fragmented ancient DNA, present in minute amounts, can be analyzed to an unprecedented depth and accuracy with NGS. Draft nuclear genomes of a 38,000-year-old Neanderthal specimen [12] and of the 30,000 to 48,000-year-old Denisova fossil [13] display milestones in the application of NGS on ancient remains and shed new light into the genetic evolution of hominoids. Further NGS based studies have allowed new insights

into the phenotype and origin of the 4,000-year-old Saqqaq individual [14] and the mummified 5,300-year-old Iceman [15]. In the near future NGS will be routinely used to analyze ancient human remains. The more ancient human genomes become available, the more we will then learn about our population history [16]. The aforementioned ancient hominoid genome studies can be regarded by definition as metagenomic analyses since they represent a community of both hominoid and microbial sequences in an environment. The amount of non-hominoid sequence material can thereby vary dramatically ranging from 15.8% of all recovered reads in the Shaggaq genome study [14] up to 99% of all reads in the Neanderthal genome study [12]. This percentage of recovered reads in metagenomes is not only affected by the complexity and diversity of the community but also by the sequencing depth.

Initially, these non-hominoid sequence reads had been declared as complex background dataset and were not subjected to further in-depth analysis. Several recently published studies, however, impressively demonstrate what precious additional information besides the human host genome can be extracted from ancient DNA datasets of human remains. The combination of DNA enrichment methodology with NGS enabled researchers to recover the first complete ancient pathogen genome of *Yersinia*

pestis from Black Death victims in Medieval England [17]. In one exceptionally well-preserved medieval individual with indicative leprosy bone lesions, it was even possible to *de novo* assemble the complete genome of an ancient *Mycobacterium leprae* strain from the ancient DNA dataset without previous DNA enrichment [18]. Both studies provided new insights into the evolution of important human diseases as these applied methodologies paved new ways for analyzing ancient DNA datasets.

Despite the reconstruction of ancient pathogen genomes in individuals with known pathologies, the analysis of ancient DNA datasets also offers the potential to detect unexpected, ancient pathogens without any previous pathological indication. The analysis of the non-human reads in the Iceman's genome study, for example, indicated the presence of *Borrelia burgdorferi* sequence reads [15]. Recently, the re-analysis of the Iceman's metagenomic dataset using a different bioinformatics pipeline proved the presence of *Borrelia* reads on the genus level [19]. Further SNP based analysis, however, indicated a *Borrelia* species divergent form *B. burgdorferi*. This different species assignment displayed how challenging it can be to identify and correctly classify endogenous ancient DNA in huge datasets of highly fragmented DNA reads and it furthermore stresses the necessity for accurate filtering steps and computational proof for ancient DNA authenticity [20,21]. For the taxonomic classification of metagenomic reads, the most basic method is to use BLAST to search for homology-based similarity to references in a database. Most previous studies provided a taxonomic profile of the non-hominoid sequence reads by using a BLASTn search against the public database [15,22]. This basic approach, however, is known to be error prone and can lead to false positive assignments [19]. The study of Zaremba-Niedz'wiedzka and Anderson [23] provided an alternative approach in taxonomically profiling the Neanderthal metagenome by BLAST searching against a designated rRNA gene sequence database. The specific sequence curation of rRNA databases and the high phylogenetic resolution of the rRNA genes allow a relatively fast read alignment and accurate taxon assignment. In addition, Zaremba-Niedz'wiedzka and Anderson's [23] computationally age classified the retrieved rRNA reads which finally brought the authors to the conclusion that the microbiome of the Neanderthal bone sample was highly dominated by *Actinobacteria* which showed no sign of DNA damage.

In this study we re-analyzed the metagenomic data of the Iceman's genomic survey. Similar to the study of Zaremba-Niedz'wiedzka and Anderson [23] we first taxonomically profiled the non-human dataset by screening for rRNA specific reads. Unexpectedly, the taxonomic assignment and further genome wide analysis indicated the presence of sequence reads most similar to a human opportunistic oral pathogen. The sequence reads were then further subjected to DNA damage analysis and used for phylogenetic assignment. Finally, we extended our survey to detect members of the human commensal oral microflora in the Iceman's oral cavity.

Materials and Methods

In silico Analysis of the Iceman's Metagenome

Identification of the non-human reads in the iceman's metagenome. All SOLiD sequence reads from the Iceman's whole genome sequencing study ([15]; ENA Study Accession No.: ERP001144) were filtered for a minimal mean quality score of 20 and, in an additional step, for paired reads both passing the mean quality filter. To separate human from non-human reads, all remaining read pairs were mapped against the human reference genome hg19 GRCh37 (Feb. 2009, GB Accession No.:

GCA_000001405.1) using the SHRiMP software package [24] with the following parameters: -p opp-in -I 10,1000-n 3-I −6, reflecting an insert size of 10 to 1000 nucleotides for the paired-end reads and a mismatch score of −6 (see Appendix S1 for details on the parameter optimization). In this manner, the retrieved unmapped reads represent the non-human fraction of the Iceman metagenome. In all further analysis, we refer to this dataset as "Iceman non-human reads".

Taxonomic profile of the iceman's non-human reads and screening for potential pathogens. The taxonomic structure of the metagenome was determined from all Iceman non-human reads that encode segments of the small subunit (SSU) or large subunit (LSU) ribosomal RNA (rRNA) genes, using sequence alignment and taxonomy assignment parameters optimized for the degraded DNA present in the Iceman sample. The Iceman non-human reads were searched against the SILVA database (SSU/ LSU ref SILVA 106 from 30th august 2011; [25]). The BLAST [26] search was performed with the following parameters: -b 1000-v 1000-e 0.1-F F. The BLAST hits were further analysed with the MEGAN 4 software package [27] and assigned according to their taxonomy based on the SILVA database. The profile was created with the options: synonyms file = silva2ncbi.map (from http:// www-ab.informatik.uni-tuebingen.de/data/software/megan4/ download/silva2ncbi.zip); usekegg = false; useseed = false. The scoring parameters were: maxmatches = 1000 minscore = 50 top-percent = 10 minsupport = 1. The abundance of bacterial genera was obtained from the resulting taxonomic profile of the Iceman non-human reads. Each genus present in the metagenome was classified according to the NCBI microbial genome database (ftp://ftp.ncbi.nih.gov/genomes/genomeprj/lproks_0.txt). Genera containing mainly human-pathogenic bacteria represent reasonable candidates for Iceman-associated pathogens and were therefore selected for further analysis at the species level.

Determination of reads originating from treponema genomes. All non-human Iceman metagenomic reads were mapped against all available complete genomes of the genus *Treponema* in the NCBI RefSeq database [28]. This mapping was performed with SHRiMP using the following parameters for unpaired mapping, reduced mismatch score for more sensitive mapping of reads, and reporting only the best mapping alignment: -n 2-i 6–report 1. To prevent unspecific alignments resulting from close sequence similarities between *Treponema* and non-*Treponema* genomes, we retained only reads having sufficiently better similarities to *Treponema* compared to non-*Treponema* genomes (details in Appendix S1). Finally, all *Treponema* specific reads were extracted from the mapping file and used for further analysis.

Reconstruction of ancient genomic fragments of treponema denticola. For the *Treponema denticola* ATCC 35405 genome (NCBI GenBank accession AE017226.1), to which most of the *Treponema*-specific reads were most similar, contiguous consensus sequences were extracted using the mpileup command implemented in samtools [29]. For each genomic region, the gene coordinates and names were obtained from the genome annotation of the *T. denticola* ATCC 35405 genome. For further phylogenetic assignment, the contig containing the 23S rRNA gene was extracted. The logarithmic coverage of the genome was plotted against CDS regions (up- and downstream) and against the tRNA and rRNA regions. The circular plot was done using DNAPlotter from the Artemis package [30].

mapDamage analysis. To assess the nucleotide misincorporation patterns along the DNA fragments, we performed a mapDamage analysis [20] using all reads mapped to the *T. denticola* ATCC 35405 reference genome (gi|41821838|gb|AE017226.1|). Results were compared to the mapDamage patterns of the Iceman

genome specific reads and reads of the human reference genome HG00101 (ENA Study Accession No.: SRP001294), which was also sequenced on a SOLID platform. The map step integrated in the mapDamage tool was performed using the following parameters: −l 70-a 10-t 4. In the plot step using the parameters −l 50-m 0.1, the y-axis range displaying the misincorporation frequencies was set to 0.5 for *Treponema* and 0.1 for the Iceman and human reference samples to display the full range of nucleotide misincorporation frequencies in the different datasets.

Phylogenetic Assignment of the *Treponema denticola* 23S rRNA Contig

The sequence analysis and phylogenetic assignment of the 23S rRNA containing contig was performed with software tools implemented in the ARB software package [31]. The partial Iceman *Treponema* 23S rRNA contig was aligned against a subset of complete 23S rRNA sequences of the genus *Treponema* within the aligned SILVA large subunit ribosomal RNA dataset (SILVA 111, LSU Ref) [32]. The alignment was afterwards manually refined using the ARB sequence editor. Phylogenetic analyses were performed using the DNA maximum-likelihood method [PhyML [33] with the JTT substitution model] implemented in the ARB software package. The partial Iceman *Treponema* 23S rRNA contig was added to the phylogenetic tree using the ARB Parsimony tool with a filter on the 107 informative positions of the contig. Selected non-*Treponema* sequences of the phylum *Spirochaetes* served as outgroups.

Iceman's Gingival Tissue and Mouth Swab Samples

To further screen for opportunistic oral pathogens using PCR, a gingival tissue biopsy (1321) and a mouth swab sample (1324) were taken from the right inner Iceman's mouth region (Fig. S2). Samples were withdrawn using a bone tissue biopsy needle (TRAPSYSTEM®SET, TRAPJ0810, HS Hospital Service S.P.A., Rome, Italy) for sample 1321 and by using a swab sampling device (Copan Diagnostics Inc., California, USA) for sample 1324. The sampling took place under sterile conditions at a temperature of 4°C in the Iceman's conservation cell at the Archaeological Museum of Bolzano, Italy. The samples were immediately stored at −20°C in the ancient DNA laboratory of the EURAC - Institute for Mummies and the Iceman.

Molecular Screening for Opportunistic Oral Pathogens

The Iceman's gingival tissue and mouth swab samples were further subjected to molecular paleomicrobiological analysis to screen for the opportunistic oral pathogens *T. denticola* and *Porphyrimonas gingivalis*. The molecular analyses were conducted at the ancient DNA Laboratory of the EURAC - Institute for Mummies and the Iceman, Bolzano, Italy. Sample preparation and DNA extraction was 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. DNA extraction was performed with approximately 40 mg of gingival tissue and 300 µl swab sample using a chloroform-based DNA extraction according to Tang and colleagues [34] with minor modifications. Three sets of newly designed primers were used in the PCR assay (Table S1). For the 16S rRNA gene PCR assay we adapted an already existing PCR based detection method for *T. denticola* [35] (for details please refer to Table S1) by shortening the fragment length from 316 bp to 68 bp. We took the forward primer from the original publication [35] and manually designed a new reverse

primer in a *T. denticola* specific region by using the ARB editor of the ARB software package and the SILVA small subunit ribosomal RNA dataset (SSURef_106_SILVA). The two newly designed IS1126 primer pairs target the repetitive element IS1126 of *P. gingivalis* by amplifying a fragment of 71 bp and 98 bp length respectively. The PCR reaction mix for all primer sets contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.875 mM MgCl$_2$, 200 µM of each deoxynucleotidetrisphosphate, 0.5 µM of each primer, 0.1 mg/ml bovine serum albumin, 0.05 U/µl AmpliTaq Gold (Applied Biosystems, Foster City, CA, USA) and 4 µl of extracted DNA to a final volume of 50 µl. Polymerase chain reaction was carried out according to the parameters in Table S1. The PCR products were initially documented by electrophoresis on 2.8% agarose TBE gels and then either used directly for Sanger sequencing or cloned into the pCR 2.1-TOPO vector (Life Technologies, Carlsbad, CA, USA) prior to sequencing. Subsequently, 5 µl of direct PCR product or M13-based PCR amplification product from the vector were treated with 1 U of Shrimp Alkaline Phosphatase (SAP) and 0.8 U of ExoI and incubated at 37°C for 60 min, followed by heat inactivation at 75°C for 15 min. Four microliters of each reaction product was sequenced on an ABI Prism 310 DNA automated sequencer, using the BigDye Terminator Cycle Sequencing Ready Reaction Kit version 3.1 (Applied Biosystems, Foster City, CA, USA). The DNA sequences reported in this work have been deposited in the GenBank database (Accession Nos. KJ543701, KJ543702).

Results

Analysis of the Iceman Non-human Reads

The Iceman genome study [15] created a massive metagenomic dataset, which consists of human and non-human reads. The metagenome was sequenced on a SOLID 4 sequencing platform producing short reads up to 50 bp with high accuracy. In this study we applied a bioinformatics pipeline on this SOLID sequencing data in order to accurately separate the human from the non-human reads and to further screen the non-human reads for potential ancient pathogens (Figure 1). To extract the non-human reads from the metagenome all reads were mapped against the human reference genome hg19 GRCh37 using the SHRiMP software package [36]. With parameters adapted to the unique nature of the Iceman metagenomic dataset (as outlined in the Appendix S1) out of the 1.1 billion total metagenomic reads, 0.9 billion could be mapped to the human reference genome. The remaining 0.2 billion non-human reads were further taxonomically classified using MEGAN [27] based on sequence-similarity searches against the small subunit (SSU) and large subunit (LSU) ribosomal RNA SILVA database. In total 371,905 reads of the 0.2 billion non-human reads were assigned to various bacterial and eukaryotic phyla down to the genus level (Figure 2). The bacterial fraction of reads, comprising 88% of all assigned reads, is highly dominated by the phylum *Firmicutes* (73% of the *Bacteria* reads) with the genus *Clostridium* being the major representative of this phylum (54% of the *Firmicutes* reads). The remaining 28% of the bacterial reads were mainly unassigned (25% of the *Bacteria* reads) or spread throughout numerous different bacterial phyla comprising 0.8% (*Proteobacteria*) or less of all assigned non-human reads. In the subsequent step, we focused on bacterial phyla with assigned reads at the genus level aiming to identify highly abundant genera which are dominated by potential human pathogens.
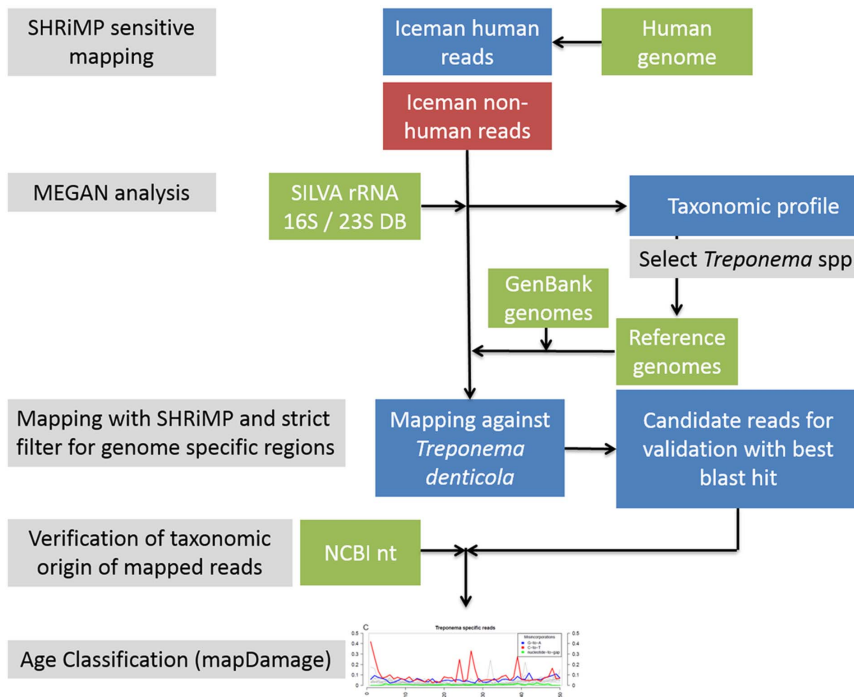
Figure 1. Schematic overview of the bioinformatics pipeline used in this study.
doi:10.1371/journal.pone.0099994.g001

Pre-selection of Bacterial Genera with Potential Human Pathogens

Based on the rRNA taxonomic profile (Figure 2) and following the pathogen classification of the NCBI microbial genome database we pre-selected assigned bacterial genera, which contain potential human pathogens and further ranked the genera by taxonomic abundance. The retrieved list was divided into a high abundance (more than 5000 assigned rRNA reads) and a low
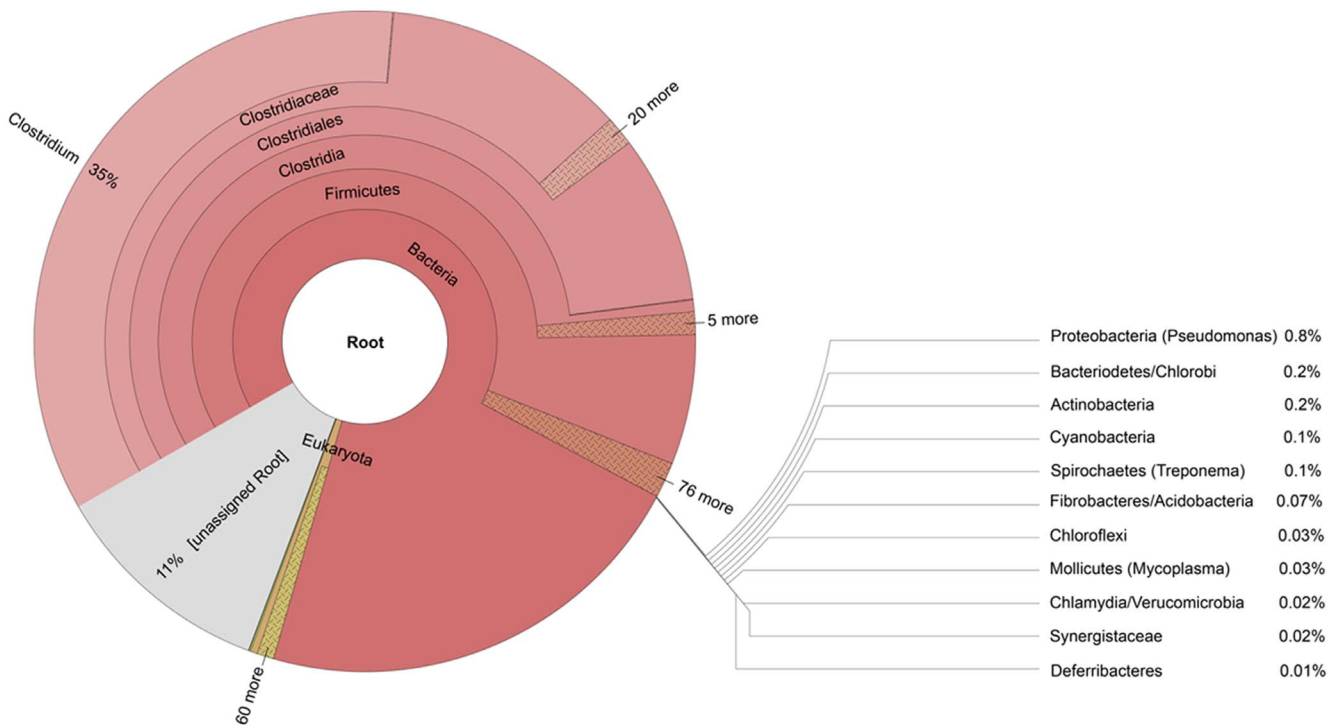
Figure 2. Taxonomic profile of the non-human Iceman metagenome. Phylogenetic assignment of the bacterial and eukaryotic rRNA reads of the Iceman's metagenome to different phyla. Indicated in brackets are the predominant assignable genera within a phylum.
doi:10.1371/journal.pone.0099994.g002

abundance group (less than 5000 assigned rRNA reads) of bacterial genera (Table 1). The listed genera containing human-pathogenic bacteria represent reasonable candidates for Iceman-associated potential pathogens and were then selected for further analysis on species level. Initially, the low abundance group was excluded from all further analysis since the low overall amount of assigned rRNA reads suggested only a minor presence of specific genomic reads which were not sufficient for a meaningful reconstruction of genomic sequences. Furthermore, the first two genera of the high abundance group, *Clostridium* and *Pseudomonas*, were dominated by environmental non-pathogenic bacterial species and thus these genera were not considered for further analysis.

The third most abundant genus with a total of 5,503 assigned rRNA reads is *Treponema*, which is known to contain different human pathogenic species [37]. Nearly all of the reads could be assigned to the human oral bacterium *T. denticola*. To further support this first rRNA-based indication for the presence of human opportunistic pathogen DNA in the metagenomic dataset we decided to focus in the following genome wide survey on the reads specific for the genus *Treponema*.

Genus specific regions of the currently available complete genomes of *Treponema* have been determined from genome alignments to all other published genome sequences from NCBI GenBank [38], retaining only regions having no similarities to any other genome outside the genus. Consequently, Iceman non-human reads mapped with SHRiMP to those of specific regions provided evidence for the presence of the respective genus in the metagenome. Interestingly, 15,670 reads were assigned to specific regions in the genomes from the genus *Treponema*, of which the majority of the reads (8,090) were mapped against the genome of the human opportunistic pathogen *T. denticola* (Figure S1).

Analysis of the *Treponema denticola* Specific Reads

For all further analyses the *T. denticola* specific reads were extracted from the Iceman non-human reads and were first mapped to the *T. denticola* ATCC 35405 genome (Figure 3). Even though the low number of reads excluded a reconstruction of the complete genome, the reads display an even distribution throughout the genome with an accumulation of reads in the two 23S rRNA gene copies (for details on the mapped gene loci please refer to the Table S2).

To assess the nucleotide misincorporation patterns along the DNA fragments, a mapDamage analysis with the *T. denticola* specific reads was performed and compared to the damage pattern of the Iceman human reads (Figure 4). Furthermore, since the Iceman genome displays the first ancient genome sequenced on a SOLID platform [15], a modern human genome sequenced on the same platform was included as a control. Compared to the modern dataset the Iceman genome displays an increased C to T misincorporation pattern at the 5′ end of the reads (the absence of mismatches directly at the 5′ end of the human reads is a technical artefact resulting from the sensitive settings of the SHRiMP mapping tool). An increase of DNA damage was also observed in the *T. denticola* specific reads. However, the damage patterns occur in contrast to the Iceman genome at an order of magnitude higher frequency and the C to T misincorporation pattern is not restricted to the 5′ end and was found additionally within the reads.

The 23S rRNA gene was the only phylogenetic marker gene sufficiently covered by *T. denticola* specific reads. Thus a contiguous

Table 1. Assignment of bacterial rRNA reads of the Icemans metagenome to different genera and pre-selection of human pathogenic or opportunistic pathogenic bacteria according to the NCBI Genome Project database (ftp://ftp.ncbi.nih.gov/genomes/genomeprj_archive/lproks_0.txt).

Genus	Assigned rRNA reads	Pathogenic species in the Iceman non-human reads	
Clostridium	Genus: 32010, all *Clostridiales*: > 1 Mio	Species mainly non-pathogenic	**High abundance group**
Pseudomonas	8939	Mostly non-pathogenic (e.g. *P. flourescens*)	
Treponema	5503	Strong signal for the opportunistic oral pathogen *T. denticola*	
Burkholderia	1092	Plant pathogens	**High abundance group**
Bacillus	845	No human pathogenic *Bacilli*	
Mycoplasma	389	Most to *M. arthritidis*	
Erysipelothrix	171	*Erysipelothrix rusiopathiae*	
Fusobacterium	78	*Fusobacterium* sp. 3_1_5R	
Legionella	70		
Borrelia	36		
Helicobacter	35		

doi:10.1371/journal.pone.0099994.t001
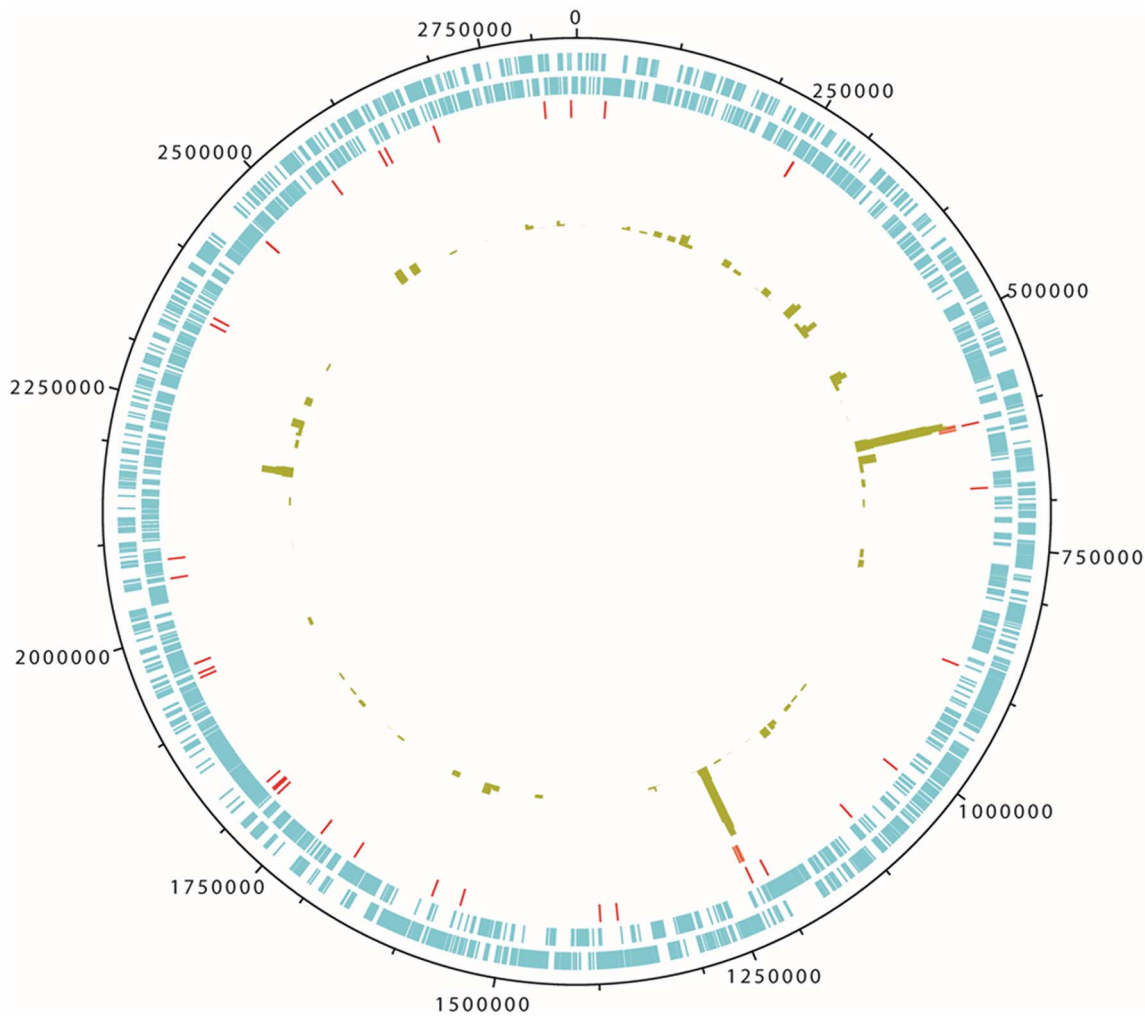
Figure 3. Gene coverage and distribution of the validated Iceman *Treponema denticola* reads mapped on the 2.8 Mb large genome of *T. denticola* ATCC35405. From outer to inner circles coding sequences forward and reverse are highlighted in blue, tRNA and rRNAs in red, and depicted by the green bars are the log scale coverage of mapped reads. For details on genes with mapped reads, please refer to Table S2.
doi:10.1371/journal.pone.0099994.g003

consensus sequence of the 23S rRNA gene was extracted and used for further phylogenetic assignment. The partial Iceman *Treponema* 23S rRNA contig was first aligned against a subset of complete 23S rRNA sequences of the genus *Treponema*. In the next step, a phylogenetic tree based on 23S rRNA genes of bacteria of the genus *Treponema* and selected bacteria of the phylum *Spirochaetes*, which served as an outgroup, was calculated and the partial Iceman metagenome 23S rRNA gene contig was added to the tree using the Parsimony tool in the ARB software package [31]. In the resulting phylogenetic framework the partial Iceman metagenome 23S rRNA gene contig clustered together with sequences of *T. denticola* strains opening a new basal branch highly related to the *T. denticola* sequences (Figure 5).

Molecular Screening for Opportunistic Pathogens in the Iceman's Oral Cavity

The extraordinary well-preserved mummy provides the unique opportunity to confirm the presence of members of the human commensal oral microflora in the Iceman's mouth region (Figure S2A). Therefore, the survey was extended to an Iceman gingival tissue sample and a mouth swab sample using a PCR-based diagnostics assay (Figure S2B). Both the 16S rRNA based PCR

assay for *T. denticola* and the PCR assay targeting the repetitive element IS1127 of *Porphyrimonas gingivalis* gave a positive result. This indicates the presence of opportunistic pathogens in the Iceman gingival tissue and in the mouth swab sample, respectively (Figure 6). In this context, it is important to highlight that only the small IS1127 fragment of both applied IS1127 PCR assays was amplifiable which suggests the degraded nature of the *Porphyrimonas gingivalis* DNA (data not shown).

Discussion

Here we report the discovery of sequence reads indicating the presence of opportunistic pathogens in Iceman's tissue biopsies. In-depth analysis of the metagenome of the Iceman's genomic survey [15] resulted in the recovery of *T. denticola* specific sequence reads which were further subjected to phylogenetic assignment and DNA damage pattern analysis. Finally, we extended our survey to an Iceman's gingival tissue biopsy and mouth swab sample to detect, via conventional PCR, members of the human commensal oral microflora.

Initially we taxonomically classified the non-human reads of the Iceman's metagenome by screening for rRNA specific reads.
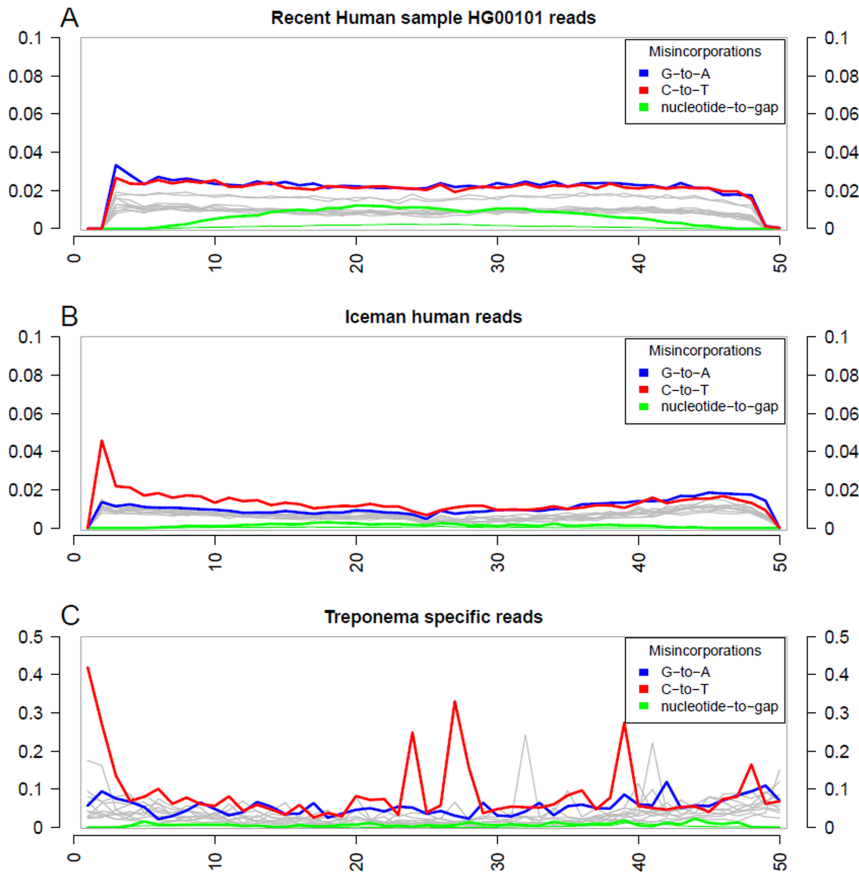
Figure 4. MapDamage analysis displaying the frequency of nucleotide misincorporations (y-axes) in SOLID reads (starting on the 5'-end, x-axes) of different datasets. (**A**) Human reference genome (ENA Experiment Accession No.: ERX008207) (**B**) Human reads of the Iceman metagenome (ENA Study Accession No.: ERP001144) (**C**) Validated *T. denticola* reads from the Iceman metagenome. Grey lines indicate all possible misincorporations; G-to-A and C-to-T misincorporations are plotted in blue and red, respectively. The green lines display all possible variants of a nucleotide-to-gap position.
doi:10.1371/journal.pone.0099994.g004

rRNA-based profiling has been increasingly applied to the taxonomic classification of bacteria in metagenomic studies of recent and ancient DNA datasets [23,39]. Due to the specific sequence curation and alignment procedures used, SILVA-derived rRNA datasets provide fast computation and accurate taxon assignment [40].

Our taxonomic profiling revealed a high abundance of bacterial rRNA reads assigned to the genus *Clostridium* in the predominant phylum *Firmicutes* (Figure 2). The presence of Clostridia in Iceman's tissue and intestinal samples was indicated in earlier molecular studies [41,42] and our finding is consistent with previous data from the Iceman genomic survey [15]. Recently,
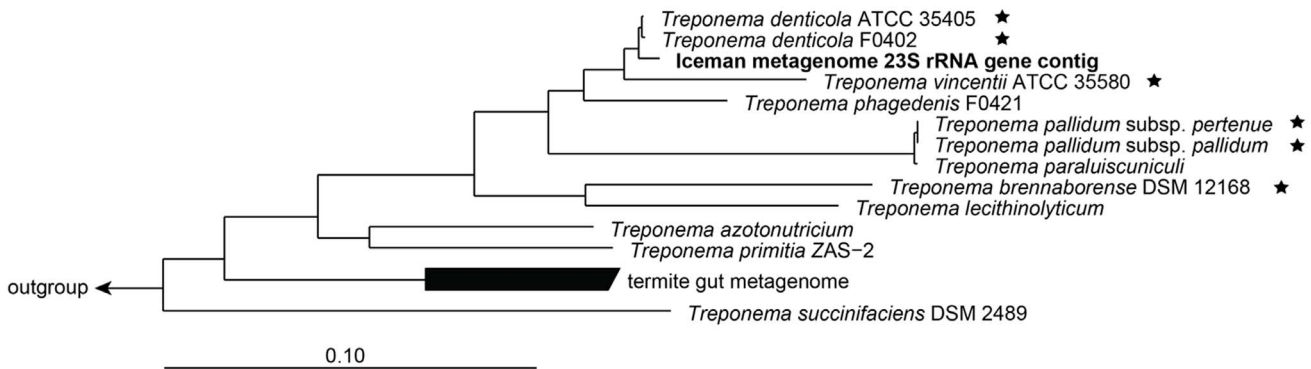
Figure 5. Phylogenetic tree based on bacterial 23S rRNA genes of the genus *Treponema* and selected bacteria of the phylum *Spirochaetes* (serving as an outgroup). The Iceman metagenome 23S rRNA contig is highlighted in bold. All sequences marked with an asterisk belong to a pathogenic or opportunistic pathogenic *Treponema* species. The scale bar indicates 10% estimated sequence divergence.
doi:10.1371/journal.pone.0099994.g005
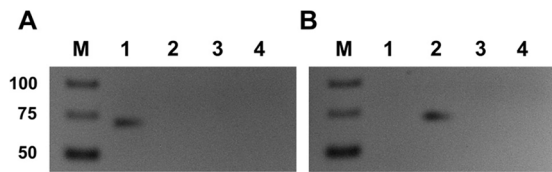
Figure 6. PCR-based detection of different opportunistic oral pathogens in Iceman's gingival tissue (1) and mouth swab (2) samples. (A) PCR assay targeting the 16S rRNA gene of *T. denticola*. **(B)** PCR assay targeting the repetitive element IS1126 of *Porphyromonas gingivalis*. All assays include a PCR negative control (3) and a PCR of the DNA extraction blank (4).
doi:10.1371/journal.pone.0099994.g006

metagenomic analysis discovered a high percentage of *Firmicutes* also in tissue samples of Egyptian mummies [43]. Moreover, a microbial survey of the mummies from the Capuchin Catacombs in Sicily, Italy, indicates that members of the order *Clostridiales* are predominant on mummified skin and muscle samples [44]. Presumably, these Clostridia-like bacteria display remnants of the post-mortem growth of the bacterial community, which is involved in the overall body decomposition process [45,46]. With the present data, however, it is impossible to predict whether these spore-forming bacteria are still viable and could start to grow under favourable environmental conditions on the mummified tissues, thus increasing the risk for biodeterioration of these precious human remains.

Unexpectedly, after ranking the bacterial rRNA reads by abundance, we detected in addition to the numerous *Clostridium*-specific reads a high number of reads affiliated to the genus *Treponema* of which nearly all were assigned to the human oral spirochete *T. denticola*. *T. denticola* is an opportunistic pathogen, which is a member of the human commensal oral microflora. Whenever the oral microbiome is in a state of disequilibrium or when environmental conditions change within the host, *T. denticola* can promote a pathology as part of a microbial consortium. To further support the initial rRNA-based results, we extended our survey to whole genomes available for the genus *Treponema*. By mapping the non-human reads against regions having no similarities to any other genome outside the genus *Treponema* we obtained 8,090 reads mapping against the genome of the human opportunistic pathogen *T. denticola*. Due to the limited dataset we observed an uneven mapping of the reads to the genome with a high read abundance within the 23S rRNA gene. One possible explanation for the accumulation of reads in the conserved and highly variable regions of the ribosomal rRNA genes could be the presence of several *T. denticola* strains in the metagenome. However, we only see this read accumulation in the two 23S rRNA gene copies and not in the 16S rRNA genes, where we would have expected a similar but less pronounced effect, due to the shorter length of the 16S rRNA gene. Thus for phylogenetic assignment we decided to use a contiguous consensus sequence of the 23S rRNA gene to verify the presence of *Treponema* sequence reads in the reads mapped to the 23S rRNA gene. Furthermore, we performed a DNA damage pattern analysis on all obtained *Treponema* reads and compared the pattern to nucleotide misincorporation frequencies in the Iceman genomic reads. In contrast to the reads of the modern human reference genome both *Treponema* reads and Iceman genomic reads display increased frequency of C to T substitutions towards the fragment ends. It is important to note that the DNA library of the Iceman genomic survey [15], similar to the library of the Shaqqaq genome study [14], was PCR amplified with a Phusion Polymerase. This is a modified Pfu Polymerase [47], which has been demonstrated to

show poor activity at uracil and/or deaminated cytosine residues [48,49]. Therefore, the observed C to T misincorporation events in the Iceman genomic reads presumably indicate both conserved ancient methylation patterns as these have been recently reported in the Shaqqaq genomic reads [50], and diminished yet detectable levels of cytosine deamination [20]. Unexpectedly, the observed nucleotide misincorporation frequencies are much higher in the *T. denticola* reads than in the Iceman genomic reads. A similar quantitative effect has been reported in a recent ancient genome study on *Mycobacterium leprae*, the causative agent of leprosy [18]. In contrast to our observation in the *Treponema* study the DNA damage patterns in the leprosy study were much less pronounced in the bacterium than in the human host. The observed different nucleotide misincorporation frequencies require further investigation having more enriched *T. denticola* sequence data available. Nevertheless, the obtained *T. denticola* DNA sequences from our study indicates an already increased frequency of C to T substitutions close to the fragment ends characteristic of ancient DNA.

The detection of metagenomic reads specific for an opportunistic oral pathogen in the Iceman bone biopsy was unexpected and we provide below different explanatory models for this finding. One possible explanation for the presence of *T. denticola* reads in the bone biopsy could be the haematogenous spread of opportunistic oral pathogens as reported in recent literature [51,52]. In addition, *T. denticola* and other members of the human commensal oral microflora seem to be associated with the formation of atherosclerotic plaques [53,54]. Interestingly, the Iceman shows strong signs of generalized arteriosclerotic disease by the observation of several calcified plaques as previously revealed by CT scan analysis [55] (Figure S3). Since sampling of a plaque would be invasive, we decided to extend our survey to the mouth region of the Iceman, the actual living environment of opportunistic oral pathogens. Using a PCR-based assay, we detected the DNA of opportunistic oral pathogens in an Iceman's gingival tissue and swab sample. Our results indicate the presence of bacteria of the oral microbiome in different Iceman tissue biopsies. The detection of opportunistic oral pathogens in the Iceman's mouth region is in accordance with the findings of Adler and colleagues [56] and Warinner and colleagues [57], who could demonstrate the presence of oral microbiota DNA in numerous ancient calcified human dental plaques. Both *T. denticola* and *P. gingivalis* belong, together with *Tannerella forsythia*, to the so-called "red complex" bacteria which are members of the dental plaque biofilm community and which are highly associated with periodontal disease [58–60]. Periodontitis is the major cause of tooth loss worldwide. Bacterially induced chronic inflammatory processes can result in localized alveolar bone loss around the tooth root surfaces [58]. Recently a re-evaluation of CT scans of the Iceman focusing on the oral cavity could show extensive alveolar bone loss [61] indicative of periodontitis. Thus, our molecular data is in accordance with the CT-based results. Another possible explanation for the presence of *T. denticola* reads in the bone biopsy is beside the above mentioned haematogenous spread of bacteria during lifetime the dissemination of commensal microflora via the bloodstream around the body shortly before death. However, with the currently available data we cannot determine whether the spread of these opportunistic pathogens occurred during the Iceman's lifetime or shortly before death.

Taken together, our data indicates the presence of bacteria of the oral microbiome in different Iceman tissue biopsies. Moreover, the obtained *T. denticola* DNA displays damage patterns characteristic for ancient DNA. The detection of the low-GC bacterium *T. denticola* may pave the way for future genome studies of

important ancient outright pathogens within the genus *Treponema* such as *Treponema pallidum*, the causative agent of syphilis. Even more importantly, this study underscores the opportunity to detect disease-associated microorganisms when applying metagenomics-enabled approaches on datasets from ancient human remains.

Supporting Information

Table S1   Detailed list of all oligonucleotide primers used in this study and the corresponding PCR conditions.
(DOCX)

Table S2   Detailed list of the gene coverage of all validated *Treponema denticola* reads.
(DOCX)

Figure S1   Number of Iceman metagenomic reads specifically mapped to all available genomes of the genus *Treponema*.
(TIF)

Figure S2   (A) Iceman's mouth region. Samples have been taken from the Iceman's right oral cavity. (B) A gingival soft tissue sample (1) and a mouth swab sample (2) have been taken.
(TIF)

Figure S3   CT image of the Iceman's abdomen. The arrows highlight two calcifications constituting aortic plaques in the aortic bifurcation.
(TIF)

Appendix S1   Supplementary Material and Methods, *In silico* analysis of the Iceman's metagenome.
(DOCX)

Author Contributions

Conceived and designed the experiments: FM AT SW TR AZ. Performed the experiments: FM AT GC. Analyzed the data: FM AT TR. Wrote the paper: FM AT TR AZ.

References

1. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456: 53–59.
2. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376–380.
3. Metzker ML (2010) Sequencing technologies - the next generation. Nat Rev Genet 11: 31–46.
4. Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, et al. (2012) Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011. Proc Natl Acad Sci U S A 109: 3065–3070.
5. Li R, Fan W, Tian G, Zhu H, He L, et al. (2010) The sequence and de novo assembly of the giant panda genome. Nature 463: 311–317.
6. Mayer KF, Waugh R, Brown JW, Schulman A, Langridge P, et al. (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491: 711–716.
7. Meyerson M, Gabriel S, Getz G (2010) Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet 11: 685–696.
8. Cho I, Blaser MJ (2012) The human microbiome: at the interface of health and disease. Nat Rev Genet 13: 260–270.
9. Green RE, Briggs AW, Krause J, Prufer K, Burbano HA, et al. (2009) The Neandertal genome and ancient DNA authenticity. EMBO J 28: 2494–2502.
10. Kirsanow K, Burger J (2012) Ancient human DNA. Ann Anat 194: 121–132.
11. Schatz MC, Delcher AL, Salzberg SL (2010) Assembly of large genomes using second-generation sequencing. Genome Res 20: 1165–1173.
12. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, et al. (2010) A draft sequence of the Neandertal genome. Science 328: 710–722.
13. Reich D, Green RE, Kircher M, Krause J, Patterson N, et al. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468: 1053–1060.
14. Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, et al. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 463: 757–762.
15. Keller A, Graefen A, Ball M, Matzas M, Boisguerin V, et al. (2012) New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. Nat Commun 3: 698.
16. Stoneking M, Krause J (2011) Learning about human population history from ancient and modern genomes. Nat Rev Genet 12: 603–614.
17. Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, et al. (2011) A draft genome of Yersinia pestis from victims of the Black Death. Nature 478: 506–510.
18. Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jager G, et al. (2013) Genome-wide comparison of medieval and modern Mycobacterium leprae. Science 341: 179–183.
19. Ames SK, Hysom DA, Gardner SN, Lloyd GS, Gokhale MB, et al. (2013) Scalable metagenomic taxonomy classification using a reference genome database. Bioinformatics.
20. Ginolhac A, Rasmussen M, Gilbert MT, Willerslev E, Orlando L (2011) mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics 27: 2153–2155.
21. Pruefer K, Stenzel U, Hofreiter M, Paabo S, Kelso J, et al. (2010) Computational challenges in the analysis of ancient DNA. Genome Biol 11: R47.
22. Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, et al. (2006) Analysis of one million base pairs of Neanderthal DNA. Nature 444: 330–336.
23. Zaremba-Niedzwiedzka K, Andersson SG (2013) No ancient DNA damage in Actinobacteria from the Neanderthal bone. PLoS One 8: e62799.
24. Rumble SM, Lacroute P, Dalca AV, Fiume M, Sidow A, et al. (2009) SHRiMP: accurate mapping of short color-space reads. PLoS Comput Biol 5: e1000386.
25. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35: 7188–7196.
26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
27. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. Genome Res 21: 1552–1560.
28. Wheeler DL, Chappey C, Lash AE, Leipe DD, Madden TL, et al. (2000) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 28: 10–14.
29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.
30. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J (2009) DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 25: 119–120.
31. Ludwig W, Strunk O, Westram R, Richter L, Meier H, et al. (2004) ARB: a software environment for sequence data. Nucleic Acids Res 32: 1363–1371.
32. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, et al. (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41: D590–596.
33. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704.
34. Tang JN, Zeng ZG, Wang HN, Yang T, Zhang PJ, 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.
35. Martinez-Pabon MC, Restrepo-Ospina DP, Isaza-Guzman DM, Orozco-Rojas LM, Tobon-Arroyave SI (2008) Detection of Treponema denticola in saliva obtained from patients with various periodontal conditions. Clin Oral Investig 12: 73–81.
36. David M, Dzamba M, Lister D, Ilie L, Brudno M (2011) SHRiMP2: sensitive yet practical SHort Read Mapping. Bioinformatics 27: 1011–1012.
37. Norris SJ, Paster BJ, Moter A, Goebel UB (2006) The Genus Treponema. In: Balows A, Trüper HG, Dworkin M, Harder W and Schleifer KH (eds)The Prokaryotes Springer: New York 7: 211–234.
38. NCBI_Resource_Coordinators (2013) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 41: D8–D20.
39. Yu K, Zhang T (2012) Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. PLoS One 7: e38183.
40. Santamaria M, Fosso B, Consiglio A, De Caro G, Grillo G, et al. (2012) Reference databases for taxonomic assignment in metagenomics. Brief Bioinform 13: 682–695.

41. Cano RJ, Tiefenbrunner F, Ubaldi M, Del Cueto C, Luciani S, et al. (2000) Sequence analysis of bacterial DNA in the colon and stomach of the Tyrolean Iceman. Am J Phys Anthropol 112: 297–309.

42. Rollo F, Luciani S, Canapa A, Marota I (2000) Analysis of bacterial DNA in skin and muscle of the Tyrolean iceman offers new insight into the mummification process. Am J Phys Anthropol 111: 211–219.

43. Khairat R, Ball M, Chang CC, Bianucci R, Nerlich AG, et al. (2013) First insights into the metagenome of Egyptian mummies using next-generation sequencing. J Appl Genet 54: 309–325.

44. Pinar G, Piombino-Mascali D, Maixner F, Zink A, Sterflinger K (2013) Microbial survey of the mummies from the Capuchin Catacombs of Palermo, Italy: biodeterioration risk and contamination of the indoor air. FEMS Microbiol Ecol 86: 341–356.

45. Daldrup T, Huckenbeck W (1984) Significance of the putrefactive bacterium Clostridium sordellii for the determination of age of the cadaver. Z Rechtsmed 92: 121–125.

46. Janisch S, Gunther D, Fieguth A, Bange FC, Schmidt A, et al. (2010) Post-mortal detection of clostridia–putrefaction or infection? Arch Kriminol 225: 99–108.

47. Wang Y, Prosen DE, Mei L, Sullivan JC, Finney M, et al. (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. Nucleic Acids Res 32: 1197–1207.

48. Greagg MA, Fogg MJ, Panayotou G, Evans SJ, Connolly BA, et al. (1999) A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. Proc Natl Acad Sci U S A 96: 9045–9050.

49. Fogg MJ, Pearl LH, Connolly BA (2002) Structural basis for uracil recognition by archaeal family B DNA polymerases. Nat Struct Biol 9: 922–927.

50. Pedersen JS, Valen E, Velazquez AM, Parker BJ, Rasmussen M, et al. (2014) Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. Genome Res 24: 454–466.

51. Inaba H, Amano A (2010) Roles of oral bacteria in cardiovascular diseases–from molecular mechanisms to clinical cases: Implication of periodontal diseases in development of systemic diseases. J Pharmacol Sci 113: 103–109.

52. Ohki T, Itabashi Y, Kohno T, Yoshizawa A, Nishikubo S, et al. (2012) Detection of periodontal bacteria in thrombi of patients with acute myocardial infarction by polymerase chain reaction. Am Heart J 163: 164–167.

53. Cavrini F, Sambri V, Moter A, Servidio D, Marangoni A, et al. (2005) Molecular detection of Treponema denticola and Porphyromonas gingivalis in carotid and aortic atheromatous plaques by FISH: report of two cases. J Med Microbiol 54: 93–96.

54. Okuda K, Ishihara K, Nakagawa T, Hirayama A, Inayama Y (2001) Detection of Treponema denticola in atherosclerotic lesions. J Clin Microbiol 39: 1114–1117.

55. Murphy WA, Jr., Nedden Dz D, Gostner P, Knapp R, Recheis W, et al. (2003) The iceman: discovery and imaging. Radiology 226: 614–629.

56. Adler CJ, Dobney K, Weyrich LS, Kaidonis J, Walker AW, et al. (2013) Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. Nat Genet 45: 450–455, 455 e451.

57. Warinner C, Rodrigues JF, Vyas R, Trachsel C, Shved N, et al. (2014) Pathogens and host immunity in the ancient human oral cavity. Nat Genet. doi:10.1038/ng.2906.

58. Darveau RP (2010) Periodontitis: a polymicrobial disruption of host homeostasis. Nat Rev Microbiol 8: 481–490.

59. Griffen AL, Beall CJ, Campbell JH, Firestone ND, Kumar PS, et al. (2012) Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. ISME J 6: 1176–1185.

60. Hormozdiari F, Hach F, Sahinalp SC, Eichler EE, Alkan C (2011) Sensitive and fast mapping of di-base encoded reads. Bioinformatics 27: 1915–1921.

61. Seiler R, Spielman AI, Zink A, Ruhli F (2013) Oral pathologies of the Neolithic Iceman, c.3,300 BC. Eur J Oral Sci 121: 137–141.