Metagenomic peek into a corn mummy

Norbert Solymosi[1,2*], Bernadett Pap[3], Sára Ágnes Nagy[2], Adrienn Gréta Tóth[1,2], Flóra Judit Kevély[4], Gergely Maróti[3], István Csabai[2], Katalin Kóthay[4], and Donát Magyar[5]

[1]Centre for Bioinformatics, University of Veterinary Medicine, Budapest, 1078, Hungary
[2]Department of Phyisics of Complex Systems, Eötvös Loránd University, Budapest, 1117, Hungary
[3]Institute of Plant Biology, HUN-REN Biological Research Center, Szeged, 6726, Hungary
[4]Egyptian Collection of the Museum of Fine Arts, Budapest, 1146, Hungary
[5]National Center for Public Health and Pharmacy, Budapest, 1097, Hungary
[*]solymosi.norbert@gmail.com

ABSTRACT

Numerous studies have shown that metagenomics has opened up a new dimension in reading the contents of archaeological remains as time capsules. Corn mummies are ritual objects made in ancient Egypt by forming human-shaped figures made of cereal grains grown in a mixture of water and earth. We performed metagenomic analyses on samples from a corn mummy dated to the second half of the 3rd century BC. Alongside a number of clearly modern contaminants, we identified organisms that cannot be excluded as being of historical origin. Besides considerable amounts of bacterial sequences belonging to the genera *Bacillus*, *Coprococcus*, *Metabacillus*, *Niallia* and *Roseburia*, we also found traces of plants, animals, and humans. Sequences assigned to the genus *Triticum* showed the highest similarity to ancient *T. turgidum* ssp. *dicoccum* specimens from Egypt and the southern Levant. Sequences classified in the genus *Populus* showed the highest identity with the genome of *P. nigra*. The fragments that were identified as being of Lepidopteran origin showed the greatest similarity to the genomes of the Sphingidae. Haplotype analysis of *H. sapiens* sequences suggests haplogroups L3 and E, based on the mitogenome and Y chromosome, respectively, which are the two common lineages in sub-Saharan Africa today.

Introduction

Metagenomics, based on next-generation sequencing, brings new and novel results to studying complex communities and ecosystems. In addition to recent, contemporary samples, this new approach also allows the detection of traces of organisms that once lived and were trapped in a sample as a time capsule based on the DNA content of archaeological remains.[1–4] Metagenomic tools have been used to identify the genome of the oldest known *Mycobacterium leprae* from the bones of an Egyptian human mummy.[5] Khairat et al.[6] also detected microbial and plant genome fragments in samples from human mummies. A special type of mummy is the corn mummy, which was made in Egypt between the 3rd century BC and the 1st century AD as part of the religious rites of the Khoiak festival. In the ceremony, the grain was grown in a mixture of 1 hin (∼0.47 L[7]) of barley and/or wheat,[8,9] 1 hin of water, and 4 hin of sand for 9 days (between 12 and 21 of the month of Khoiak, around 22-31 December), and it was watered by river water.[10] Then, the formed mummy was placed on dried myrrh, embalmed, painted, and wrapped in papyrus and clothed with textiles, and finally placed in a sycamore coffin.[11] The results of metagenomic research on ancient Mesopotamian clay bricks by Arbøll et al.[12] inspired us to conduct a similar study on a corn mummy. In order to gain insight into the organism composition of corn mummies, metagenomic analyses were carried out on samples from one of the two corn mummies in the Egyptian Collection of the Museum of Fine Arts, Budapest, Hungary.

Material and methods

Sampling

The Egyptian Collection of the Budapest Museum of Fine Arts owns two corn mummies. One of them (inv. no. 60.22-E) is completely intact and has a sarcophagus, while the other (inv. no. 68.2-E) is without a sarcophagus and is broken crosswise. Object no. 68.2-E was acquired by the Museum in 1968. It is estimated by experts to be from Central Egypt and dated to the second half of the 3rd century BC on stylistic grounds.[11]

Due to the high value and rarity of the specimen, we took samples from the fractured mummy (68.2-E) with great care, disturbing only the fracture surface. On the very first occasion (Sample 0), specimens were taken from the surface shown in Figure 1 by "thin needle biopsy". This was done by inserting a sterile long blood collection needle deep into part A of the mummy and withdrawing the needle with a syringe to create a vacuum. The remains in the lumen of the needle were then collected in a sterile tube. It was not possible to isolate enough DNA from this sample for sequencing. In the next sample,

we used the fragments that had fallen off the mummy to see if sufficient DNA could be obtained from this material. From this Sample 1 (Figure 1), we could extract enough DNA to allow sequencing. Since our question had been answered, i.e., a sufficient amount of DNA could be isolated from the mummy, we decided to use a different solution instead of a needle biopsy for the next sample. Using a sterile syringe cone, we scraped 1-2 mm of the surface of the fracture from part B of the split mummy. Continuing this procedure, a sample was taken in a sampling tube from the material deeper to the surface (Sample 2).

DNA extraction, library preparation and sequencing

Kistler's CTAB Protocol[13] was used on DNA samples isolated from 200 mg material of Sample 0. For Sample 1, cell pellets were re-suspended in 567 μL of 1xTE (10 mM Tris, 1 mM EDTA pH 8.0) buffer (pH 8.0) containing 2 mg/mL of lysozyme and incubated for 30 minutes at 37°C. After adding 30 μL 10% sodium dodecyl sulfate (SDS) and 3 μL 20 mg/mL Proteinase K incubated for 1 hour at 37°C. Samples were then incubated for 20 minutes at 65°C with 150 μL of 5 M NaCl and 80 μL of CTAB/NaCl solution (700 mM NaCl/10% CTAB). Following incubation, extracts were purified using phenol-chloroform extraction, and DNA was recovered by isopropanol precipitation. Pelleted DNA was washed with 70% ethanol, allowed to air dry, and re-suspended in 32 μL of 1xTE buffer. In the case of Sample 2, following the Basic Protocol 1 published by Latorre et al.,[14] 20 and 200 mg samples were used to isolate aDNA.

Paired-end libraries were prepared using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (Cat.Num.: E7645L). Paired-end fragment reads were generated on an Illumina NextSeq sequencer using TG NextSeq® 500/550 High Output Kit v2 (300 cycles). Primary data analysis (base-calling) was done with Bbcl2fastq software (v2.17.1.14, Illumina).

Bioinformatics

The raw FASTQ files were trimmed and filtered by TrimGalore (v.0.6.6, https://github.com/FelixKrueger/TrimGalore, accessed on 12/6/2024), with a quality threshold of 20, minimal length 50 bp. The remained reads were taxonomically classified by Kraken2 (v2.1.3, `--confidence 0.5`)[15] using the NCBI NT[16] database (created: 3/1/2024). The de novo assembly was performed by MEGAHIT (v1.2.9)[17] with default settings. The assembled contigs were taxon classified by local BLAST[18] on NCBI NT database (updated: 10/5/2024). The blast results of the contigs were further evaluated by targeting the selected genus. We collected all the contigs for which the blast gave a hit to a genus. For each of these contigs, we then assessed which match gave the highest bit-score. Since the same maximum bit-score can be obtained for different taxa, the results were further filtered by selecting the hit with the minimum difference between the contig and the length of the match. Even though this can result in different taxa giving the same score, these were treated as undecidable. In metagenome binning, clusters were generated from the assembled contigs using MaxBin2 (v2.2.7),[19] MetaBAT2 (v2.17-21-g2740d8d),[20] MetaBinner (v1.4.4),[21] MetaDecoder (v1.0.18)[22] and SemiBin2 (v2.0.2).[23] From these, the best bins found by DAS Tool (v1.1.6)[24] were evaluated using NCBI prokaryotic genome annotation pipeline (v2024-04-27.build7426, PGAP).[25]

For haplogroup classification of probable *Homo sapiens* origin sequences, cleaned reads were aligned to mitochondrial (Reconstructed Sapiens Reference Sequence, RSRS) and Y chromosome reference sequences using Bowtie2 (v2.5.3, `--very-sensitive-local`)[26]. Duplums were removed from the RSRS-based BAM file using Picard (v2.26.6, https://broadinstitute.github.io/picard/), and samtools (v1.19, `consensus` with default settings)[27] was used to generate the mitochondrial consensus sequence, which was haplogroup-predicted by Haplogrep3 (v3.2.1, metric: Kulczyński, PhyloTree Build: 17)[28]. Contamination analysis of the mitochondrial aligned reads was performed using Haplocheck (v1.3.2)[29], which requires alignment to the Revised Cambridge Reference Sequence (rCRS), so the cleaned reads were also aligned to the NC_012920.1 sequence. The contamination of reads aligned to the mtDNA was checked by schmutzi (v1.5.7)[30] and ContamMix (v1.0-11)[31], too. By ContamMix 311 worldwide mitogenomes[32] were compared to our mitochondrial consensus sequence based on multiple sequence alignment by MAFFT (v7.490, `--auto`).[33]

Further contamination analysis was performed using the contamination (`X:5000000-154900000, -doCounts 1 -iCounts 1 -minMapQ 30 -minQ 20`) scripts of the ANGSD (v0.940)[34] package based on X chromosome (hg19, NC_000023.10) alignments. Based on the alignment of the cleaned reads to the human genome (GRCh38), we tested the XY karyotype.[35] Human Y chromosome-based haplogroup prediction was performed by pathPhynder[36] and yHaplo (v2.1.12.dev9+geb252e7, `--anc_stop_thresh 10`)[37], using the Bowtie2 aligned cleaned reads on the GRCH38.p12 and hs37d5 genomes, respectively. The VCF input of yHaplo was called by bcftools (v1.14, `--min-MQ 30 --min-BQ 30 --ploidy 1`)[27]. To illustrate the geographic prevalence of L3 and E haplotypes within Africa, we generated maps based on the data published by Cabrera et al. in their Supplementary Table S2.[38]

Post-mortem DNA damage was analysed by estimating the frequency of C>T and G>A misincorporation using MapDamage (v2.3.0a0)[39]. For this analysis, sequences were cleaned using AdapterRemoval (v2.3.3, `--trimns --trimqualities --minlength 30 --minquality 20 --minadapteroverlap 1 --collapse`)[40]. The cleaned reads were aligned to reference genomes (with at least 1000 reads from the Kraken2 classification, Figure S2-S4) using BWA (v0.7.17-r1198, `-n 0.01 -l 1024 -k 2 -o 2`)[41].

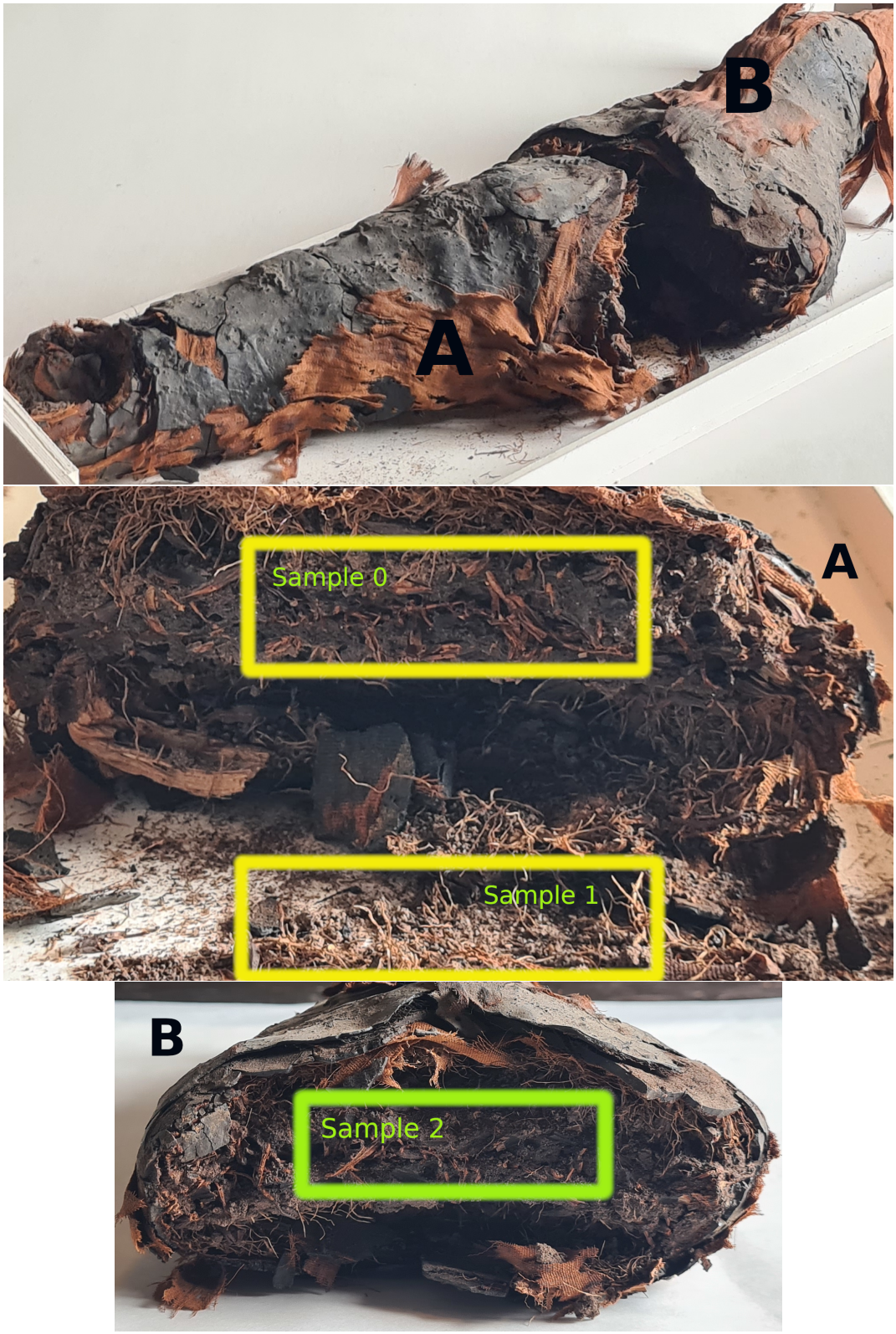

Figure 1. Corn mummy 68.2-E sampled. The top image shows the whole mummy, the middle image shows part A with sampling areas 0 and 1, and the bottom image shows part B with sampling area 2.

The phylogenetic analysis of modern and ancient samples of the genus *Triticum* and *Hordeum* was based on the longest contig of the genus *Triticum*, which we generated by de novo assembly. The data for modern and ancient samples from the Illumina sequencing aligned to the longest contig of the genus *Triticum* can be found in Table S2. From the generated BAM files, we constructed consensus sequences using samtools, keeping those with the N ratio of less than 5%. The gene-tree was constructed[42] based on multiple sequence alignment by MAFFT. The best substitution model was selected by functions of phangorn (v2.11.1) package[43] based on the Bayesian information criterion (BIC). The generated neighbor-joining tree was optimized by the maximum likelihood method. Bootstrap values were produced by 100 iterations. All data processing and plotting were done in the R-environment.[44]

Results

Sample 1

Of the 5,472,755 raw read pairs, we have 5,426,107 reads left after trimming and cleaning. Of the 4,317,771 reads classified, 99.856% were bacteria, 0.139% were eukaryotes, and the rest were viruses and archaea. The bacterial genera with more than 0.5% were as follows: *Cupriavidus* (89.1%), *Stenotrophomonas* (5.1%), *Caulobacter* (4.9%). The detected genomes in Sample 2 were present in Sample 1: *Acinetobacter* (with read count=540), *Bacillus* (130), *Bacteroides* (2), *Coprococcus* (0), *Cutibacterium* (808), *Faecalibacterium* (0), *Kocuria* (16), *Metabacillus* (15), *Microbacterium* (19), *Micrococcus* (449), *Niallia* (8), *Paenibacillus* (18), *Roseburia* (0), *Staphylococcus* (166). In Sample 1, 5559 reads were assigned to the *H. sapiens* genome and 5 to the genus *Triticum*. Although there were reads of human origin, they matched the mitochondrial and Y chromosome reference sequences only very fragmentarily, so that haplogroups could not be predicted from them. By binning the contigs generated by de novo assembly of the reads from Sample 1, three clusters were identified (Table S1). Figure S2 summarises the results of the DNA damage analysis of Sample 1, none of the genomes detected can be considered as historical, but rather as recent contamination.

Sample 2

From 1,502,824 raw read pairs, 1,338,027 reads remained after trimming and cleaning. Out of the 276,104 reads classified (conf: 50), 36.374% were bacteria, 62.751% were eukaryotes, and the rest were viruses and archaea. Bacterial genera with a relative abundance of at least 0.5%: *Cutibacterium* (5.7%), *Metabacillus* (2.7%), *Kocuria* (2.5%), *Niallia* (2%), *Coprococcus* (1.9%), *Roseburia* (1.9%), *Bacillus* (1.8%), *Micrococcus* (1.5%), *Acinetobacter* (1.2%), *Faecalibacterium* (1%), *Bacteroides* (0.8%), *Microbacterium* (0.6%), *Staphylococcus* (0.6%), *Paenibacillus* (0.5%). The bacterial species classified with a relative abundance of at least 0.1% and their classification results are shown in Table 1. Presence of genera identified from Sample 1 in Sample 2: *Caulobacter* (read count=7), *Cupriavidus* (12), *Stenotrophomonas* (53).

Eukaryota orders classified at 1% or higher: Primates (49.8%), Lepidoptera (7.86%), Poales (1.83%), Malpighiales (1.54%).

Within Primates, *Homo sapiens* was identifiable at the species level (47.5%). When these reads were aligned to the human genome (GRCh38), the average coverage of the 22 chromosomes (excluding Y and MT) was 1.16% (SD: 0.11), with a depth of 0.028 (SD: 0.003). Alignment to the RSRS sequence resulted in 688 reads with 85.53% coverage and 4.56× depth of coverage. The top 5 Haplogrep3 hits of the consensus sequence obtained after deduplication are summarised in table 2. Based on Haplocheck analysis, mtDNA did not contain any detectable contamination, X-chromosome based analysis estimated a contamination level of 0.007%. Neither schmutzi nor ContamMix gave valuable results, the former due to insufficient depth of mtDNA aligments, the latter due to incomplete coverage. From the ratio of reads matching the X and Y chromosomes, an XY karyotype was obtained ($R_y = 0.085$, 95% CI: 0.082-0.089). However, the karyotype could not be predicted with the default minimum mapping quality of 30 but only with a threshold of 10. A coverage of 6.50% and an average depth of 0.45 was obtained by aligning the reads to the Y chromosome (GRCH38.p12). The best path elements identified by pathPhynder are A1b, BT, C1a2, CT, E2, E2b1a1∼, yHaplo predicts haplogroup E-M96. The relative frequencies of haplogroups L3 and E within Africa are shown in Figure S1. *H. sapiens* received the maximum bitscore of 550 (total length: 237,647 bp, longest contig: 3,715 bp) contigs generated by de novo assembly.

Findings belonging to the order Lepidoptera were mainly from the genus *Bombyx* (7.84% of the classified reads), with *B. mori* (2.76%) and *B. mandarina* (0.097%) detected at the species level. In addition, two other species, *Eugnorisma glareosa* and *Campaea margaritaria*, had 3 and 2 reads, respectively. Among the contigs generated by de novo assembly, 24 showed a maximum BLAST match to *B. mori* (total length: 10,565 bp, longest contig: 1377 bp). For three contigs, *B. mandarina* had the maximum bitscore (total length: 1,092 bp, longest contig: 427 bp). When the reads were aligned to the *B. mori* reference genome (ASM3026992v2), the average coverage was 1.08% (SD: 3.27), depth 0.033 (SD: 0.096). In Sample 1, a single read was assigned to the genus *Bombyx*.

Within the order Poales, the genus *Triticum* dominates the hits (1.72%), with other genera appearing in less than 100 reads in the sample. At the species level, *T. aestivum* had the highest number of hits (0.41%). Based on the BLAST results of the contigs for *T. aestivum* 207 sequences (total length: 84,959 bp, longest contig: 1554 bp), 59 for *T. monococcum* (total length:

Table 1. Bacterial species with a relative abundance of at least 0.1% from Sample 2. The results of DNA damage analyses for species identified with at least 1000 reads are shown in Table S3.

Species	Abundance (%)	Read count
Cutibacterium acnes	4.79	13,226
Metabacillus sp. KUDC1714	2.19	6,057
Niallia circulans	1.38	3,819
Coprococcus sp. ART55/1	1.19	3,296
Roseburia intestinalis	1.12	3,095
Bacillus sp. Y1	0.41	1,140
Bacillus paralicheniformis	0.40	1,113
Acinetobacter junii	0.29	800
Faecalibacterium prausnitzii	0.26	725
Kocuria rhizophila	0.23	644
Moraxella osloensis	0.22	596
Staphylococcus epidermidis	0.18	491
Phocaeicola vulgatus	0.15	418
Odoribacter splanchnicus	0.15	413
Dorea longicatena	0.14	393
Methylosinus sp. C49	0.12	334
Micrococcus luteus	0.12	331
Bacteroides caccae	0.12	323
Alistipes shahii	0.12	319
Dialister succinatiphilus	0.11	315
Oceanobacillus iheyensis	0.10	286
Chryseobacterium gambrini	0.10	282
Parabacteroides merdae	0.10	282
Alistipes onderdonkii	0.10	280

Table 2. Top 5 haplogroup assigments of RSRS-based mtDNA consensus sequence analysis with Haplogrep3 (metric: Kulczyński, PhyloTree Build: 17).

Haplogroup	Quality	Detected polys	Undetected polys
L3	0.8870	146T, 152T, 182C!, 1018G, 2758G, 2885T, 3594C, 4104A, 4312C, 7146A, 7256C, 7521G, 8468C, 8655C, 10688G, 10810T, 10915T, 11914G, 13506C, 16129G, 16187C, 16189T, 16230A, 16278C	182T
L3c'd	0.8754	146T, 152T, 182C!, 1018G, 2758G, 2885T, 3594C, 4104A, 4312C, 7146A, 7256C, 7521G, 8468C, 8655C, 10688G, 10810T, 10915T, 11914G, 13105G!, 13506C, 16129G, 16187C, 16189T, 16230A, 16278C	152T!, 182T
L3b'f	0.8716	146T, 152T, 182C!, 1018G, 2758G, 2885T, 3594C, 4104A, 4312C, 7146A, 7256C, 7521G, 8468C, 8655C, 10688G, 10810T, 10915T, 11914G, 13506C, 16129G, 16187C, 16189T, 16230A, 16278C	182T, 15944d
L3'4	0.8663	146T, 152T, 182C!, 2758G, 2885T, 3594C, 4104A, 4312C, 7146A, 7256C, 7521G, 8468C, 8655C, 10688G, 10810T, 10915T, 11914G, 13506C, 16129G, 16187C, 16189T, 16230A, 16278C	182T
L3e'i'k'x	0.8656	146T, 152T, 182C!, 1018G, 2758G, 2885T, 3594C, 4104A, 4312C, 7146A, 7256C, 7521G, 8468C, 8655C, 10688G, 10810T, 10915T, 11914G, 13506C, 16129G, 16187C, 16189T, 16230A, 16278C	150T, 182T, 10819G

22,486 bp, longest contig: 907 bp), to *T. urartu* 21 (total length: 8,398 bp, longest contig: 697 bp), to *Aegilops tauschii* 17 (total length: 7,164 bp, longest segment: 694 bp), to *T. tugridum* 10 (total length: 3,978 bp, longest segment: 599 bp), to *Aegilops umbellulata* 7 (total length: 2,887 bp, longest segment: 508 bp) received the highest bitscore. The results of the phylogenetic

analysis (with the best substitution model TPM2u) based on the longest contig (GenBank accession PP971119: 1554 bp) classifiable in the genus *Triticum* are shown in Figure 2. Tables S3 and S4 were obtained by BLAST the PP971119 sequence onto the representative reference genomes of *T. aestivum* and *Hordeum vulgare*, respectively.

Within the order Malpighiales, the genus *Populus* dominates the hits (1.54%), with other genera accounting for less than 100 reads in the sample. At the species level, *P. nigra* had the highest number of hits (0.93%). Based on the BLAST result of the contigs, *P. nigra* received the highest bitscore for 9 sequences (total length: 3,444 bp, longest contig: 771 bp).

According to Figures S3-S4, which summarise the results of DNA damage analysis, the smiley plots obtained for *Bacillus paralicheniformis*, *B. mori*, *Coprococcus* sp. ART55/1, *H. sapiens*, *Metabacillus* sp. KUDC1714, *Niallia circulans*, *P. nigra*, *Roseburia intestinalis*, *T. aestivum* from Sample 2 show patterns suggesting that they are historical rather than contemporary sequences. The *Cutibacterium acnes* sequences seem more contemporary.

Discussion

The large number of bacterial reads identified in Sample 1 can be reasonably assumed to be from modern contamination, partly because of the absence of DNA damage and partly because they could be used to generate almost complete genomes. Since all three bacteria, whose genomes we have assembled are almost complete, can survive under extreme conditions, with minimal nutrients and in the presence of toxic metals,[45–47] it is reasonable to assume that they are bacteria living on the surface of the corn mummy. Swab samples collected from the surfaces of the corn mummy did not result in any colony after incubation on blood agar at room temperature for 2 weeks (data not shown).

There is also the possibility of contamination during sample preparation in the laboratory, but no work with the identified bacteria was carried out in the laboratory at the same time. The appearance of bacterial reads identified in Sample 1 may also be due to barcode cross-talk, as other samples were sequenced in the same run as the sample. However, due to this mixing, only a few hundred or thousand reads may be transferred from one sample to another, but not millions as in our case.

For the bacteria found in Sample 2 for which DNA damage could be tested and which showed a historical pattern of *Bacillus* spp.[48,49] *Coprococcus* spp.,[50] *Metabacillus* sp. KUDC1714,[51] *N. circulans*[52], their occurrence in the soil is supported by the literature. There is no such data for *R. intestinalis*, but as it is a well-known bacterium in the human and animal intestinal tract, it may be present in faecally contaminated soils.[53] *C. acnes* is also known to occur in human skin[54] and soil[55], but as we do not see a historical pattern of DNA damage for this bacterium, it is more likely to be contemporary human contamination. Bacteria with less than 1000 reads but at least 0.1% abundance may be related to soil, rhizosphere (*A. junii*,[56] *C. gambrini*,[57] *K. rhizophila*,[58] *Methylosinus* sp. C49,[59] *M. luteus*,[60] *M. osloensis*,[61] *O. iheyensis*,[62] *P. merdae*[63]) or human intestinal tract, faeces (*Alistipes* spp.,[64,65] *B. caccae*,[66] *D. succinatiphilus*,[67] *D. longicatena*,[68] *F. prausnitzii*,[69] *O. splanchnicus*,[70] *P. vulgatus*[71]). Similar to *C. acnes*, *S. epidermidis*[72] can be found on human skin and soil. The presence of soil-associated, non-historic bacteria in the sample is difficult to speculate. It is possible that the mummy was exposed to some form of sewage before it was deposited in the museum. Bacteria were introduced at that time and have been living and multiplying in the mummy ever since. However, we cannot ignore the possibility that human/animal fecal bacteria may be present in the sample because the soil has been fertilized for seed germination.

The presence of Poales sequences is not surprising as there is literature data on both *Triticum* and *Hordeum* content in corn mummies.[8,9] A previous archaeobotanical analysis of the same corn mummy we studied identified barley as corn content.[11] In our results, more sequences were identified for *Triticum* with minimal presence of *Hordeum*. These two results suggest that it may have been used as a mixed grain for rituals. Although we were unable to use the genes[73] commonly used for *Triticum* phylogenetic analysis, based on the PP971119 contig, our sample most closely resembles ancient Egyptian and southern Levantine samples. The ratio of wheat to human reads in Sample 1 is much lower than in Sample 2.

Populus was the other most abundant plant taxon, and the best match was with the *P. nigra* genome. As *P. nigra* is only native to the Egyptian north coast, and it has a higher probability the origin of the sequences found belonged to another species of the genera. In their study, Khairat et al.[6] also found *Populus* sequences in Egyptian human mummies. In that work, the authors interpreted the presence of *Populus* sequences as a component influencing the smell of mummified bodies. It is possible that it was also used for this purpose in the corn mummy. It is also possible that the pollen grains or seeds of the poplar tree, which are easily blown by the wind, were deposited in the mummy during the ceremony. Poplar pollen was detected in resins used for coffins in Egypt from the second century BC,[74] mostly with other airborne pollen types of late spring and summer. But if the corn mummy was created during the Khoiak Festival, in December, then it is too early for any poplar tree to flower or fruit.

The interpretation of the *Bombyx* results may be linked to the *Niallia* findings, as *N. circulans* is known to be present in the gut of *B. mori*, among other environments.[75] The DNA damage of the reads matched to two species suggests that they are likely historical. The larvae of *Bombyx* and its related species feed on the leaves of various plants and pupate either on the plant, on the soil surface, or by digging into the soil. There might be two ways in which a developmental form of the moth could have been introduced into Sample 2. One is that a larva (or pupae) was introduced into the corn mummy with the soil used to create

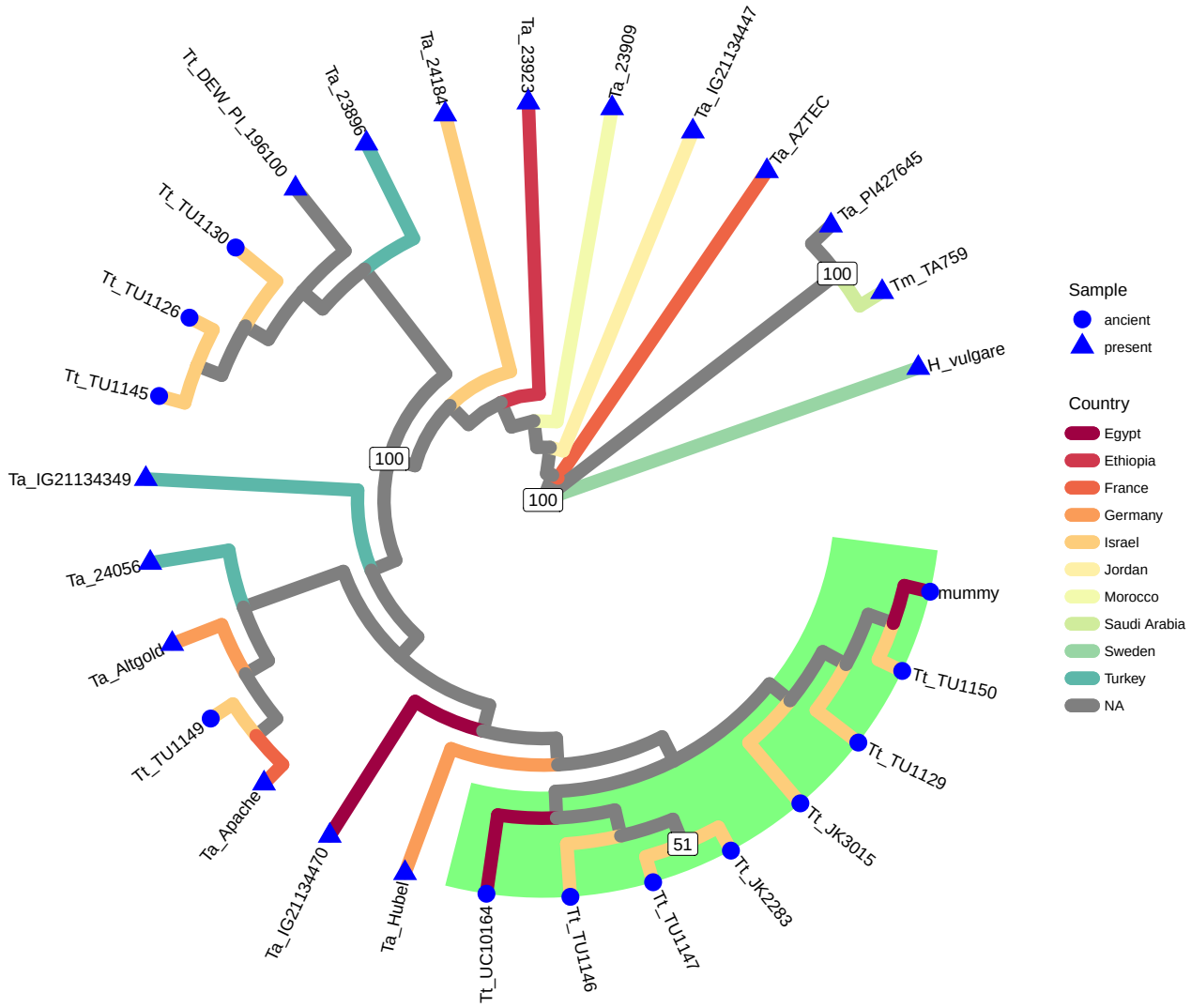

Figure 2. Cladogram of *Triticum* and *Hordeum* samples based on the PP971119 sequence. In the labels the H_vulgare is *H. vulgare* (outer group), Ta is *T. aestivum*, Tm is *T. monococcum*, Tt is *T. turgidum* ssp. *dicoccum*, mummy is the PP971119 contig from the corn mummy sample. The green band indicates the clade containing the corn mummy sequence. Numbers at branches indicate bootstrap support levels (100 replicates). The detailed information of the used samples is summarized in Table S2.

the corn mummy. The other one is that the larva was placed on the open surface of the corn munny prior to mummification. There is little probability of the larva entering the aged corn mummy through the fractured surface of the already broken corn mummy and digging into it. This latter case could have happened several decades ago, raising the question of whether the larva would have chosen the old corn mummy for pupation. A further issue with this opportunity is that no related species has been described in museum conditions. If the sequences are indeed from *B. mori*, we need to investigate whether this species may have come into contact with the mummy when it was made. As we may know, silk was not introduced into Egypt until the Ptolemaic period (332-30 BC),[76] but it was possible at the time of the mummy's creation. Although the export of *B. mori* moths, larvae or pupae from China was banned until the end of the Middle Ages, if silk had appeared in Egypt, the moth might have had a chance, and in this way *B. mori* might have the opportunity to be incorporated into the mummy. However, it should also be considered that the sample contains not *B. mori* genome fragments but those of a related moth species that occurred in the environment where the mummy was made. Several species of Bombycoidea/Sphingidae are known to be native to Egypt: *Acherontia atropos*, *Agrius convolvuli*, *Daphnis nerii*, *Hippotion celerio*, *Hyles euphorbiae*, *Hyles lineata*, *Hyles livornica*, *Hyles tithymali deserticola*, *Macroglossum stellatarum*, *Theretra alecto*, *Theretra oldenlandiae*. From these species, the NCBI NT database contains complete genomes only for the *H. euphorbiae* and the *H. lineata*. It is also possible that a related moth lived in the area when the mummy was made but later extinct, so we don't even know a trace of its genome. On this basis, we cannot rule out the possibility that the sequences we identified as *B. mori* originated from a related species.

The haplogrouping of the human sequences found in Sample 2 suggests that both the mtDNA and the Y chromosome might be of African, sub-Saharan origin. These haplogroups may have been present in both modern and ancient Egypt. Within the L3 mitochondrial lineage we found, the L3b, L3d, L3e, and L3f subclades are the most common in Central and West Africa.[77]. L3c is very rare in Soares et al., who report one sample from East Africa and one from the Middle East.[77] Rishishwar and King[78] in their global survey of MT haplotype groups report their distribution by continent based on 2534 samples. Of these, 272 were L3 haplotypes, of which 74.6% (n=203) were from African populations, 0.37% (n=1) from European populations, 0.74% (n=2) from Indian populations, and 24.3% (n=66) from American populations. However, this latter figure should be supplemented by the fact that 50 of the L3 US samples had known African ancestry (African Caribbean in Barbados n=26, African ancestry in the South West US n=24). None of the 504 East Asian samples included in the study belonged to the L3 haplogroup. DNA damage analysis of the reads aligned to the human genome does not rule out their historical origin. In Hallast et al.[79], the global prevalence of Y-haplotypes was investigated using 1208 samples. Of these, 150 samples belonged to the E-haplogroup, of which 83.3% (n=125) were from Africa, 8% (n=12) from Central/West Asia, 6% (n=9), 2% (n=3) from South Asia, and 0.7% (n=1) from North Asia. E2 lineage included 8 samples of exclusively African origin, including 2 E2a∼ (M41) samples from Kenya, one E2a∼ from Sudan, one E2b1a (M85) from Gambia, one E2b1a1 (M200) sample each from Congo, Gambia and Kenya, and one E2b1a1∼ (CTS155) sample from Angola.[79]

However, we cannot be more precise about the age of these sequences. If Sample 2 had been taken from the surface of the mummy, one might think that DNA sequences belonging to the haplogroups mentioned could have remained on it from its handling in Egypt during the modern period. However, as this sample was taken from a depth of only a few millimeters, it is difficult to imagine that human DNA would be present due to human touch. However, as with the bacteria mentioned above, it is possible that some human DNA present in the fluid may have infiltrated the mummy (particularly under the fracture surface) and been found. This could have been the case, for example, if the crypt in which the mummy was kept had been contaminated with human sewage, which could have soaked the mummy. Since this possibility could have occurred before the mummy was placed in the museum, the DNA damage that has occurred in the decades since could have produced the smiley plot patterns we have seen. In addition to these possibilities, the possibility that human DNA was introduced during the making and moulding of the mummy cannot be completely ruled out. Since the mummy is described as having been moulded, it was obviously done by human hands, so the genetic material of the epithelial cells that were shed may have been preserved in a lucky case. At the same time, it cannot be ruled out that the human DNA may have come from human feces (e.g. night soil), which may have been mixed with the seed/soil/water mixture to assist the growth of the grain. It is worth noting that the mummy was dried in the sun during its preparation, creating a dry medium to help the DNA survive. Obviously, it would be better to have a deeper and more complete sequencing, but others have used hashing quality data, Ziesemer et al.[80] used 3-97× depth samples, with coverage between 62-97%.

Conclusion

Studying the corn mummy offers a unique opportunity to peek into an ancient time capsule. However, even with the latest genetic dating methods, interpreting the results requires caution. Even though the samples taken from the corn mummy are likely to contain both modern and ancient genetic material, sequences that are more likely to be from Egypt in the 3rd century BC were identified. Most of the bacterial sequences appear to be of modern origin, while sequences from an ancient wheat cultivar, a poplar tree, a silkmoth-related insect with its gut bacteria, and a North African/sub-Saharan human male appear to be from when the corn mummy was created.

Declarations

Acknowledgements
In memory of Vendel Dolmány OCist. We want to thank Lilla Füleki, a colleague of the National Museum Conservation and Storage Centre (OMRRK) – Museum of Fine Arts, Budapest, Hungary, for her enthusiastic suggestions and support.

Author contribution statement
István Csabai: conceptualization (equal); formal analysis (supporting); funding acquisition (lead); methodology (supporting); writing – review and editing (equal). Flóra Judit Kevély: writing – review and editing (equal). Katalin Kóthay: conceptualization (equal); writing – review and editing (equal). Donát Magyar: conceptualization (equal); writing – review and editing (equal). Gergely Maróti: investigation (supporting); writing – review and editing (equal). Sára Ágnes Nagy: funding acquisition (supporting); investigation (supporting); writing – original draft (equal); writing – review and editing (equal). Bernadett Pap: investigation (lead). Norbert Solymosi: conceptualization (lead); data curation (lead); formal analysis (lead); methodology (lead); writing – original draft (lead); writing – review and editing (lead). Adrienn Gréta Tóth: investigation (supporting); writing – original draft (equal); writing – review and editing (equal).

Competing interests
The authors declare that they have no competing interests.

Data Availability
The raw short read data are publicly available and accessible through the PRJNA1128416 (Sample 1: SAMN42050348; Sample 2: SAMN42050349) from the NCBI Sequence Read Archive (SRA).

Ethics approval and consent to participate
Not applicable.

Funding
The study was supported by the European Union's Horizon 2020 research and innovation program (No. 874735, VEO) and by the European Union project RRF-2.3.1-21-2022-00004 within the framework of the MILAB Artificial Intelligence National Laboratory.

References

1. Seersholm, F. V. *et al.* DNA evidence of bowhead whale exploitation by Greenlandic Paleo-Inuit 4,000 years ago. *Nat. Commun.* **7**, 13389 (2016).
2. Brealey, J. C. *et al.* Dental calculus as a tool to study the evolution of the mammalian oral microbiome. *Mol. Biol. Evol.* **37**, 3003–3022 (2020).
3. Nishimura, L. *et al.* Identification of ancient viruses from metagenomic data of the Jomon people. *J. Hum. Genet.* **66**, 287–296 (2021).
4. Jäger, H. Y. *et al.* Metagenomic analysis reveals mixed Mycobacterium tuberculosis infection in a 18th century Hungarian midwife. *Tuberculosis* **137**, 102181 (2022).
5. Neukamm, J. *et al.* 2000-year-old pathogen genomes reconstructed from metagenomic analysis of Egyptian mummified individuals. *BMC Biol.* **18**, 1–18 (2020).
6. Khairat, R. *et al.* First insights into the metagenome of Egyptian mummies using next-generation sequencing. *J. Appl. Genet.* **54**, 309–325 (2013).
7. Selin, H. Encyclopaedia of the history of science, technology, and medicine in non-western cultures springer-verlag berlin heidelberg new york 2008. (2008).
8. Enany, A. A Corn-Mummy from the Egyptian Museum in Cairo. *J. Assoc. Arab. Univ. for Tour. Hosp.* **15**, 52–62 (2018).
9. Hassan, R. R. A. A Recently Discovered Corn Mummy In Ibis, South Of Abu Sir, North of Saqqara 2018: A Case Study. *Int. J. Conserv. Sci.* **12**, 879–892 (2021).
10. Centrone, M. C. *Egyptian corn-mummies: a class of religious artefacts catalogued and systematically analysed* (VDM, Verlag Dr. Müller, 2009).
11. Liptay, E., Mekis, T. & Varga, E. Supplements for a better understanding of the Osiris rites of the month of Khoiak. *Bulletin du Musée Hongrois des Beaux-Arts* **127–128**, 33–60 (2021-2022).

12. Arbøll, T. P. *et al.* Revealing the secrets of a 2900-year-old clay brick, discovering a time capsule of ancient DNA. *Sci. Reports* **13**, 13092 (2023).

13. Kistler, L. *Ancient DNA Extraction from Plants*, 71–79 (Humana Press, Totowa, NJ, 2012).

14. Latorre, S. M., Lang, P. L., Burbano, H. A. & Gutaker, R. M. Isolation, library preparation, and bioinformatic analysis of historical and ancient plant DNA. *Curr. Protoc. Plant Biol.* **5**, e20121 (2020).

15. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome biology* **20**, 1–13 (2019).

16. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D61–D65 (2007).

17. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).

18. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinforma.* **10**, 1–9 (2009).

19. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2015).

20. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

21. Wang, Z., Huang, P., You, R., Sun, F. & Zhu, S. MetaBinner: a high-performance and stand-alone ensemble binning method to recover individual genomes from complex microbial communities. *Genome Biol.* **24**, 1 (2023).

22. Liu, C.-C. *et al.* MetaDecoder: a novel method for clustering metagenomic contigs. *Microbiome* **10**, 46 (2022).

23. Pan, S., Zhao, X.-M. & Coelho, L. P. SemiBin2: self-supervised contrastive learning leads to better MAGs for short-and long-read sequencing. *Bioinformatics* **39**, i21–i29 (2023).

24. Sieber, C. M. *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836–843 (2018).

25. Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624 (2016).

26. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

27. Danecek, P. *et al.* Twelve years of samtools and bcftools. *Gigascience* **10**, giab008 (2021).

28. Schönherr, S., Weissensteiner, H., Kronenberg, F. & Forer, L. Haplogrep 3-an interactive haplogroup classification and analysis platform. *Nucleic Acids Res.* **51**, W263–W268 (2023).

29. Weissensteiner, H. *et al.* Contamination detection in sequencing studies using the mitochondrial phylogeny. *Genome Res.* **31**, 309–316 (2021).

30. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 1–18 (2015).

31. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).

32. Gleize, Y. *et al.* Early Medieval Muslim Graves in France: First Archaeological, Anthropological and Palaeogenomic Evidence. *PLOS One* **11**, 1–13 (2016).

33. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

34. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinforma.* **15**, 1–13 (2014).

35. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using dna shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013).

36. Martiniano, R., De Sanctis, B., Hallast, P. & Durbin, R. Placing ancient DNA sequences into reference phylogenies. *Mol. Biol. Evol.* **39**, msac017 (2022).

37. Poznik, G. D. Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. *bioRxiv* (2016).

38. Cabrera, V. M., Marrero, P., Abu-Amero, K. K. & Larruga, J. M. Carriers of mitochondrial DNA macrohaplogroup L3 basal lineages migrated back to Africa from Asia around 70,000 years ago. *BMC Evol. Biol.* **18**, 1–16 (2018).

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

40. Schubert, M., Lindgreen, S. & Orlando, L. Adapterremoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 1–7 (2016).

41. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

42. Xu, S. *et al.* Ggtree: a serialized data object for visualization of a phylogenetic tree and annotation data. *IMeta* **1**, e56 (2022).

43. Schliep, K., Potts, A. J., Morrison, D. A. & Grimm, G. W. Intertwining phylogenetic trees and networks. *Methods Ecol. Evol.* **8**, 1212–1220 (2017).

44. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2023).

45. Hu, P., Brodie, E. L., Suzuki, Y., McAdams, H. H. & Andersen, G. L. Whole-genome transcriptional analysis of heavy metal stresses in Caulobacter crescentus. *J. Bacteriol.* **187**, 8437–8449 (2005).

46. Pages, D. *et al.* Heavy metal tolerance in Stenotrophomonas maltophilia. *PLOS One* **3**, e1539 (2008).

47. Janssen, P. J. *et al.* The complete genome sequence of Cupriavidus metallidurans strain CH34, a master survivalist in harsh and anthropogenic environments. *PLOS One* **5**, e10433 (2010).

48. Guo, F. *et al.* An effective degumming enzyme from Bacillus sp. Y1 and synergistic action of hydrogen peroxide and protease on enzymatic degumming of ramie fibers. *BioMed Res. Int.* **2013** (2013).

49. Asif, M. *et al.* Comprehensive genomic analysis of Bacillus paralicheniformis strain BP9, pan-genomic and genetic basis of biocontrol mechanism. *Comput. Struct. Biotechnol. J.* **21**, 4647–4662 (2023).

50. Li, H., Su, J.-Q., Yang, X.-R. & Zhu, Y.-G. Distinct rhizosphere effect on active and total bacterial communities in paddy soils. *Sci. Total. Environ.* **649**, 422–430 (2019).

51. Lee, S.-Y., Son, J.-S., Hwang, Y.-J., Shin, J.-H. & Ghim, S.-Y. Metabacillus elymi sp. nov., isolated from the Rhizosphere of Elymus tsukushiensis, a plant native to the Dokdo Islands, Republic of Korea. *Antonie van Leeuwenhoek* **114**, 1709–1719 (2021).

52. Aziz, N. M. A., Goda, D. A., Abdel-Meguid, D. I., El-Sharouny, E. E. & Soliman, N. A. A comparative study of the biosynthesis of CuNPs by Niallia circulans G9 and Paenibacillus sp. S4c strains: characterization and application as antimicrobial agents. *Microb. Cell Factories* **23**, 156 (2024).

53. Nie, K. *et al.* Roseburia intestinalis: a beneficial gut organism from the discoveries in genus and species. *Front. Cell. Infect. Microbiol.* **11**, 757718 (2021).

54. Nakase, K. *et al.* Cutibacterium acnes phylogenetic type ic and ii isolated from patients with non-acne diseases exhibit high-level biofilm formation. *Int. J. Med. Microbiol.* **311**, 151538 (2021).

55. Gohar, D. *et al.* Global diversity and distribution of mushroom-inhabiting bacteria. *Environ. Microbiol. Reports* **14**, 254–264 (2022).

56. Jiang, S. *et al.* Biodegradation of Oil by a Newly Isolated Strain Acinetobacter Junii WCO-9 and Its Comparative Pan-Genome Analysis. *Microorganisms* **11**, 407 (2023).

57. Zhu, J. *et al.* Isolation, identification, and carbonate mineralization characteristics of a newly carbonic anhydrase–producing strain. *Appl. Biochem. Biotechnol.* 1–17 (2024).

58. Takarada, H. *et al.* Complete genome sequence of the soil actinomycete Kocuria rhizophila. *J. Bacteriol.* **190**, 4139–4146 (2008).

59. Yasuda, S., Suenaga, T. & Terada, A. Complete genome sequence of Methylosinus sp. strain C49, a methane-oxidizing bacterium harboring phaABC genes for polyhydroxyalkanoate synthesis. *Microbiol. Resour. Announc.* **9**, 10–1128 (2020).

60. Huang, C.-H. *et al.* Reclassification of Micrococcus aloeverae and Micrococcus yunnanensis as later heterotypic synonyms of Micrococcus luteus. *Int. J. Syst. Evol. Microbiol.* **69**, 3512–3518 (2019).

61. Mello, A. *et al.* Bacterial and fungal communities associated with tuber magnatum-productive niches. *Plant Biosyst.* **144**, 323–332 (2010).

62. d Santos, A. F., Pacheco, C. A., Valle, R. d. S., Seldin, L. & d Santos, A. L. S. Enzymes produced by halotolerant spore-forming gram-positive bacterial strains isolated from a resting habitat (Restinga de Jurubatiba) in Rio de Janeiro, Brazil: focus on proteases. *Appl. Biochem. Biotechnol.* **174**, 2748–2761 (2014).

63. Quaiser, A. *et al.* Diversity and comparative genomics of Microviridae in Sphagnum-dominated peatlands. *Front. Microbiol.* **6**, 375 (2015).

64. Song, Y. *et al.* Alistipes onderdonkii sp. nov. and Alistipes shahii sp. nov., of human origin. *Int. J. Syst. Evol. Microbiol.* **56**, 1985–1990 (2006).

65. Parker, B. J., Wearsch, P. A., Veloo, A. C. & Rodriguez-Palacios, A. The genus Alistipes: gut bacteria with emerging implications to inflammation, cancer, and mental health. *Front. Immunol.* **11**, 906 (2020).

66. Nilsen, M. *et al.* A globally distributed Bacteroides caccae strain is the most prevalent mother-child shared Bacteroidaceae strain in a large Scandinavian cohort. *Appl. Environ. Microbiol.* **89**, e00789–23 (2023).

67. Morotomi, M., Nagai, F., Sakon, H. & Tanaka, R. Dialister succinatiphilus sp. nov. and Barnesiella intestinihominis sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* **58**, 2716–2720 (2008).

68. Taras, D., Simmering, R., Collins, M. D., Lawson, P. A. & Blaut, M. Reclassification of Eubacterium formicigenerans Holdeman and Moore 1974 as Dorea formicigenerans gen. nov., comb. nov., and description of Dorea longicatena sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* **52**, 423–428 (2002).

69. Zou, Y. *et al.* Characterization and description of Faecalibacterium butyricigenerans sp. nov. and F. longum sp. nov., isolated from human faeces. *Sci. Reports* **11**, 11340 (2021).

70. Hiippala, K. *et al.* Novel Odoribacter splanchnicus strain and its outer membrane vesicles exert immunoregulatory effects in vitro. *Front. Microbiol.* **11**, 575455 (2020).

71. Evans, J. C. *et al.* A proteolytically activated antimicrobial toxin encoded on a mobile plasmid of Bacteroidales induces a protective response. *Nat. Commun.* **13**, 4258 (2022).

72. He, J., Zhang, Q. & Achal, V. Heavy metals immobilization in soil with plant-growthpromoting precipitation in support of radish growth. *Microbiol. Biotechnol. Lett.* **48**, 223–229 (2020).

73. Golovnina, K. *et al.* Molecular phylogeny of the genus triticum l. *Plant Syst. Evol.* **264**, 195–216 (2007).

74. Lippi, M. M. & Mercuri, A. Palynology of a resin from an Egyptian coffin of the second century BC. *Rev. Palaeobot. Palynol.* **71**, 207–218 (1992).

75. Prem Anand, A. A. *et al.* Isolation and characterization of bacteria from the gut of Bombyx mori that degrade cellulose, xylan, pectin and starch and their impact on digestion. *J. Insect Sci.* **10**, 107 (2010).

76. Robins, G. *Women in ancient Egypt* (Harvard University Press, 1993).

77. Soares, P. *et al.* The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Mol. Biol. Evol.* **29**, 915–927 (2011).

78. Rishishwar, L. & Jordan, I. K. Implications of human evolution and admixture for mitochondrial replacement therapy. *BMC Genomics* **18**, 1–11 (2017).

79. Hallast, P., Agdzhoyan, A., Balanovsky, O., Xue, Y. & Tyler-Smith, C. A Southeast Asian origin for present-day non-African human Y chromosomes. *Hum. Genet.* **140**, 299–307 (2021).

80. Ziesemer, K. A. *et al.* The efficacy of whole human genome capture on ancient dental calculus and dentin. *Am. J. Phys. Anthropol.* **168**, 496–509 (2019).

81. Scott, M. F. *et al.* A 3,000-year-old Egyptian emmer wheat genome reveals dispersal and domestication history. *Nat. Plants* **5**, 1120–1128 (2019).

Supplementary information

Table S1. Summary statistics of bacterial genomes assembled from sample 1 by metagenome binning. Mapped reads represent the number of reads that can be mapped to the metagenome assembly contigs. The average nucleotide identity (ANI) value is given in parentheses after the assembly name with the best match based on prokaryotic genome annotation pipeline (PGAP) analysis.

	Cupriavidus metallidurans	*Caulobacter segnis*	*Stenotrophomonas maltophilia*
N50	284,106	98,488	24,502
Contigs	62	167	254
Genome size (bp)	6,244,478	4,755,198	4,428,340
GC content	0.64	0.68	0.66
Average coverage (fold)	203	13	23
Protein-coding genes	5,594	4,431	4,048
5S rRNA	1	0	1
16S rRNA	1	0	0
23S rRNA	1	0	0
tRNAs	52	47	59
ncRNAs	4	4	4
Putative CRISPR count	0	0	0
Completeness	94.68	90.73	93.41
Contamination	1.78	5.22	0.66
Strain heterogeneity	0	3.17	0
PGAP best match (ANI)	ASM19601v1 (99.22)	ASM9228v1 (84.98)	XMT-16JAN14 (92.95)

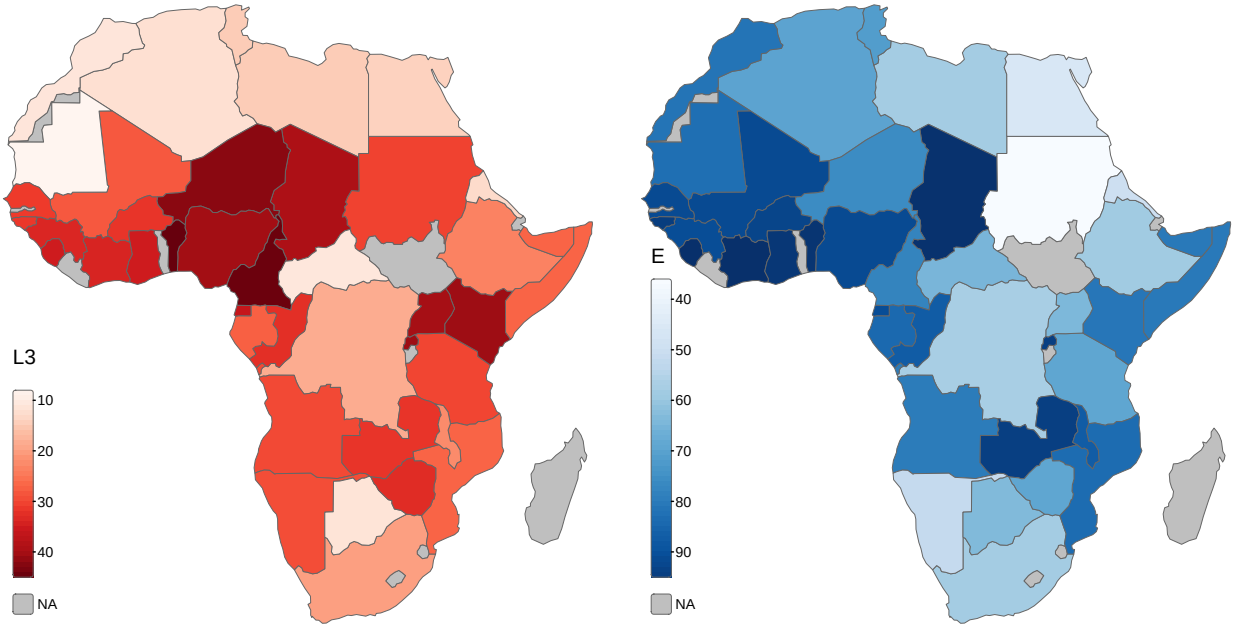

Figure S1. Mean frequencies of mtDNA L3 and Y-chromosome E haplogroup in African countries mapped based on Supplementary Table S2 of Cabrera et al.[38]

Table S2. The *Triticum* and *Hordeum* samples were used in phylogenetic analysis of the trit_max_contig sequence. The ancient sample Tt_UC10164 dated to 1130–1000 BC,[81] while similar estimation for the age of southern Levant samples is not available (BioProject PRJEB64853: *"The history of domesticated tetraploid wheat has been studied extensively for many decades, yet key questions in the evolutionary history of this crop remain elusive. To elucidate the evolutionary trajectory of tetraploid wheat, we sequenced ancient tetraploid wheat samples excavated at different locations along the Judean desert and the Arava valley in Israel."*).

Biosample	Run	ID	Origin	Species
Ancient				
SAMEA5328956	UC10164	Tt_UC10164	Egypt	*T. turgidum* ssp. *dicoccum*
SAMEA114240248	TU1126	Tt_TU1126	southern Levant	*T. turgidum* ssp. *dicoccum*
SAMEA114240249	TU1128	Tt_TU1128	southern Levant	*T. turgidum* ssp. *dicoccum*
SAMEA114240250	TU1129	Tt_TU1129	southern Levant	*T. turgidum* ssp. *dicoccum*
SAMEA114240251	TU1130	Tt_TU1130	southern Levant	*T. turgidum* ssp. *dicoccum*
SAMEA114240252	TU1131	Tt_TU1131	southern Levant	*T. turgidum* ssp. *dicoccum*
SAMEA114240253	TU1145	Tt_TU1145	southern Levant	*T. turgidum* ssp. *dicoccum*
SAMEA114240254	TU1146	Tt_TU1146	southern Levant	*T. turgidum* ssp. *dicoccum*
SAMEA114240255	TU1147	Tt_TU1147	southern Levant	*T. turgidum* ssp. *dicoccum*
SAMEA114240256	TU1148	Tt_TU1148	southern Levant	*T. turgidum* ssp. *dicoccum*
SAMEA114240257	TU1149	Tt_TU1149	southern Levant	*T. turgidum* ssp. *dicoccum*
SAMEA114240258	TU1150	Tt_TU1150	southern Levant	*T. turgidum* ssp. *dicoccum*
SAMEA114240259	TU686	Tt_TU686	southern Levant	*T. turgidum ssp. dicoccum*
SAMEA114240260	TU692	Tt_TU692	southern Levant	*T. turgidum* ssp. *dicoccum*
SAMEA114240246	JK2283	Tt_JK2283	southern Levant	*T. turgidum* ssp. *dicoccum*
SAMEA114240247	JK3015	Tt_JK3015	southern Levant	*T. turgidum* ssp. *dicoccum*
Present				
SAMEA13856358	ERR11272062	Tm_TA759	Saudi Arabia	*T. monococcum*
SAMEA112910807	SAMEA112910807	Tt_DEW_PI_196100	?	*T. turgidum* ssp. *dicoccum*
SAMN13640588	SRR10762479	Ta_Altgold	Germany	*T. aestivum*
SAMN13640593	SRR10762510	Ta_Apache	France	*T. aestivum*
SAMN13640590	SRR10762513	Ta_Hubel	Germany	*T. aestivum*
SAMN13640700	SRR10812078	Ta_24056	Turkey	*T. aestivum*
SAMN13640694	SRR10812119	Ta_23923	Ethiopia	*T. aestivum*
SAMN13640693	SRR10812130	Ta_23909	Morocco	*T. aestivum*
SAMN13640701	SRR10812140	Ta_24184	Israel	*T. aestivum*
SAMN13640692	SRR10812141	Ta_23896	Turkey	*T. aestivum*
SAMN13640786	SRR10838845	Ta_AZTEC	France	*T. aestivum*
SAMN13640758	SRR10838857	Ta_IG21134470	Egypt	*T. aestivum*
SAMN13640757	SRR10838868	Ta_IG21134447	Jordan	*T. aestivum*
SAMN13640756	SRR10838869	Ta_IG21134349	Turkey	*T. aestivum*
SAMN16131686	SRR12640151	Ta_PI427645	?	*T. monococcum*
SAMN20819193	SRR15494265	Ta_PI345242	Macedonia	*T. monococcum*
SAMN40712268	SRR28520609	H_vulgare	Sweden	*H. vulgare*

Table S3. Chromosomal hits (scaffoldes excluded) obtained by BLAST aligment the trit_max_contig sequence against the reference genome of *Triticum aestivum* (cultivar Chinese Spring, IWGSC CS RefSeq v2.1).

Chromosome	Max Score	Total Score	Query Cover	E value	Per. ident
7A	2697	4958000	100%	0	98.01
2B	2691	6071000	100%	0	97.94
3A	2691	5362000	100%	0	97.94
1B	2689	5429000	100%	0	97.94
4A	2686	6131000	100%	0	97.88
6B	2686	7340000	100%	0	97.88
4B	2686	6610000	100%	0	97.88
7B	2686	6607000	100%	0	97.88
5A	2680	4896000	100%	0	97.81
5B	2675	5528000	100%	0	97.75
3B	2673	7147000	100%	0	97.75
2A	2662	6089000	100%	0	97.62
6A	2662	5210000	100%	0	97.62
1A	2656	4162000	100%	0	97.55

Table S4. Chromosomal hits (scaffoldes excluded) obtained by BLAST aligment the trit_max_contig sequence against the reference genome of *Hordeum vulgare* subsp. *vulgare* (MorexV3_pseudomolecules_assembly).

Chromosome	Max Score	Total Score	Query Cover	E value	Per. ident
2H	579	869700	94%	3E-162	78.39
4H	569	906000	94%	2E-159	78.29
6H	564	812300	96%	8E-158	78.25
7H	564	778400	96%	8E-158	78.12
5H	562	727300	95%	3E-157	74.72
3H	556	777100	94%	1E-155	78.01
1H	555	665700	94%	5E-155	78.03

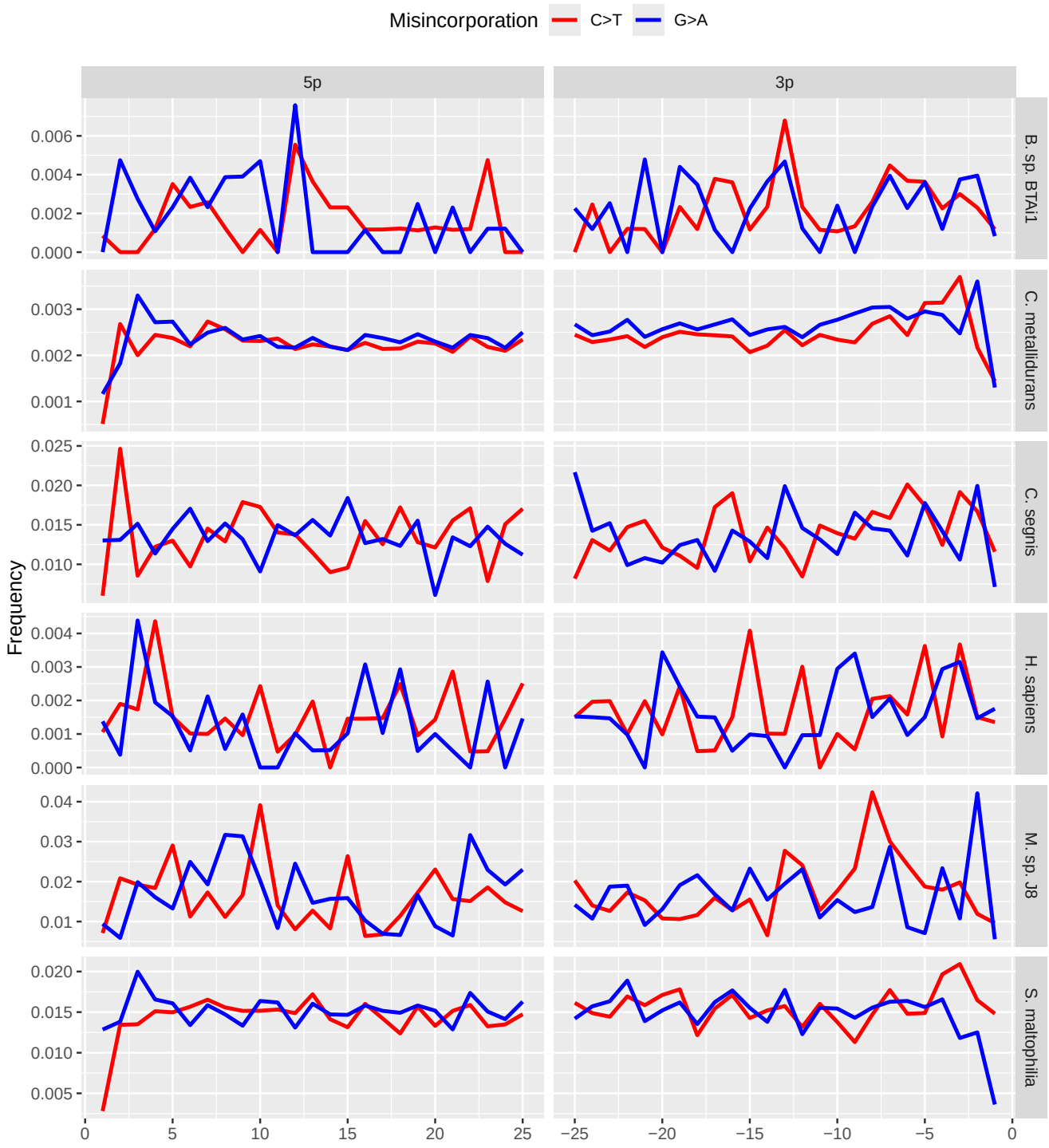

Figure S2. DNA damage curves for genomes in Sample 1. The species analyzed: *Bradyrhizobium* sp. BTAi1 (reference genome: GCF_000015165.1), *Caulobacter segnis* (GCF_003015125.1), *Cupriavidus metallidurans* (GCF_000196015.1), *Homo sapiens* (GCA_000001405.15), *Mesorhizobium* sp. J8 (GCF_016591715.1), *Stenotrophomonas maltophilia* (GCF_900475405.1).

.

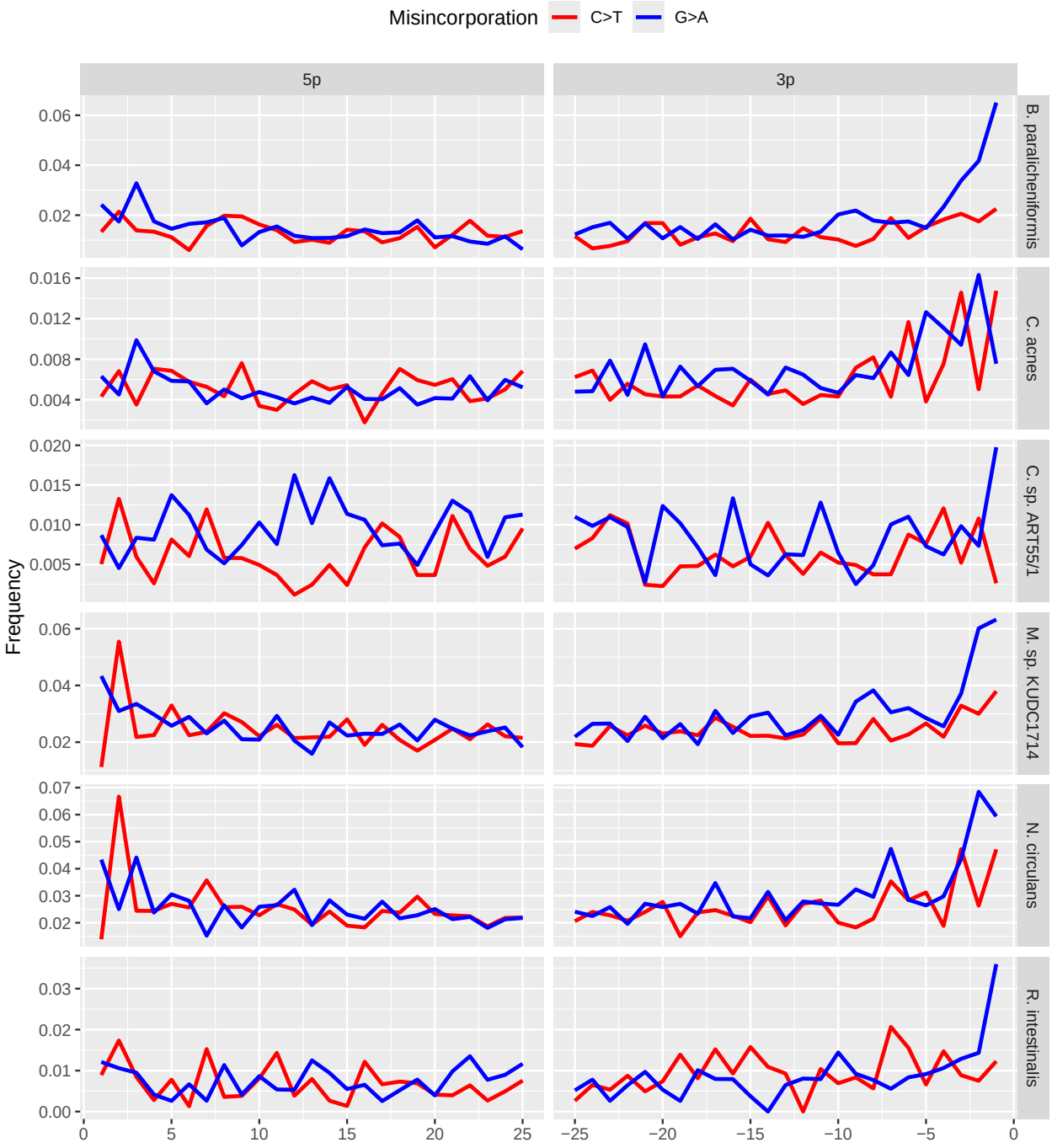

Figure S3. DNA damage curves for bacteria in Sample 2. The species analyzed: *Bacillus paralicheniformis* (reference genome: GCF_002993925.1), *Coprococcus* sp. ART55/1 (GCF_000210595.1), *Cutibacterium acnes* (GCF_000376705.1), *Metabacillus* sp. KUDC1714 (GCF_014217835.1), *Niallia circulans* (GCF_013267435.1), *Roseburia intestinalis* (GCF_900537995.1).

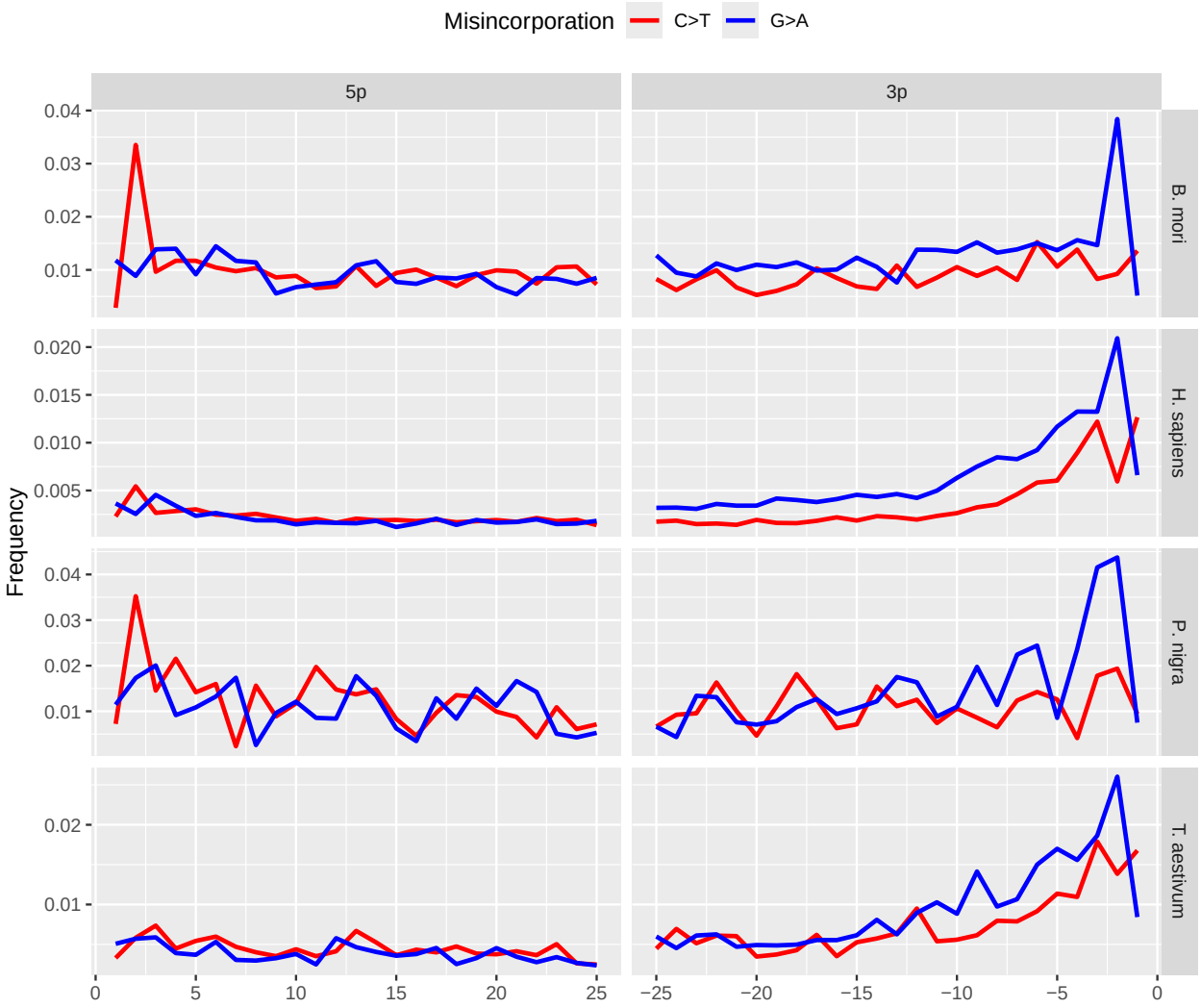

Figure S4. DNA damage curves for eukaryotes in Sample 2. The species analyzed: *Bombyx mori* (reference genome: GCF_030269925.1), *Homo sapiens* (GCA_000001405.15), *Populus nigra* (GCF_951802175.1), *Triticum aestivum* (GCF_018294505.1).