

RESEARCH ARTICLE

Ancient DNA from latrines in Northern Europe and the Middle East (500 BC–1700 AD) reveals past parasites and diet

Martin Jensen S \ddot{o} e^{1,2*}, Peter Nejs \ddot{u} m³, Frederik Valeur Seersholm², Brian Lund Fredensborg¹, Ruben Habraken⁴, Kirstine Haase^{5,6}, Mette Marie Hald⁷, Rikke Simonsen⁸, Flemming H \ddot{o} jlund⁹, Louise Blanke¹⁰, Inga Merkyte¹¹, Eske Willerslev^{2,12,13}, Christian Moliin Outzen Kapel^{1*}

1 Department of Plant and Environmental Sciences, University of Copenhagen, Frederiksberg, Denmark, **2** Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen K, Denmark, **3** Department of Veterinary Disease Biology, University of Copenhagen, Frederiksberg, Denmark, **4** BioArchaeological Research Bureau, Den Haag, The Netherlands, **5** Odense Bys Museer, Odense, Denmark, **6** Centre for Urban Network Evolutions, School of Culture and Society, Aarhus University, H \ddot{o} jbjerg, Denmark, **7** Environmental Archaeology and Materials Science, National Museum of Denmark, Kgs. Lyngby, Denmark, **8** Museum of Copenhagen, Copenhagen V, Denmark, **9** Moesgaard Museum, H \ddot{o} jbjerg, Denmark, **10** Department of Archaeology, School of Culture and Society, Aarhus University, Aarhus, Denmark, **11** The Saxo Institute, University of Copenhagen, Copenhagen S, Denmark, **12** Department of Zoology, University of Cambridge, Downing St, Cambridge, United Kingdom, **13** Sanger Institute, Hinxton, Cambridge, United Kingdom

* martin.j.soe@gmail.com (MJS); chk@plen.ku.dk (CMOK)



OPEN ACCESS

Citation: S \ddot{o} e MJ, Nejs \ddot{u} m P, Seersholm FV, Fredensborg BL, Habraken R, Haase K, et al. (2018) Ancient DNA from latrines in Northern Europe and the Middle East (500 BC–1700 AD) reveals past parasites and diet. PLoS ONE 13(4): e0195481. <https://doi.org/10.1371/journal.pone.0195481>

Editor: Dong Hoon Shin, Seoul National University College of Medicine, REPUBLIC OF KOREA

Received: October 31, 2017

Accepted: March 24, 2018

Published: April 25, 2018

Copyright: © 2018 S \ddot{o} e et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available through GenBank, accession numbers: KY368754 to KY368775.

Funding: This work was funded through The University of Copenhagen to M.J.S., and all laboratory expenses were funded through the KU2016 initiative, 'The Genomic History of Denmark'. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

High-resolution insight into parasitic infections and diet of past populations in Northern Europe and the Middle East (500 BC–1700 AD) was obtained by pre-concentration of parasite eggs from ancient latrines and deposits followed by shotgun sequencing of DNA. Complementary profiling of parasite, vertebrate and plant DNA proved highly informative in the study of ancient health, human-animal interactions as well as animal and plant dietary components. Most prominent were finding of soil-borne parasites transmitted directly between humans, but also meat-borne parasites that require consumption of raw or undercooked fish and pork. The detection of parasites for which sheep, horse, dog, pig, and rodents serves as definitive hosts are clear markers of domestic and synanthropic animals living in closer proximity of the respective sites. Finally, the reconstruction of full mitochondrial parasite genomes from whipworm (*Ascaris lumbricoides*) and roundworm species (*Trichuris trichiura* and *Trichuris muris*) and estimates of haplotype frequencies elucidates the genetic diversity and provides insights into epidemiology and parasite biology.

Introduction

Human infections with intestinal worms were widespread in Europe until the middle of the last century, but are today primarily confined to rural communities in developing countries with low levels of sanitation, insufficient water refinement, and where humans and animals

Competing interests: The authors have declared that no competing interests exist.

live in close proximity. Under such settings, the giant roundworm *Ascaris lumbricoides* and the human whipworm *Trichuris trichiura* are still found at high prevalence. These two soil-transmitted parasites have been widely identified in archaeological samples associated with human remains and coprolites in areas where they are no longer endemic [1].

Traditional palaeoparasitological methods rely on morphological examinations of recovered parasite eggs. While a few species may reliably be identified using microscopy, others can only be identified to genus level due to overlapping morphological characteristics of the eggs of closely related species [2,3]. This prevents further elucidation of specific host-parasite associations. Contrary, DNA based analysis may identify parasites at the species level and, as most parasites are host specific with distinct life cycles, they can serve as markers for the presence of their intermediate and/or definitive hosts. This is particularly the case for the *Trichuris* genus, which comprises more than 10 distinct species with narrow host specificity where e.g. *T. trichiura* infects humans, *T. suis* infects pigs and *T. muris* infects mice [4–7]. *Trichuris* spp., as well as *Ascaris* spp., exhibits direct life cycles where worms residing in the host intestine, excrete eggs to the environment with host faeces. Transmission to a new host occurs through ingestion of infective eggs from the environment or through contaminated foods. High host specificity of *T. trichiura* along with the global distribution and wide detection of eggs in archaeological samples makes this parasite a suitable indicator of human migrations [8]. This is supported by a recent study on populations of modern *T. trichiura* isolates, which suggested that worms were distributed with human migration first from Africa to Asia and then further to South America whereas *T. suis* might have been distributed through a combination of pig migration and transport of domesticated pigs,—and their parasites [9]. *Ascaris* in humans (*A. lumbricoides*) and pigs (*A. suum*) are very closely related and cannot be distinguished based solely on mitochondrial DNA markers [10]. However, as *Ascaris* and *Trichuris* most often are co-detected in latrine samples and human intestinal contents from Europe and the Middle East [11–15], we use the term *A. lumbricoides* to describe *Ascaris* based on DNA findings when co-identified with *T. trichiura*.

Identifying ancient parasites, which exhibit indirect life cycles involving an animal as intermediate host and humans as definitive hosts, may reveal our ancestors' dietary preferences and animal exploitation history. Tapeworms, such as *Taenia solium* and *T. saginata*, exemplify this, as tissue cysts are transmitted to humans by consumption of undercooked pork or beef, respectively. Notably, in ancient Greenlandic kitchen middens only the tapeworm species *T. hydatigena*, *T. multiceps* and *E. canadensis* were identified by DNA analysis, inferring the presence of canid and ungulate hosts, typically sheep and reindeer, and thereby indicating human resource economy [16].

Currently, the implementation of DNA based analysis in palaeoparasitology is limited to first-generation sequencing of PCR amplicons [17–19], a next-generation barcoding approach for identifying 8 genera of helminths [20] and our recent publication on the resource economy in ancient Greenland [16]. Here we extend the shotgun DNA sequencing analysis on ancient parasite eggs; identify parasite infections to the species level, use them as markers of diet and specifically as markers for which animals who lived in the immediate proximity of the sample point. We identify vertebrate DNA, for which the majority of taxa are domesticated animals, typically livestock, and animals associated with hunting or fishing practices. Diet is further addressed through plant DNA with focus on edible plants. Finally, shotgun sequencing has allowed us to reconstruct multiple ancient mitochondrial genomes of *A. lumbricoides*, *T. trichiura* and *T. muris*, analyse their phylogenies and estimate the number of haplotypes present in a sample. The approach shows obvious potential for future reconstructions of nuclear genomes from ancient parasites.

Methods

Samples and egg extraction

Samples were collected from archaeological excavations of human latrines and deposits with high levels of organic material. Initially, samples were screened for the presence of *Ascaris* spp. or *Trichuris* spp. eggs to identify human fecal material, by dissolving 10 gram of sample in 50 mL tubes filled with flotation buffer (FB) (glucose monohydrate 375 g/L + sodium chloride 250 g/L). Tubes were centrifuged (15–100 g, 5 min) and the top 1 mL of supernatant was transferred to McMaster counting chambers and examined with a microscope at 100/400x magnification. Samples chosen for further processing are illustrated in Fig 1, while specimen numbers are listed in S1 Table. No permits were required for the described study, which complied with all relevant regulations. The Danish samples were taken from excavations carried out within the Danish Museum Act, which includes the undertaking of scientific sampling and analysis by non-written agreement with the local archaeologists. Permit for analysis of the Jordanian samples is included in the export permit of the samples by the Amman and Jarash offices of the Department of Antiquities. The Dutch and Lithuanian samples were analysed at the discretion of the local archaeologists.

From each sample, 75–503 g (average 183 g) were processed by selecting particles based on flotation in high density liquid and size, as previously described [19]. In brief, samples were suspended in FB and centrifuged. The sediment was discarded after wet sieving the supernatant on stacked 100 and 22.4 µm filters. The material collected from the 22.4 µm filter was re-suspended in FB, centrifuged and supernatant wet sieved on stacked 35.5 and 22.4 µm filters. The samples were collected and termed ‘sample ID’ followed by ‘A’ for 35.5–100 µm particles and ‘B’ for 22.4–35.5 µm particles, unless otherwise specified. 10% of the sample was subjected to morphological examination and the remaining 90% for DNA analysis. Morphological

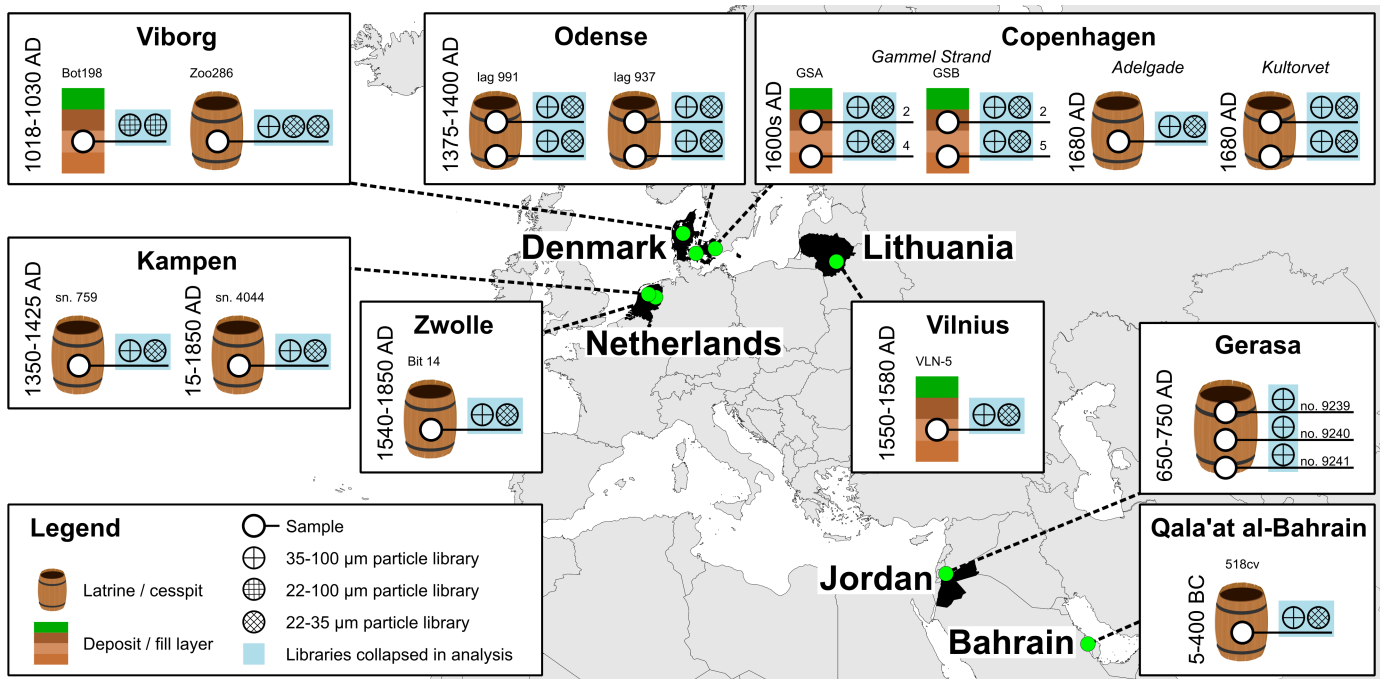


Fig 1. Sampling locations for archaeologically classified sample types. Samples were processed for particle size selection by wet sieving on filters of designated sizes. Individual DNA libraries were prepared for the illustrated combinations of sample and particle sizes, then sequenced and collapsed to sample level for further analysis. The 22–35 µm particle isolates of the Jordan, Gerasa samples were devoid of any particles and not processed for sequencing.

<https://doi.org/10.1371/journal.pone.0195481.g001>

examination was performed on microscope slides at 100x and 400x magnification and quantification of eggs performed either on fixed microscope slides or in McMaster counting chambers. The reported 'number of eggs in sample' was extrapolated from egg count and proportion of total analysed sample. All egg extractions were carried out in a dedicated palaeoparasitological laboratory at the Department of Plant and Environmental Sciences (PLEN), University of Copenhagen.

DNA sequencing

Filtered samples, (0.02 g–0.58 g, average of 0.11 g) were processed for DNA extraction and library preparation in dedicated aDNA laboratories at the Centre for GeoGenetics (CGG), Natural History Museum, University of Copenhagen, in accordance with strict aDNA-specific requirements as previously described [19], although eluted in 60 µL elution buffer. Blunt end DNA libraries were prepared using NEBNext DNA Sample Prep Master Mix Set for 454 (E6070) and Illumina-specific adapters [21] following established protocols [22,23], hence without the ssDNA isolation module. The following exceptions to the protocols were made: intermittent reaction clean-ups were performed using the MinElute PCR purification kit (Qiagen), with an improved binding buffer that has proved highly efficient in recovering very short DNA fragments as described elsewhere [23]; adaptors were used at final concentration of 0.5 µM and the fill-in reaction was performed for 20 min at 60° C and inactivated for 20 min at 80° C. DNA libraries were amplified using a nested PCR approach in a first round reaction of 50 µL for 12 PCR cycles, 5 µL of product was used as template in the second reaction run for 10–16 cycles. PCR cycling and post-PCR handling was performed in modern DNA laboratories physically separated from the aDNA laboratories. Second round PCR's were examined on a 2% agarose gel, purified using the MinElute kit following manufacturer's instructions, then quantified on a Qubit 2.0 using the dsDNA HS kit (Thermo Fischer) and finally on a Bioanalyzer (Agilent Technologies) using the Agilent High Sensitivity DNA kit. For libraries not visible on the gel, or if the concentration after purification was <5 nM, the second round PCR was repeated with a higher number of cycles (up to 16). Purified libraries were pooled at concentrations of 5–20 nM before shotgun sequencing using 100-bp single read chemistry on a HiSeq 2000/2500 platform at The Danish National High-Throughput DNA Sequencing Centre. Negative controls were prepared for egg extraction, library preparation and PCR amplification steps and processed along with samples in the subsequent steps. Extraction controls were sequenced along with samples and processed together in subsequent analysis.

Pre-processing sequence reads

Sequencing data was base called using Illumina software CASAVA 1.8.2 and sequences were de-multiplexed with a requirement of full match for the 6-nucleotide index that was used for library preparation. Sequences were trimmed using Adapter Removal v2 [24] to remove adapter sequences and stretches of mixed low quality bases and/or Ns from both the 5' and 3' end as well as discarding reads shorter than 30 nt. Sequence data was merged to library level and an additional filtering of low-quality reads was performed using the SGA-String Graph Assembler [25] *preprocess* command, option *dust-threshold* set to 3. Exact match read duplicates were removed, by first indexing the reads using *sga index* and then filtering using *sga filter* with the *no-kmer-check* disabled.

Identifying helminth, vertebrate and plant DNA

The 'lowest common ancestor' (LCA) approach as described in detail elsewhere [16], was used to identify hits to mitochondrial and plastid genomes (<https://github.com/frederikseersholm/>

getLCA). The mitochondrion reference database was acquired from NCBI (<ftp://ftp.ncbi.nlm.nih.gov/refseq/release/mitochondrion/>) and supplemented with *Trichuris* sp. sequences (KC461179, KT449822, KT449823, KT449826) and *Ascaris* sp. sequences (KY045800, KY45801, KY045804 and KY045805) to better encompass sequence variation in these genera. Similarly the plastid reference database was obtained from <ftp://ftp.ncbi.nlm.nih.gov/refseq/release/plastid/>. Taxonomic IDs were added to each fasta header in the databases using the NCBI file `gi_taxid_nucl.dmp` (ftp.ncbi.nlm.nih.gov/pub/taxonomy/gi_taxid_nucl.dmp.gz), `names.dmp` and `nodes.dmp` (<ftp.ncbi.nlm.nih.gov/pub/taxonomy/taxdump.tar.gz>), and custom scripts [16]. Sequences were mapped to the two databases using bowtie2 version 2.2.4 [26], reporting up to 500 alignments per read. The LCA algorithms (<https://github.com/frederikseersholm/getLCA>) were applied to the sam output files and the lowest common ancestor from the best hits for each read was returned, as in Seersholm et al (2016) [16]. Briefly, each read is assigned to the lowest common ancestor of the best hits to the database, based on edit distance. If a read has only a single best alignment the read is assigned directly to the taxa of that reference sequence. If the read has several equally good alignments (same edit distances), on the other hand, the read is assigned to the lowest common ancestor of the best hits, based on the NCBI taxonomy files `names.dmp` and `nodes.dmp` (<ftp.ncbi.nlm.nih.gov/pub/taxonomy/taxdump.tar.gz>). Reads identified as the best hit(s) were discarded if sequence identity to the reference was less than 95%.

Mitochondrial sequence hits are reported as ‘helminths’, which includes the ‘platyhelminthes’ and ‘nematoda’ phylum assignments, whereas ‘vertebrates’ includes ‘vertebrata’ sub-phylum assignments. Reads were mapped to sample level prior to reporting in Fig 2. The 11 most abundant helminth species of humans and animals are reported, while reads assigned to the *Ascaris* genus or the species, *A. lumbricoides* or *A. suum*, were reported as *A. lumbricoides*. The most abundant vertebrate taxa (species or genus) with more than 3 uniquely assigned reads are reported in Fig 2. Reads mapping to the following genus and species have been collapsed to genus level as follows: *Canis* contains *Canis*, *Canis lupus*, *Canis lupus familiaris* and *Canis lupus chanco*; *Equus* contains *Equus*, *Equus (ferus) przewalskii* and *Equus (ferus) caballus*; *Felis* contains *Felis*, *Felis silvestris*, *Felis catus* and *Felis margarita*; *Ovis* contains *Ovis* and *Ovis orientalis*.

The following taxa have been excluded from the reported vertebrate results: *Homo sapiens* and related primates; *Bos*, *Gallus* and *Sus* genera—observed many samples and DNA extraction controls and known to be common laboratory contaminants [27]; *Meleagris gallopavo* (Wild Turkey)—was observed in all DNA extraction controls and more so in highly degraded samples from the Middle East; *Pelodiscus sinensis* (Chinese Softshell Turtle), *Phelsuma guimbeaui* and *Phyllodactylus unctus* (both Geckos)—which were identified in many samples although with a maximum of 5 reads.

Plastid sequence hits are reported at the genus level, as the database in its current form is lacking many species. Genus level hits as shown in Fig 3, are collapsed as follows: *Vitis* contains reads assigned to *Vitis* genus and *Vitis vinifera* and *Vitis rotundifolia* species; *Humulus* contains *Humulus lupulus*; *Prunus* contains *Prunus* and *Prunus maximowiczii*; *Pyrus* contains *Pyrus* and *Pyrus spinosa*; *Pisum* contains *Pisum sativum*; *Fagopyrum* contains *Fagopyrum* and *Fagopyrum esculentum*; *Rheum* contains *Rheum palpatum*; *Lupinus* contains *Lupinus luteus*; *Lotus* contains *Lotus japonicus*; *Vaccinium* contains *Vaccinium macrocarpon*; *Hordeum* contains *Hordeum* and *Hordeum vulgare*. Species with reads mapping to more than 1 of 4 extraction controls and more than 2 hits in a single control are not reported in Fig 4. Full lists of helminth, vertebrate and plastids hits are reported in S3, S4 and S6 Tables.

Mitochondrial genome analysis

Reads assigned to the species *T. trichiura* and *T. muris* and the *Ascaris* genus by the LCA algorithm, were extracted using *seqtk* version 1.0-r31 and the *subseq* command. Reads were re-

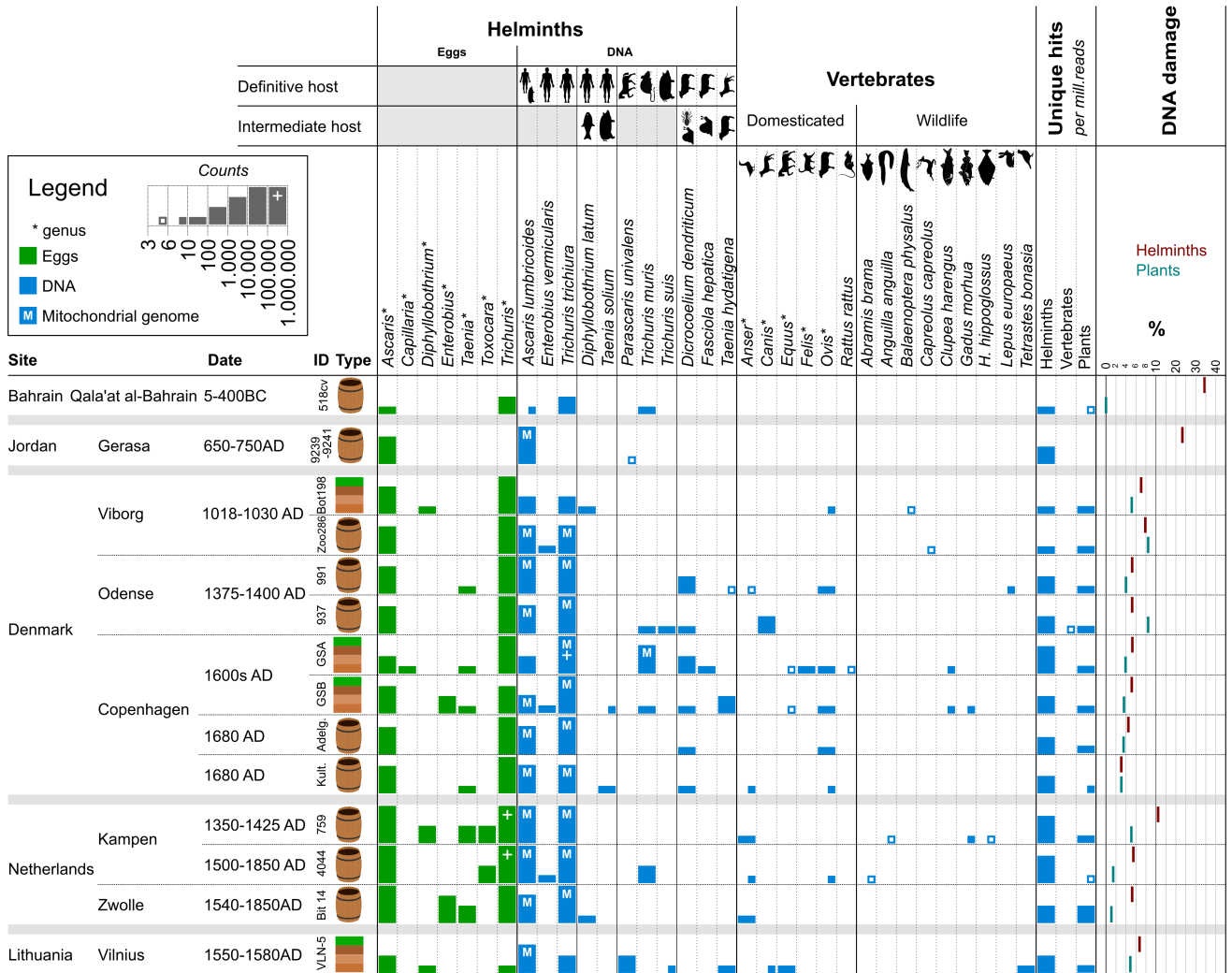


Fig 2. The most abundant taxa of helminths and vertebrates identified in the samples. Bars are logarithmic in size and show the estimated number of eggs in the sequenced sample and the number of DNA reads uniquely assigned to species and/or genus level. Helminths and vertebrates were identified based on sequence similarity to mitochondrial references, while plants were identified based on sequence similarity to plastid references. The typical hosts for the identified helminths are shown as the definitive host (*i.e.* the host where the parasite reproduces sexually and from which eggs are excreted) and intermediate hosts (*i.e.* host required for completion of the parasite life cycle). DNA damage is illustrated as the average frequency of C to T transition on the first 5' position and the G to A transition on the first 3' position of reads uniquely assigned to the helminth taxa presented here or the edible plant taxa presented in Fig 3.

<https://doi.org/10.1371/journal.pone.0195481.g002>

mapped to the *T. trichiura* sequence from humans in Uganda (KT449826), the *Ascaris* sp. sequence from haplotype 07 (PUG3-1-H07) or the *T. muris* sequence (NC_028621) using bow-tie2. Sam output was converted to bam format, merged to library level and sorted using command *sort* in *samtools* version 1.2 [28]. Alignment coverage and depths were calculated using *Paleomix* version 1.1.1, commands *coverage* and *depths* [29]. Mitochondrial consensus sequences were called using the commands *samtools mpileup*, *bcftools call* (version 1.2) and *vcftutils vcf2fq*.

The consensus sequences from *T. trichiura* and *T. muris* were aligned together with all mitochondrial genome sequence of *Trichuris* genera publicly available at NCBI, as was done for the *Ascaris* genus sequences. Certain reference sequences were re-arranged to start with the *cox1* gene. Alignment was performed with *mafft v7*[30], using the *globalpair* setting at 1000 iterations. A tree was built through 1000 bootstraps using the Neighbor Joining method and

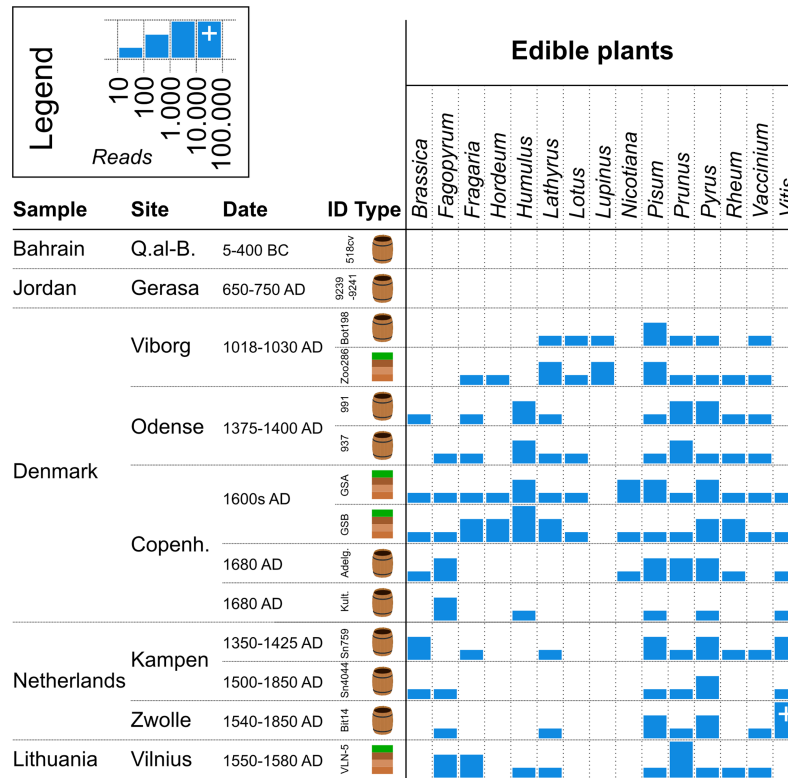


Fig 3. Edible plant identifications. The 15 most abundant plants, which are consumed by humans and identified in the samples are shown. Bars represent the number of uniquely assigned DNA reads, *i.e.* assigned to species and/or genus level, then collapsed to genus level. Only assignments with more than 10 reads and only samples for which the combined reads exhibit ancient DNA damage profiles are shown.

<https://doi.org/10.1371/journal.pone.0195481.g003>

Jukes-Cantor nucleotide distance model in the CLC Sequence Viewer 7 program, which was also used for visualization. The consensus mitochondrial genome sequences were annotated using Mitos WebServer (<http://mitos.bioinf.uni-leipzig.de/index.py>) and made publicly available (Genbank accession number will be inserted here).

DNA damage

Ancient DNA damage patterns were assessed using mapDamage v2 [31], for helminth, vertebrate and plastid identified hits. The analyses were performed as an average of all reads assigned accordingly (Fig 2 and S2–S6 Tables). For each library the 5' C to T transition percentage and the 3' G to A transition percentage were calculated and the average damage percentages are reported at the sample level.

Characterization of haplotypes

In order to identify potential SNPs, heterozygous bases were identified from pileup files of each mitochondrial genome assembly. Values for consensus coverage depend on the DNA sequencing error rate and the number of heterozygous bases of the genome analysed. However, as the number of heterozygous bases is markedly lower than the number of homozygous bases, the vast majority of data points will fall around the DNA sequencing error rate. Hence, with an error rate of 5% (sequencing and ancient DNA damage), values for consensus coverage for each position forms a distributions around the average of 95%. To distinguish potential

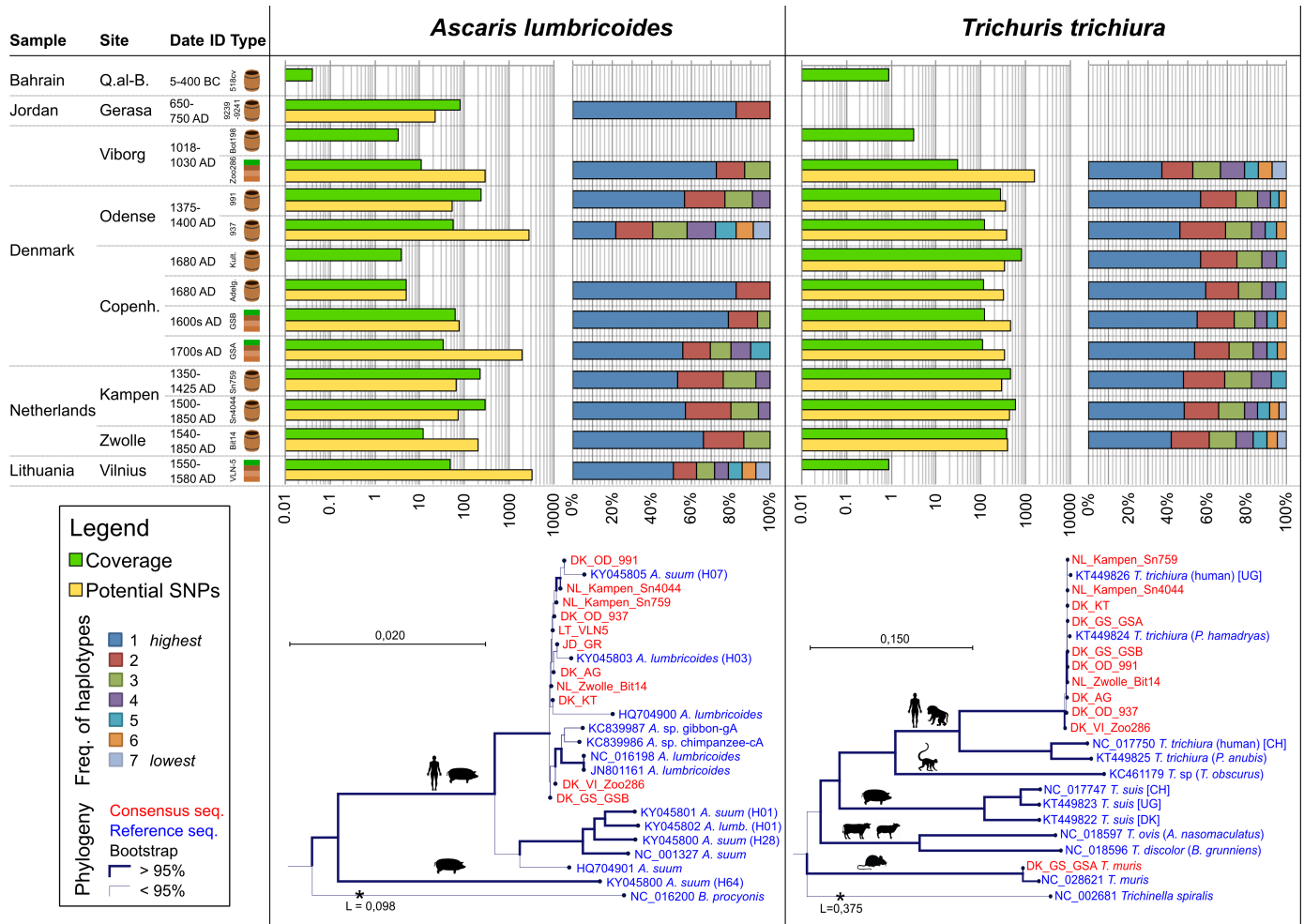


Fig 4. Sequence depth, genetic variation and phylogeny of *Ascaris lumbricoides* and *Trichuris trichiura* mitochondrial genomes. Top left: Depth of coverage and number of potential SNPs when mapping unique hits (Fig 2) to the closest related mitochondrial genomes of *Ascaris lumbricoides* (KY045805) and *Trichuris trichiura* (KT449826). Top right: Frequency of identified haplotypes. Average depth coverage of minimally 5–10x was required to estimate the number of potential SNPs and the frequency of haplotypes. Bottom: Phylogenies of consensus mitochondrial genomes and all published mitochondrial genomes. The *T. muris* consensus sequence was generated by remapping unique hits to NC_028621. The scale bar indicates number of base substitutions per site.

<https://doi.org/10.1371/journal.pone.0195481.g004>

SNPs from the background of the error rate, the average and spread of the consensus coverage data was calculated across each mitochondrial genome and only positions where the consensus coverage could be identified as outliers, i.e. values outside the main body of the data (<6* inter quartile range (IQR)) were used for the analysis. This threshold was also used to distinguish 'true haplotypes' from haplotypes that could be an effect of sequencing errors below.

Next, clusters of potential SNPs were characterized using a window size of the read length (100bp). Sam files were then parsed with respect to the SNP clusters identified: for each cluster, reads covering all potential SNPs of the given cluster were analysed, counting all different haplotypes at that particular SNP cluster. In order to avoid a potential bias from ancient DNA damage, the following bases did not count towards the final haplotype assessment: Thymidine bases within the first ten base pairs of a read where the consensus base was Cytosine and Adenine bases within the last ten bases of a read where the consensus base was Guanine.

To retain only high confidence clusters, clusters were excluded if they were covered by less than 10 reads or less than one fourth of the mean coverage of the mitochondrial genome

assembly. Furthermore, to filter out haplotypes that could be a result of sequencing errors, low abundance haplotypes constituting less than the threshold percentage described above ($<1-6 \times \text{IQR}$) were discarded. After filtering, the estimated (minimum) number of haplotypes for each mitochondrial genome was characterized, defined as the highest number of haplotypes identified in at least two different clusters. The fraction of each haplotype presented in Fig 4 was calculated as the mean fractions across the clusters where the estimated number of haplotypes were identified. This approach was benchmarked using shotgun sequencing data generated from single, modern *T. trichiura* worm isolates which all yield a single haplotype (data not shown).

Results

Samples apparently rich in organic material were collected from recent archaeological excavations and museum collections (S1 Appendix). Samples were screened for the presence of *Ascaris* spp. and *Trichuris* spp. eggs, typically associated with human faeces, of which 21 subsamples from 14 distinct locations were chosen for further processing (Fig 1). The majority of samples were characterized as latrines or cesspits (10/14) and the remaining as deposits or fill layers (4/14). Subsamples were concentrated for 22–35 and 35–100 μm particles by wet sieving; from which 1/10th was subjected to traditional morphological assessment of parasite eggs whereas the remaining material was used to generate 40 blunt-end DNA libraries. Shotgun sequencing generated a total of 1,062,232,288 DNA reads, which after rigorous pre-processing by removing adapter sequences, low complexity reads and PCR duplicates, yielded 839,575,890 reads (details in S1 and S2 Tables). Pre-processed reads were mapped against a mitochondrial reference database and assigned to the lowest common ancestor. After removal of hits to common laboratory contaminant in particular *Homo sapiens* and related primates as well as cattle (*Bos*), chicken (*Gallus*) and pig (*Sus*), which were present in most samples and in some controls, a dataset was compiled and used in the subsequent presentation. This dataset contains 605,635 unique hits (721 hits per million sequenced reads) to helminths (phylum *Nematoda* and *Platyhelminthes*) and 549 unique hits (0.6 per million) to vertebrates as well as 28,141 unique hits to edible plants when mapped against a plastid reference database (33 per million).

Helminths, vertebrates and edible plants

Taxa assignments, as identified by morphological examination and DNA assignment, were initially compared for individual particle size fractions. In general, parasites which excrete large eggs such as *A. lumbricoides*, *Fasciola hepatica*, *Parascaris univalens*, *Toxocara* spp., were identified in 35–100 μm particle size fractions, whereas parasites excreting smaller eggs, i.e. *Dicrocoelium dendriticum*, *Enterobius vermicularis*, *Taenia* spp. and *Trichuris* spp., were mainly identified in the 22–35 μm particle fractions. However, few discrepancies were noted, most often identification of parasites with small eggs among the larger particles (35–100 μm fraction), resulting from insufficient washing of filters.

Taxa assignments from different particle size fractions were then merged for the 14 samples for which the most abundant taxa are presented in Fig 2 (for complete lists see S3 and S4 Tables). Morphological examination identified 7 different helminth genera while DNA analysis identified 8 genera including 11 distinct species. DNA identified 78% of the morphologically identified taxa (32 of 41), with the 9 taxa missed by DNA being: *Capillaria* (1 sample, no reference sequence available), *Diphyllobothrium* (2), *Enterobius* (1), *Taenia* (3) and *Toxocara* (2). In contrast, DNA provided species level identifications for 6 different genera missed by morphological identification: *Dicrocoelium dendriticum* (6 samples), *Diphyllobothrium latum*

(1), *Enterobius vermicularis* (2), *Fasciola hepatica* (1), *Parascaris univalens* (2) and *Taenia hydatigena* (1), and additionally provided species assignment for two distinct *Taenia* species (*T. solium* and *T. hydatigena*) and three distinct *Trichuris* species (*T. trichiura*, *T. suis* and *T. muris*). DNA from both helminth and its typical definitive host were identified in a number of cases: *D. dendriticum* and *Ovis* (sheep) (identified in 4 of 5 cases), *F. hepatica* and *Ovis* (sheep) (1/1), *P. univalens* and *Equus* (horse) (1/2), *Taenia hydatigena* and *Canis* (dog) (1/3), *T. muris* and rodents (*Rattus rattus* (Black rat)) (1/5). Ancient DNA damage profiles are shown for helminth identified reads (Fig 2). Damage profiles for vertebrate identified reads could not be determined due to the very low number of reads.

The most abundant edible plants show a large diversity of taxa present in samples from the Northern European areas (Fig 3 and S6 Table). Identified plant taxa exhibit ancient DNA damage profiles that are comparable with those presented for the helminth findings, although lower rates were observed for the two youngest samples from the Netherlands (Kamper 4044 and Zwolle) (Fig 2.).

Genetic diversity and phylogenetic relationships

Reads assigned to *A. lumbricoides*, *T. muris* and *T. trichiura* were extracted and remapped (as the database contained multiple genomes of each species) to their closest reference mitochondrial genomes for further analysis. This yielded 11 samples with more than 90% bases covered for *A. lumbricoides* (average 96.6%), 10 samples for *T. trichiura* (average 100%) and one sample for *T. muris* (98.4%) (Fig 4, S3 and S4 Tables). Average read depth of coverage was 96x for *A. lumbricoides* genomes (range 5x–284x), 303x for *T. trichiura* genomes (range 30x–818x) and 7x for *T. muris*. In particular *A. lumbricoides* genome coverage was low in areas of low sequence complexity (i.e. immediately up/downstream of bp 2860, 6420, 6700, 7500, 7800, 12020). This is possibly a result of: 1) the rigorous pre-processing of sequence data or 2) that these regions are conserved between closely related species, resulting in read assignment to family level or higher. In contrast, the *T. trichiura* mitochondrial genome coverage only exhibited one problematic region (i.e. AT repeats up/downstream of bp 3230).

The number of potential SNPs, identified as positions with low consensus coverage, was estimated for all reconstructed mitochondrial genomes. *A. lumbricoides* sequences exhibited some very high (Lithuania—VLN5 = 3205) and very low (Denmark, Gammel Strand B = 5, Jordan Gerasa = 22) estimated values, while averaging 792 potential SNPs across all samples. In contrast, the number of potential SNPs in *T. trichiura* sequences ranged 297–457 when excluding the Zoo286 sample from Viborg, Denmark (1605 SNPs), averaging 492 potential SNPs across all samples.

The number of potential haplotypes was then counted in all samples, based on clusters of potential SNPs separated by no more than 100 bp (read length). This provides a conservative estimate, as it is restricted to haplotypes of SNPs spanned by a single sequencing read. Various thresholds were implemented to distinguish 'true haplotypes' from those originating from potential sequencing errors or DNA damage. Frequencies of haplotypes were estimated, with *T. trichiura* exhibiting a more uniform distribution than *A. lumbricoides*, as well as a higher number of identified haplotypes 6.0 (ranging 5–7) compared to 4.0 (ranging 2–7) for *A. lumbricoides*. To rule out that the estimated number of potential SNPs and haplotypes were influenced by the number of eggs in the sample or the sequencing depth, the following linear regression analysis were performed showing no statistical correlations: the number of eggs and number of potential SNPs in a sample ($R^2 = 0.11$ —*A. lumbricoides*; 0.02—*T. trichiura*), the number of sequencing reads and number of potential SNPs ($R^2 = 0.41$ —*A. lumbricoides*; 0.04—*T. trichiura*), the number of potential SNPs and average depth of coverage ($R^2 = 0.10$ —*A.*

lumbricoidea; 0.14—*T. trichiura*) and the number of haplotypes and average depth of coverage ($R^2 = 0.003$ —*A. lumbricoidea*; 0.04—*T. trichiura*).

Individual haplotype sequences could not be called as a result of the short read length; hence consensus sequences were called to investigate phylogenetic affiliations (Fig 4). *A. lumbricoidea* genomes group together with haplotypes H03 and H07 in a cluster which elsewhere is designated cluster B[10]. *T. trichiura* sequences grouped with reference sequence from human isolates from Uganda, distinctly different to the other *T. trichiura* cluster observed in humans from China.

Discussion

Using a novel approach of applying shotgun sequencing on ancient parasite eggs that have been purified by filtering, we have obtained a new and much more detailed insight into parasitic infections of human populations of the past. In addition, the complementary profiling of helminth, vertebrate and plant DNA proved highly informative in the study of ancient health, human-animal interactions as well as the animals and plants they exploited. Finally, the reconstruction of full mitochondrial parasite genomes from *Ascaris lumbricoidea*, *Trichuris trichiura* and *Trichuris muris*, elucidated the genetic diversity and provided insights into epidemiology and parasite biology.

We identified helminths which are directly transmitted between humans (*A. lumbricoidea*, *E. vermicularis* and *T. trichiura*) by faecal contamination of the immediate environment, but also human helminths with indirect transmission where fish (*D. latum*) and pig (*T. solium*) are required as intermediate host, indicating human consumption of raw or undercooked fish and pork. The identification of parasites for which humans are not the definitive host e.g. *D. dendriticum* (sheep definitive host), *F. hepatica* (sheep), *P. univalens* (horse), *T. hydatigena* (dog), *T. muris* (rodent) and *T. suis* (pig) indicates that faeces from these domestic and synanthropic animals are present in the samples and that they were living in proximity to humans at the respective sites. Notably, we identified two species of *Trichuris* that do not infect humans, *T. suis* from pigs and *T. muris* from rodents, which most likely would have been wrongly assigned to *T. trichiura* using traditional morphological approaches [1,14,15]. Similarly, DNA identification of *T. hydatigena* eggs is interesting as it reveals the presence of a canid definitive host. Using classical methodology, such eggs would very likely have been assigned to human infection with *T. solium*, as also found in the present study, or *T. saginata* after consumption of pork or beef, respectively [14,15]. Microscopy did identify *Capillaria* eggs (not in reference database) in one sample and *Toxocara* eggs in two samples that were not identified in the DNA analysis.

We acknowledge that the taxonomic status of *A. lumbricoidea* and *A. suum* is contentious as close similarities in mitochondrial and single nuclear markers have been identified in isolates from pigs and humans, including shared genotypes and haplotypes [10,32]. In spite of this, we are confident that the reads assigned to *A. lumbricoidea* in this study stems almost solely from human *Ascaris* infections due to co-detection with the highly host specific *T. trichiura*. However, *T. suis* was identified in two samples (54 reads in sample 937 from Odense, Denmark and 6 reads in the Lithuania VLN-5 sample) indicating that pig faeces were present in these samples. Assuming co-detection of *T. suis* and *A. suum* in pigs, similar to the level observed for *T. trichiura* and *A. lumbricoidea* in humans, would imply that *A. suum* was also be present at those sites and in those samples. This may explain why these two samples exhibited the highest number of potential SNPs in their consensus mitochondrial sequences.

Reconstructing complete mitochondrial genomes of ancient parasites is novel and analysis of the associated genetic variation, opens for a more detailed level of information regarding

the transmission biology of parasites (Fig 4). As the sequencing data stems from many different eggs, originating from many different worms and potentially many different hosts, the reconstructed mitochondrial genomes should be considered consensus sequences of the population. Doing so, we identified a higher number of haplotypes in *T. trichiura* (average of 6.0 across samples) compared to *A. lumbricoides* (4.0). We interpret this as a result of worm biology, particularly in relation to the reproductive potential of an individual worm. The effective worm population of an infection, hence the worm burden, is typically 1–15 worms for *Ascaris* infections [33] and 10–60 worms for *Trichuris* infections [34]. The larger worm burden provides a larger gene pool in *T. trichiura* infections and may explain the higher number of haplotypes.

Despite that the mitochondrial genome sequence of *A. lumbricoides* cannot be used to distinguish *Ascaris* from humans and pigs, we show mitochondrial sequence affiliation to cluster B for all samples, particularly close resemblance to haplotype 03 and 07. Although, H03 and H07 are identified among isolates from humans and pigs in Africa, South America, and Asia they are more frequently found among samples from China [10,35]. In contrast, haplotype H01 belonging to cluster A is the most common type among samples from humans in Africa [10]. This suggests that *A. lumbricoides* from Northern Europe and Jordan (650–1700AD) are more closely related to worms in Asia than to worms in Africa, which is in contrast to what we see for *T. trichiura* as outlined below. However, comparison of frequencies of haplotypes between geographical regions has to be interpreted with some caution due to the low number of genomes generated in the current study and as the phylogeny of *Ascaris* based on mitochondrial DNA is complex (Fig 4).

In sheer contrast, the *Trichuris* sp. mitochondrial sequences can provide identification to species level and has recently been used to address phylogeographic distribution of *T. trichiura* and *T. suis* [9,36]. The identified phylogeny suggests that Northern European *T. trichiura* (1000–1700 AD) was very closely related to worms found in present day Uganda while it has significantly diverged from present day China. This surprising finding indicates very high degrees of genetic mixing between *T. trichiura* in Uganda and Northern Europe (1000–1700AD) which does not fully align with the hypothesis that *T. trichiura* was dispersed globally along with human migrations [8,9]. Further, the clear segregation between isolates from Africa+Europe versus China+Americas indicates that potentially two subspecies of *T. trichiura* should be considered.

Identification of DNA in modern faecal samples has previously been used as indicators of host diet [37,38]. Vertebrate DNA in ancient latrines thus reflects either the diet of the people who used it for defecation, or concurrent deposition of animal remains or excrements. The vertebrate DNA findings here could not be authenticated through ancient DNA damage profiles due to a very few number of reads. However, authentication was obtained by inclusion of negative controls, which in no instances resulted in assignment of reads to the positively identified taxa. Furthermore, the presented vertebrate findings are sound based on historical contexts, giving validity to the use of shotgun metagenomic approaches.

In particular, we show that vertebrate DNA from domesticated animals is more abundant compared to that from wildlife animals. Wildlife findings are interesting as they clearly indicate hunting and fishing practices, as shown in samples from Denmark (1018–1400 AD), where *Baleaenoptera physalus* (finwhale), *Capreolus capreolus* (roe deer) and *Lepus europaeus* (European hare) were identified. Large whales, including fin whales, are still recurrently stranding on beaches in Denmark, which would have provided easy access to large amounts of meat and fat, while catching live fin whales would have been a major accomplishment in the Viking-age. A comparable exploitation of dead whale by ancient cultures in the Arctic was recently suggested by us [16]. Another example of consumption of marine mammals, although below threshold for inclusion in Fig 2, is identification of harbour porpoise DNA in a latrine

sample from Odense, Denmark (1375–1400 AD), a species that has been exploited locally and is still native to Danish waters. Five fish species were identified in samples from Kampen in the Netherlands (1350–1850 AD) and Copenhagen, Denmark (1600s AD). Only saltwater fish species were identified in Copenhagen (*Clupea harengus* (Atlantic herring) and *Gadus morhua* (Atlantic cod)), while both fresh- (*Abramis brama* (common bream) and *Anguilla anguilla* (European eel)) and saltwater species (*G. morhua* and *Hippoglossus hippoglossus* (Atlantic halibut)) were identified in Kampen. This corresponds with their respective locations, Copenhagen at the sea and Kampen on the banks of the river IJssel, just upstream of its estuary.

Earlier studies have used ancient environmental DNA analysis to study past ecosystems and human activities [16,39–42]. The plant material found in latrines and deposits in the present study should mainly be considered a result of human deposition, through faecal excrements, general waste management (e.g. kitchen refuse and as part of animal faeces) or as a means to initiate biological degradation processes in the latrine (e.g. leaves, ferns), but only minimally as a result of natural deposition. The most abundant edible plants found in the samples from Northern Europe included cabbages (*Brassica*), buckwheat (*Fagopyrum*), wheat (*Hordeum*), sweet peas (*Lathyrus*), legumes (*Lotus*), peas (*Pisum*) and more seasonal foods herein mainly fruits: strawberries (*Fragaria*), plums/cherries (*Prunus*), pears (*Pyrus*), rhubarb (*Rheum*), and berries possibly cranberry/blueberry (*Vaccinium*). Notably, carrots and potatoes which are consumed in large quantities today, were not identified, in agreement with their gradual introduction in Northern Europe from the 16–1700s [43,44]. The present study did not allow for a detailed analysis of geographic preferences for the above mentioned plants, but grapes (*Vitis*), as often used in wine production, were identified in high concentration in the Netherlands and less so in Copenhagen, Denmark. In contrast, hop (*Humulus*), as often used in beer production, was identified in Odense and Copenhagen, Denmark and in Lithuania. Finally, tobacco (*Nicotiana*) was identified at the docks in Copenhagen (Gammel Strand), which was also the site of highest sample diversity, plant as well as helminth and vertebrate.

In conclusion, sequencing of ancient samples often involves a DNA target enrichment step in order to obtain sufficient concentrations of endogenous DNA for sequencing [45]. As a novel approach we have combined pre-concentration of parasite eggs with shotgun sequencing to conduct metagenomic studies and to reconstruct mitochondrial genomes. In addition, the approach employed has an obvious potential for reconstructing nuclear genomes of ancient parasites.

Supporting information

S1 Appendix. Description of sampling locations.

(DOCX)

S1 Table. Sequencing libraries information.

(XLSX)

S2 Table. Sample level information.

(XLSX)

S3 Table. Sequencing reads assigned to helminths.

(XLSX)

S4 Table. Sequencing reads assigned to vertebrates.

(XLSX)

S5 Table. Mitochondrial genome coverage and damage.

(XLSX)

S6 Table. Sequencing read assignment using plastid database.
(XLSX)

Acknowledgments

We thank: the University of Copenhagen for funding to M.J.S. and all laboratory expenses through the KU2016 initiative, ‘The Genomic History of Denmark’; Professor Kurt H. Kjær, for facilitating cross-institutional collaborations which were critical for completing the project; the Jordanian Department of Antiquities, Amman and Jarash offices, for their invaluable support and collaboration and the city archaeologist of Zwolle for making the parasitological research possible.

Author Contributions

Conceptualization: Martin Jensen Søe, Peter Nejsum, Eske Willerslev, Christian Moliin Outzen Kapel.

Data curation: Martin Jensen Søe, Peter Nejsum, Frederik Valeur Seersholm, Brian Lund Fredensborg.

Formal analysis: Martin Jensen Søe, Peter Nejsum, Frederik Valeur Seersholm, Brian Lund Fredensborg.

Funding acquisition: Eske Willerslev, Christian Moliin Outzen Kapel.

Investigation: Martin Jensen Søe, Brian Lund Fredensborg, Ruben Habraken, Kirstine Haase, Mette Marie Hald, Rikke Simonsen, Flemming Højlund, Louise Blanke, Inga Merkyte.

Methodology: Martin Jensen Søe, Peter Nejsum, Ruben Habraken, Mette Marie Hald.

Project administration: Martin Jensen Søe, Eske Willerslev, Christian Moliin Outzen Kapel.

Resources: Martin Jensen Søe, Ruben Habraken, Kirstine Haase, Mette Marie Hald, Rikke Simonsen, Flemming Højlund, Louise Blanke, Inga Merkyte, Christian Moliin Outzen Kapel.

Software: Martin Jensen Søe, Frederik Valeur Seersholm.

Supervision: Peter Nejsum, Brian Lund Fredensborg, Eske Willerslev, Christian Moliin Outzen Kapel.

Validation: Martin Jensen Søe.

Visualization: Martin Jensen Søe.

Writing – original draft: Martin Jensen Søe, Frederik Valeur Seersholm, Christian Moliin Outzen Kapel.

Writing – review & editing: Martin Jensen Søe, Peter Nejsum, Brian Lund Fredensborg, Ruben Habraken, Kirstine Haase, Mette Marie Hald, Rikke Simonsen, Flemming Højlund, Louise Blanke, Inga Merkyte, Eske Willerslev, Christian Moliin Outzen Kapel.

References

1. Goncalves MLC, Araujo A, Ferreira LF. Human intestinal parasites in the past: New findings and a review. *Mem Inst Oswaldo Cruz*. 2003; 98: 103–118. Available: <Go to ISI>://WOS:000180686100016
2. Yoshikawa H, Yamada M, Matsumoto Y, Yoshida Y. Variation in egg size of *Trichuris trichiura*. *Parasitol Res*. 1989; 75: 649–654. PMID: [2771930](https://pubmed.ncbi.nlm.nih.gov/2771930/)

3. Jimenez JA, Rodriguez S, Moyano LM, Castillo Y, García HH. Differentiating *Taenia* eggs found in human stools: Does Ziehl-Neelsen staining help? *Trop Med Int Heal*. 2010; 15: 1077–1081. <https://doi.org/10.1111/j.1365-3156.2010.02579.x> PMID: 20579318
4. Betson M, S e MJ, Nejsum P. Human Trichuriasis: Whipworm Genetics, Phylogeny, Transmission and Future Research Directions. *Curr Trop Med Reports*. US; 2015; 2: 209–217. <https://doi.org/10.1007/s40475-015-0062-y>
5. Liu GH, Gasser RB, Su A, Nejsum P, Peng L, Lin RQ, et al. Clear genetic distinctiveness between human- and pig-derived trichuris based on analyses of mitochondrial datasets. *PLoS Negl Trop Dis*. 2012; 6: e1539. <https://doi.org/10.1371/journal.pntd.0001539> PMID: 22363831
6. Ravasi DF, O'Riain MJ, Davids F, Illing N. Phylogenetic evidence that two distinct *Trichuris* genotypes infect both humans and non-human primates. *PLoS One*. 2012; 7: e44187. <https://doi.org/10.1371/journal.pone.0044187> PMID: 22952922
7. Hansen TVA, Nejsum P, Olsen A, Thamsborg SM. Genetic variation in codons 167, 198 and 200 of the beta-tubulin gene in whipworms (*Trichuris* spp.) from a range of domestic animals and wildlife. *Vet Parasitol*. 2013; 193: 141–149. <https://doi.org/10.1016/j.vetpar.2012.12.003> PMID: 23352105
8. Araujo A, Reinhard KJ, Ferreira LF, Gardner SL. Parasites as probes for prehistoric human migrations? *Trends Parasitol*. 2008; 24: 112–115. <https://doi.org/10.1016/j.pt.2007.11.007> PMID: 18262843
9. Hawash MBF, Betson M, Al-Jubury A, Ketzis J, LeeWillingham A, Bertelsen MF, et al. Whipworms in humans and pigs: origins and demography. *Parasit Vectors*. Parasites & Vectors; 2016; 9: 37. <https://doi.org/10.1186/s13071-016-1325-8> PMID: 26800683
10. Betson M, Nejsum P, Bendall R, Deb RM, Stothard JR. Molecular epidemiology of ascariasis: A global perspective on the transmission dynamics of *Ascaris* in people and pigs. *J Infect Dis*. 2014; 210: 932–941. <https://doi.org/10.1093/infdis/jiu193> PMID: 24688073
11. Yeh H-Y, Prag K, Clamer C, Humbert J-B, Mitchell PD. Human intestinal parasites from a Mamluk Period cesspool in the Christian quarter of Jerusalem: Potential indicators of long distance travel in the 15th century AD. *Int J Paleopathol*. Elsevier Inc.; 2015; 9: 69–75. <https://doi.org/10.1016/j.ijpp.2015.02.003> PMID: 29539442
12. Yeh HY, Pluskowski A, Kalejs U, Mitchell PD. Intestinal parasites in a mid-14th century latrine from Riga, Latvia: Fish tapeworm and the consumption of uncooked fish in the medieval eastern Baltic region. *J Archaeol Sci*. 2014; 49: 83–89. <https://doi.org/10.1016/j.jas.2014.05.001>
13. R acz SE, De Ar ujo EP, Jensen E, Mostek C, Morrow JJ, Van Hove ML, et al. Parasitology in an Archaeological Context: Analysis of Medieval Burials in Nivelles, Belgium. *J Archaeol Sci*. 2014; 53: 304–315. <https://doi.org/10.1016/j.jas.2014.10.023>
14. Mitchell PD. Human Parasites in Medieval Europe: Lifestyle, Sanitation and Medical Treatment. *Adv Parasitol*. Elsevier Ltd; 2015; 90: 1–32.
15. Piers D. Mitchell. Human parasites in the Roman World: health consequences of conquering an empire. *Parasitology*. 2016; Jan 8: 1–11. <https://doi.org/10.1017/S0031182015001651> PMID: 26741568
16. Seersholm FV, Pedersen MW, S e MJ, Taher H, Mak SST, Ruter A, et al. DNA evidence of bowhead whale exploitation by Greenlandic Paleo-Inuit 4000 years ago. *Nat Commun*. 2016; 13389. <https://doi.org/10.1038/ncomms13389> PMID: 27824339
17. Loreille O, Roumat E, Verneau O, Bouchet F, H anni C. Ancient DNA from *Ascaris*: Extraction amplification and sequences from eggs collected in coprolites. *Int J Parasitol*. 2001; 31: 1101–1106. [https://doi.org/10.1016/S0020-7519\(01\)00214-4](https://doi.org/10.1016/S0020-7519(01)00214-4) PMID: 11429174
18. Oh CS, Seo M, Chai JY, Lee SJ, Kim MJ, Park JB, et al. Amplification and sequencing of *Trichuris trichiura* ancient DNA extracted from archaeological sediments. *J Archaeol Sci*. Elsevier Ltd; 2010; 37: 1269–1273. <https://doi.org/10.1016/j.jas.2009.12.029>
19. S e MJ, Nejsum P, Fredensborg BL, Kapel CMO. DNA typing of ancient parasite eggs from environmental samples identifies human and animal worm infections in Viking-age settlement. *J Parasitol*. US; 2015; 101: 57–63. <https://doi.org/10.1645/14-650.1> PMID: 25357228
20. C ot e NML, Daligault J, Pruvost M, Bennett EA, Gorg e O, Guimaraes S, et al. A New High-Throughput Approach to Genotype Ancient Human Gastrointestinal Parasites. *PLoS One*. 2016; 11: e0146230. <https://doi.org/10.1371/journal.pone.0146230> PMID: 26752051
21. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc*. 2010; 6: pdb.prot5448. <https://doi.org/10.1101/pdb.prot5448> PMID: 20516186
22. Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspina A-S, Manica A, Moltke I, et al. Genomic structure in Europeans dating back at least 36,200 years. *Science*. 2014; 346: 1113–1118. <https://doi.org/10.1126/science.aaa0114> PMID: 25378462

23. Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, et al. Population genomics of Bronze Age Eurasia. *Nature*. 2015; 522: 167–172. <https://doi.org/10.1038/nature14507> PMID: 26062507
24. Lindgreen S. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res Notes*. 2012; 5: 337. <https://doi.org/10.1186/1756-0500-5-337> PMID: 22748135
25. Simpson JT, Durbin R. Efficient construction of an assembly string graph using the FM-index. *Bioinformatics*. 2010; 26: 367–373. <https://doi.org/10.1093/bioinformatics/btq217> PMID: 20529929
26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9: 357–359. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
27. Leonard JA, Shanks O, Hofreiter M, Kreuz E, Hodges L, Ream W, et al. Animal DNA in PCR reagents plagues ancient DNA research. *J Archaeol Sci*. 2007; 34: 1361–1366. <https://doi.org/10.1016/j.jas.2006.10.023>
28. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
29. Schubert M, Ermini L, Der Sarkissian C, Jónsson H, Ginolhac A, Schaefer R, et al. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc*. 2014; 9: 1056–82. <https://doi.org/10.1038/nprot.2014.063> PMID: 24722405
30. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol*. 2013; 30: 772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
31. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. 2013; 29: 1682–4. <https://doi.org/10.1093/bioinformatics/btt193> PMID: 23613487
32. Søre MJ, Kapel CMO, Nejsum P. Ascaris from Humans and Pigs Appear to Be Reproductively Isolated Species. *PLoS Negl Trop Dis*. 2016; 10: e0004855. <https://doi.org/10.1371/journal.pntd.0004855> PMID: 27583548
33. Elkins DB, Haswell-Elkins M, Anderson RM. The epidemiology and control of intestinal helminths in the Pulicat Lake region of Southern India. I. Study design and pre- and post-treatment observations on *Ascaris lumbricoides* infection. *Trans R Soc Trop Med Hyg*. 1986; 80: 774–792. [https://doi.org/10.1016/0035-9203\(86\)90384-6](https://doi.org/10.1016/0035-9203(86)90384-6) PMID: 3603617
34. Bundy DAP, Thompson DE, Golden MHN, Cooper ES, Anderson RM, Harland PSE. Population distribution of *Trichuris trichiura* in a community of Jamaican children. *Trans R Soc Trop Med Hyg*. 1985; 79: 232–237. [https://doi.org/10.1016/0035-9203\(85\)90343-8](https://doi.org/10.1016/0035-9203(85)90343-8) PMID: 4002293
35. Peng W, Yuan K, Hu M, Zhou X, Gasser RB. Mutation scanning-coupled analysis of haplotypic variability in mitochondrial DNA regions reveals low gene flow between human and porcine *Ascaris* in endemic regions of China. *Electrophoresis*. 2005; 26: 4317–4326. <https://doi.org/10.1002/elps.200500276> PMID: 16287175
36. Hawash MBF, Andersen LO, Gasser RB, Stensvold CR, Nejsum P. Mitochondrial Genome Analyses Suggest Multiple *Trichuris* Species in Humans, Baboons, and Pigs from Different Geographical Regions. *PLoS Negl Trop Dis*. 2015; 9: e0004059. <https://doi.org/10.1371/journal.pntd.0004059> PMID: 26367282
37. Murray DC, Bunce M, Cannell BL, Oliver R, Houston J, White NE, et al. DNA-based faecal dietary analysis: A comparison of qPCR and high throughput sequencing approaches. *PLoS One*. 2011; 6. <https://doi.org/10.1371/journal.pone.0025776> PMID: 21998697
38. Kartzinel TR, Chen P a., Coverdale TC, Erickson DL, Kress WJ, Kuzmina ML, et al. DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proc Natl Acad Sci*. 2015; 112: 8019–8024. <https://doi.org/10.1073/pnas.1503283112> PMID: 26034267
39. Giguët-Covex C, Pansu J, Arnaud F, Rey P-J, Griggo C, Gielly L, et al. Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nat Commun*. 2014; 5: 3211. <https://doi.org/10.1038/ncomms4211> PMID: 24487920
40. Hebsgaard MB, Gilbert MTP, Arneborg J, Heyn P, Allentoft ME, Bunce M, et al. ‘The Farm Beneath the Sand’—an archaeological case study on ancient ‘dirt’ DNA. *Antiquity*. 2009; 83: 430–445.
41. Pedersen MW, Ruter A, Schweger C, Friebe H, Staff RA, Kjeldsen KK, et al. Postglacial viability and colonization in North America’s ice-free corridor. *Nature*. Nature Publishing Group; 2016; 537: 1–15. <https://doi.org/10.1038/nature19085> PMID: 27509852
42. Pedersen MW, Overballe-Petersen S, Ermini L, Sarkissian C Der, Haile J, Hellstrom M, et al. Ancient and modern environmental DNA. *Philos Trans R Soc Lond B Biol Sci*. 2015; 370: 20130383. Available: <http://dx.doi.org/10.1098/rstb.2013.0383> <https://doi.org/10.1098/rstb.2013.0383> PMID: 25487334

43. Ames M, Spooner DM. DNA from herbarium specimens settles a controversy about origins of the European potato. *Am J Bot.* 2008; 95: 252–257. <https://doi.org/10.3732/ajb.95.2.252> PMID: 21632349
44. Simon PW. Domestication, Historical Development and Modern Breeding of Carrot. *Plant breeding reviews.* 2000. pp. 157–190.
45. Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, Sikora M, et al. Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *Am J Hum Genet.* 2013; 93: 1–13. <https://doi.org/10.1016/j.ajhg.2013.10.002> PMID: 24568772