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 Physics 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*, *Mesobacillus*, *Metabacillus*, *Neobacillus*, *Niallia*, *Peribacillus* and *Paenibacillus*, 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. The fragments identified as of Lepidopteran origin showed the greatest similarity to the genomes of the Sphingidae. Analysis of the human-derived sequences revealed L3 (mtDNA), E and J (Y chromosome) haplotypes, which are common lineages in 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 ($\sim$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. Our study aimed to see if ancient DNA could be preserved in the mummy and, if so, what organisms it might have come from. To this end, metagenomic analysis was performed 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. Experts estimate it to be from Central Egypt, dating back 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 following 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 following 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 resuspended 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 resuspended 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 paired-end FASTQ files were merged, trimmed and filtered by AdapterRemoval (v3.0.0-alpha2, `--trim-ns --trim-qualities --trim-min-quality 20 --min-length 30 --min-complexity 0.3`).[15] The remaining reads were taxonomically classified by Kraken2 (v2.1.3, `--confidence 0.2`)[16] on the NCBI Core NT[17] database (built: 28/12/2024). The taxon abundance results were adjusted by Bracken (v3.1).[18] The de novo assembly was performed by MEGAHIT (v1.2.9)[19] with default settings. The assembled contigs to check the most probable origin were aligned by local BLAST[20] on NCBI Core NT database (updated: 11/2/2025). For all contig and reference genome based aligment BWA (v0.7.17-r1188)[21] was used with the settings: `aln -n 0.01 -l 16500 -o 2`.[22] Duplums of the reads were removed from the alignments using Picard (v2.26.6, https://broadinstitute.github.io/picard/). Post-mortem DNA damage was analysed by estimating the frequency of C>T and G>A misincorporation using MapDamage (v2.3.0a0).[23] The alignments used for this purpose were produced on the de novo assembled contigs and reference genomes of the most abundant species of genera with at least 1% relative abundance.

The phylogenetic analysis of modern and ancient samples of the genus *Triticum* and *Hordeum* was based on the longest contig (PV405787) 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 PV405787 sequence of the genus *Triticum* can be found in Table 1. From the generated BAM files, we constructed consensus sequences using samtools, keeping those with an N ratio of less than 5%. The gene-tree was constructed[24] based on multiple sequence alignment by MAFFT (v7.490, `--auto --adjustdirection`).[25] The best substitution model was selected by functions of phangorn (v2.12.1) package[26] 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 (v4.4.3).[27]

To estimate the contamination of human genome originated fragments, reads were aligned to mitochondria and chromosome X sequences. The reads aligned to NC_012920.1 sequence (Revised Cambridge Reference Sequence, rCRS) were analysed by Haplocheck (v1.3.3)[29] and ContamMix (v1.0-11).[30] By ContamMix 311 worldwide mitogenomes[31] were compared to our mitochondrial consensus sequence based on multiple sequence alignment by MAFFT. The schmutzi (v1.5.7)[32] applied contamination analysis was performed on the Reconstructed Sapiens Reference Sequence (RSRS) alignment. For the X chromosome contamination analysis ANGSD (v0.940, `X:5000000-154900000`, `-doCounts 1 -iCounts 1 -minMapQ 30 -minQ 20`)[33] package, based on NC_000023.10 (hg19) alignment was used. Karyotype assignment was done by alignment on X and Y chromosomes (hs37d5) with a minimum mapping quality of 30.[34] For the mtDNA haplogroup classification using the RSRS-based alignment, the mitochondrial consensus sequence is called by schmutzi, and assigned by Haplogrep3 (v3.2.2, metric: Kulczyński, PhyloTree Build: 17).[35] The Y chromosome-based haplogroup prediction was performed by pathPhynder[36] and yHaplo (v2.1.12.dev9+geb252e7, `--anc_stop_thresh 10`)[37], using the hs37d5 genomes. The VCF input of yHaplo was called by bcftools (v1.14, `--min-MQ 30 --min-BQ 30 --ploidy 1`).[38]

Results

The number of cleaned, merged reads was 2,933,736 in the Sample 1 sample and 3,506,737 in the Sample 2 sample. The length of the contigs generated by de novo assembly was longer in Sample 1 (n: 12,479, median: 482, min: 201, max: 540,009,

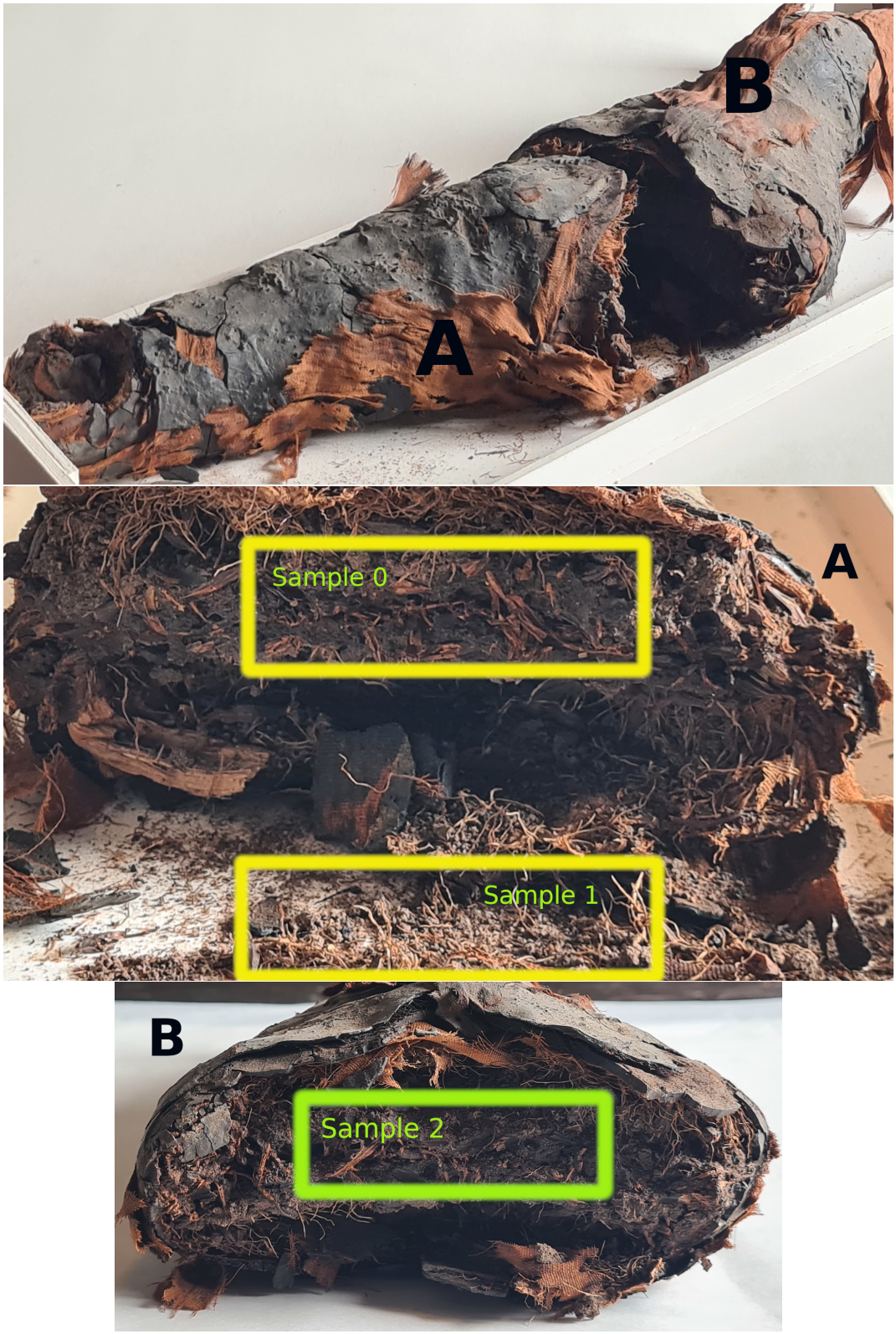

Figure 1. Corn mummy 68.2-E sampled. The top image shows the whole mummy, the middle image shows part A with sampling areas 0 and 1, and the bottom image shows part B with sampling area 2.

IQR: 453) than in Sample 2 (n: 12,851, median: 384, min: 257, max: 8,314, IQR: 147). The length distributions of the reads

Table 1. The *Triticum* and *Hordeum* samples were used in phylogenetic analysis by the longest wheat origin contig (PV405787). The ancient sample Tt_UC10164 dated to 1130–1000 BC,[28] while a similar estimation for the age of southern Levant samples obtained from the BioProject PRJEB64853 is not available. Of the samples marked with *, the contigs overlapping the mummy sequence generated from the reads contained more than 5% N and were not used in the phylogenetic analysis.

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	Tm_PI427645	?	*T. monococcum*
SAMN20819193*	SRR15494265	Ta_PI345242	Macedonia	*T. monococcum*
SAMN40712268	SRR28520609	H_vulgare	Sweden	*H. vulgare*

realigned on these contigs, and the misincorporation pattern is shown in Figure 2. While the read length distribution of Sample 2 has a peak of 84 bp, that of Sample 1 is 170. The contig and read lengths distribution with the DNA damage pattern shown in Figure 2 suggest that Sample 1 is rather a modern contamination than historical DNA. The length distribution and the misincorporation pattern of Sample 2 suggest its ancient DNA origin. Considering this, only the analysis results of Sample 2 are presented below.

Taxon classification of Sample 2 with 0.2 confidence resulted in 57.9% of hits being bacteria, 41.6% eukaryote, and the remaining viruses and archea below 1%. Within the Bacteria, Bacilliota accounted for 83.2%, Actinomycetota for 7.7%, Pseudomonadota for 6.5%, and Bacteroidota for 2.3%. Within the Eukaryota, 88.2% are Chordata, 7.8% Streptophyta, and 2.7% Arthropoda. The genera with a minimal relative abundance of 1% were *Homo* (36.6%), *Niallia* (19.3%), *Bacillus* (8.1%), *Metabacillus* (8%), *Triticum* (2.6%), *Peribacillus* (2.3%), *Neobacillus* (2.2%), *Paenibacillus* (1.8%), *Cutibacterium* (1.7%), *Mesobacillus* (1.4%), *Bombyx* (1%). Within these genera, the most abundant taxa identified at the species level were *Homo sapiens*, *Niallia circulans*, *Bacillus paralicheniformis*, *Metabacillus* sp. KUDC1714, *Triticum aestivum*, *Cutibacterium acnes*, *Peribacillus frigoritolerans*, *Mesobacillus maritimus*, *Bombyx mori*, *Neobacillus* sp. DY30, *Paenibacillus* sp. FSL R5-0766. The DNA damage and length distribution patterns of reads aligned on the reference genomes of these species are shown in

Figure 3-5.

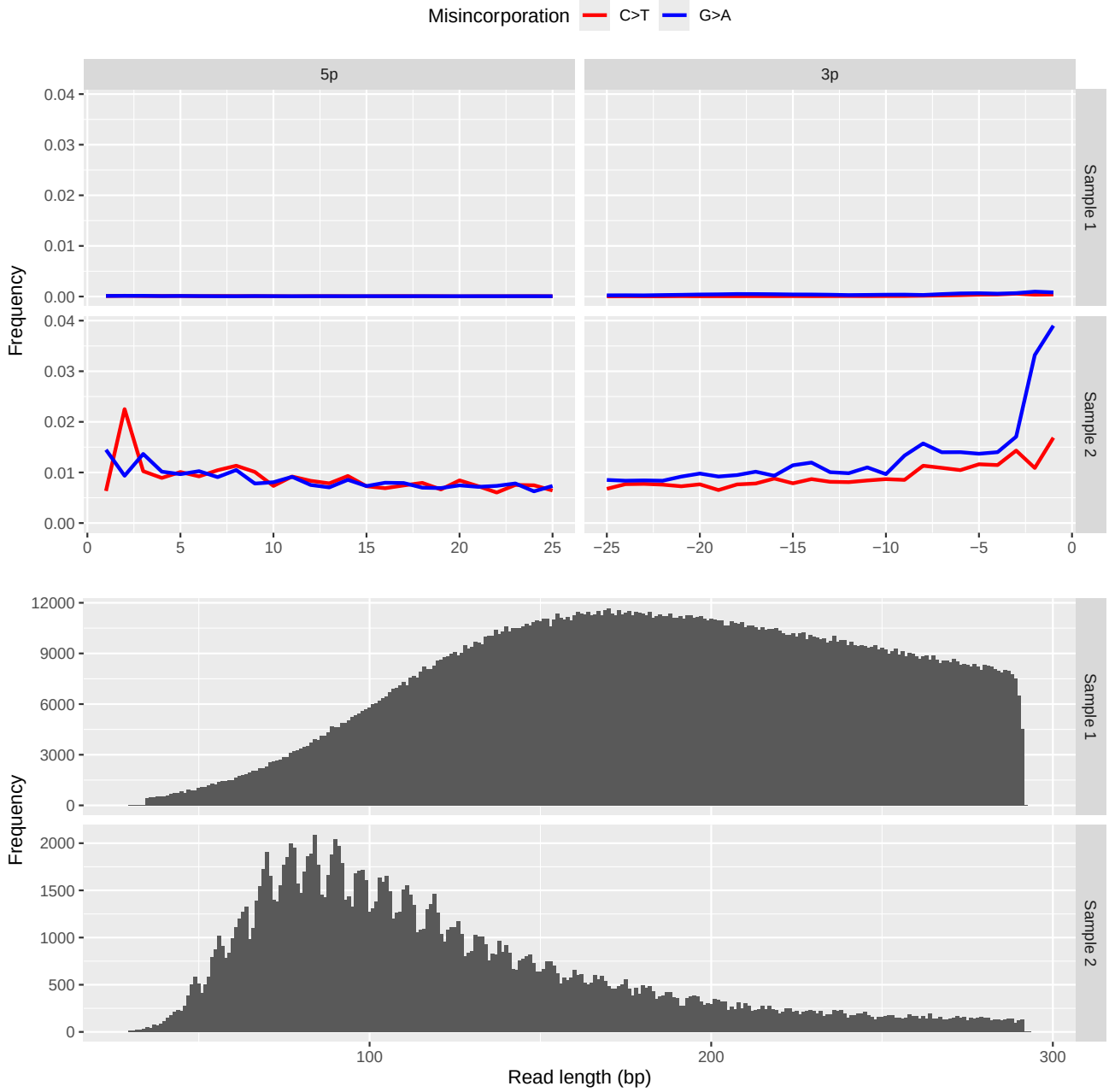

Figure 2. Misincorporation and read length plots. The number of aligned reads on the de novo assembled contigs were 2,010,050 and 170,184 in Samples 1 and 2, respectively.

None of the genes used in the literature for the phylogenetic analysis of *Triticum* could be reconstructed from the reads. Therefore, the longest contig (1241 bp, GenBank: PV405787) with high similarity to Poales was used for this purpose. From the BLAST matches of this sequence against the reference genomes of the genera *Hordeum*, *Triticum*, those covering 100% length were the following. It showed sequence identity of 71.1-72.6% with chromosome 1H, 2H, 3H, 4H, 5H, 6H, 7H parts of the genome of *Hordeum vulgare* subsp. *vulgare* (CAJHDD000000000.1). While the sequence identity to chromosome fragments 1A, 1B, 2A, 2B, 3A, 3B, 4A, 4B, 5A, 5B, 6A, 6B, 7A, 7B of the *Triticum aestivum* genome (IWGSC CS RefSeq v2.1, JAGHKL00000000000.1) was between 98.2 and 99.1%. As the contigs overlapping the mummy sequence generated from the reads from sample SAMN20819193, SAMEA114240249, SAMEA114240252, SAMEA114240256, SAMEA114240259,

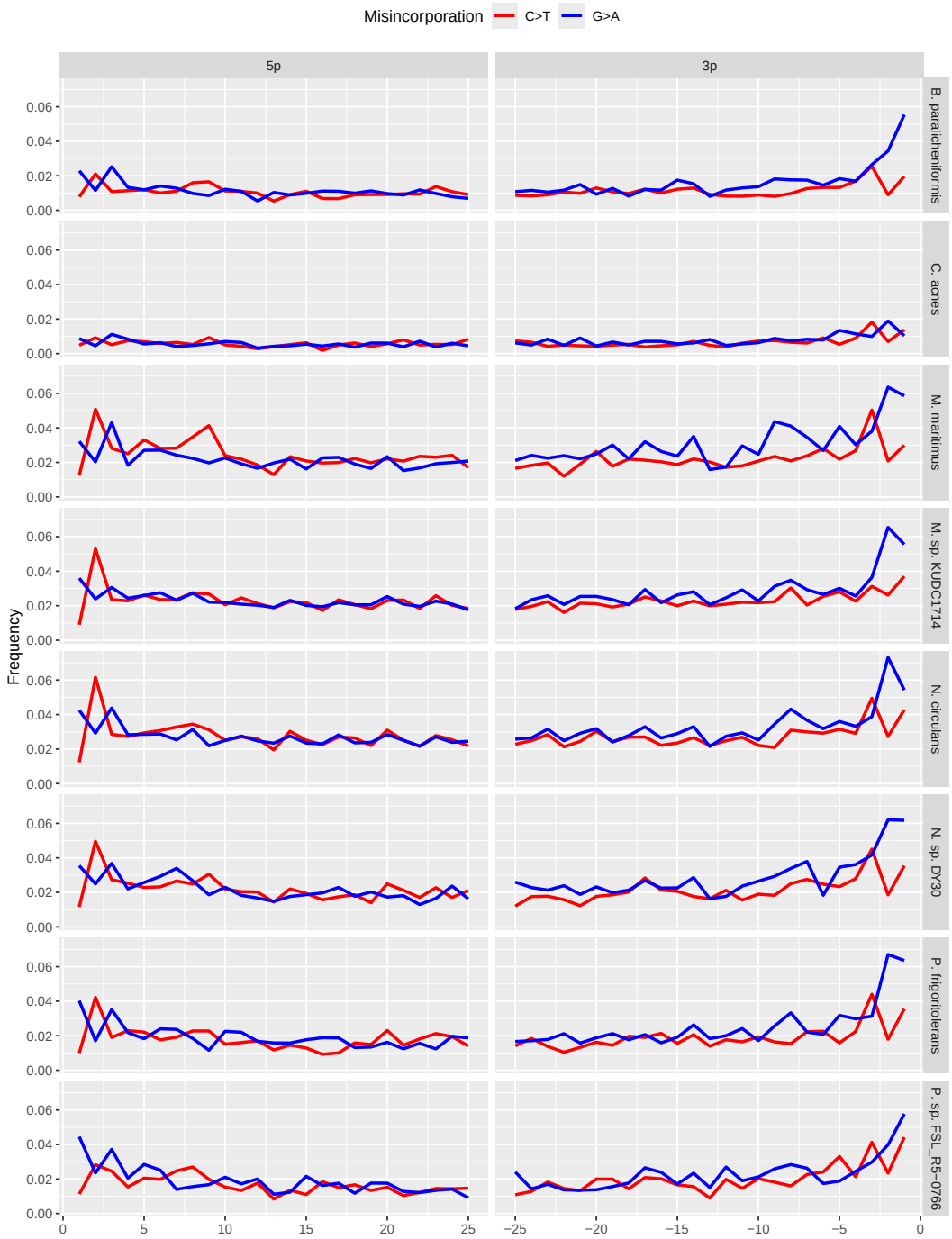

Figure 3. Damage plots of representative bacteria of the genera with at least 1% relative abundance. The genomes were used *Bacillus paralicheniformis* (GCF_002993925.1), *Cutibacterium acnes* (GCF_006739385.1), *Mesobacillus maritimus* (GCF_044803185.1), *Metabacillus* sp. KUDC1714 (GCF_014217835.1), *Neobacillus* sp. DY30 (GCF_030123065.1), *Niallia circulans* (GCF_013267435.1), *Peribacillus frigoritolerans* (GCF_022394675.1), *Paenibacillus* sp. FSL_R5-0766 (GCF_037971845.1).

SAMEA114240260 contained more than 5% N, they were not used in the multiple sequence alignment. The results of the phylogenetic analysis (with the best substitution model F81) based on the sequence PV405787 classifiable in the genus *Triticum* are shown in Figure 6.

Aligning cleaned reads to the human genome, the average coverage of the 22 chromosomes (excluding Y and MT)

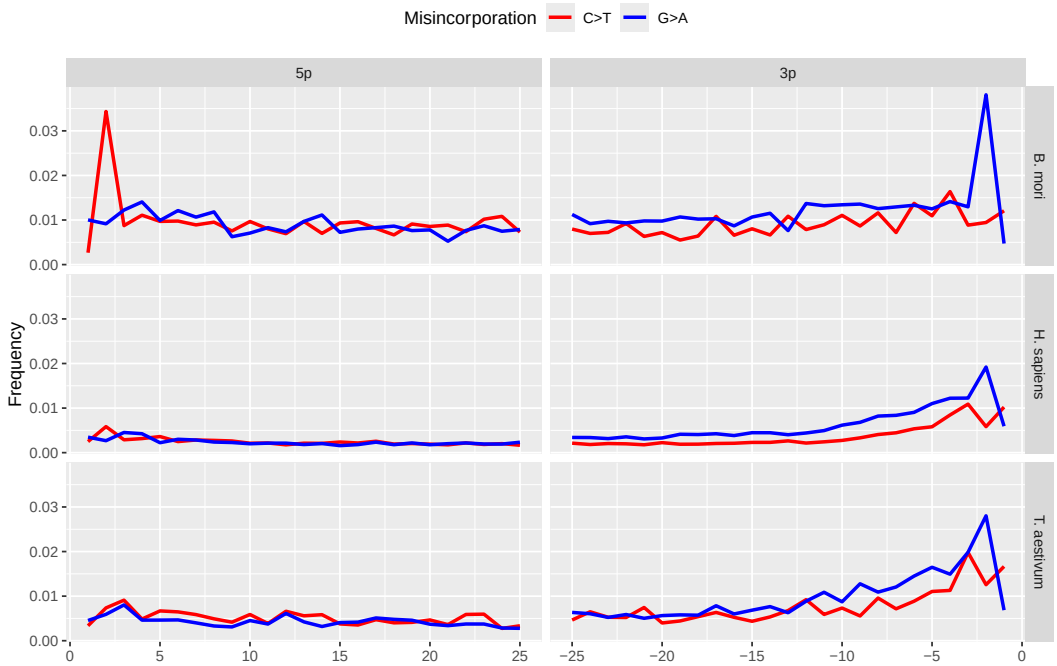

Figure 4. Damage plots of representative eukaryote species of the genera with at least 1% relative abundance. The genomes were used *Bombyx mori* (GCF_030269925.1), *Homo sapiens* (hs37d5), and *Triticum aestivum* (GCF_018294505.1).

was 0.81% (SD: 0.09), with a depth of 0.017 (SD: 0.002). Alignment to the RSRS sequence resulted in 349 reads with 82.02% coverage and 2.56× depth. Based on Haplocheck analysis, mtDNA did not contain any detectable contamination. According to the insufficient depth of mtDNA alignments, schmutzi did not give valuable results. ContamMix estimated mtDNA contamination was 0.79%. The top 5 Haplogrep3 hits of the consensus sequence obtained after deduplication are summarised in table 2. X-chromosome-based analysis estimated a contamination level of 0.007%. From the ratio of reads matching the X and Y chromosomes, an XY karyotype ($R_y = 0.099$, 95% CI: 0.093-0.104) was assigned. Aligning the read to Y chromosome, a coverage of 0.62% and an average depth of 0.01 was obtained, and haplotype was predicted as J1a2a1a1a∼ (path: A1b, BT, C1a2, CT) and E1a2b1a2 (path: BT, CT, E) by pathPhynder and yHaplo, respectively.

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

Haplogroup	Quality	Detected polys	Undetected polys
L3	0.863	146T 182C! 1018G 2758G 2885T 4104A 4312C 7146A 7256C 7521G 8468C 8655C 10688G 10810T 10915T 11914G 13506C 13650C 15301A 16129G 16187C 16189T 16230A 16278C 16311T	
L3c'd	0.863	146T 152T! 182C! 1018G 2758G 2885T 4104A 4312C 7146A 7256C 7521G 8468C 8655C 10688G 10810T 10915T 11914G 13105G! 13506C 13650C 15301A 16129G 16187C 16189T 16230A 16278C 16311T	
L3h	0.850	146T 182C! 1018G 2758G 2885T 4104A 4312C 7146A 7256C 7521G 8468C 8655C 10688G 10810T 10915T 11914G 13506C 13650C 15301A 16129G 16187C 16189T 16230A 16278C 16311T	7861C
L3b'f	0.847	146T 182C! 1018G 2758G 2885T 4104A 4312C 7146A 7256C 7521G 8468C 8655C 10688G 10810T 10915T 11914G 13506C 13650C 15301A 16129G 16187C 16189T 16230A 16278C 16311T	15944d
L3e'i'k'x	0.841	146T 182C! 1018G 2758G 2885T 4104A 4312C 7146A 7256C 7521G 8468C 8655C 10688G 10810T 10915T 11914G 13506C 13650C 15301A 16129G 16187C 16189T 16230A 16278C 16311T	150T 10819G

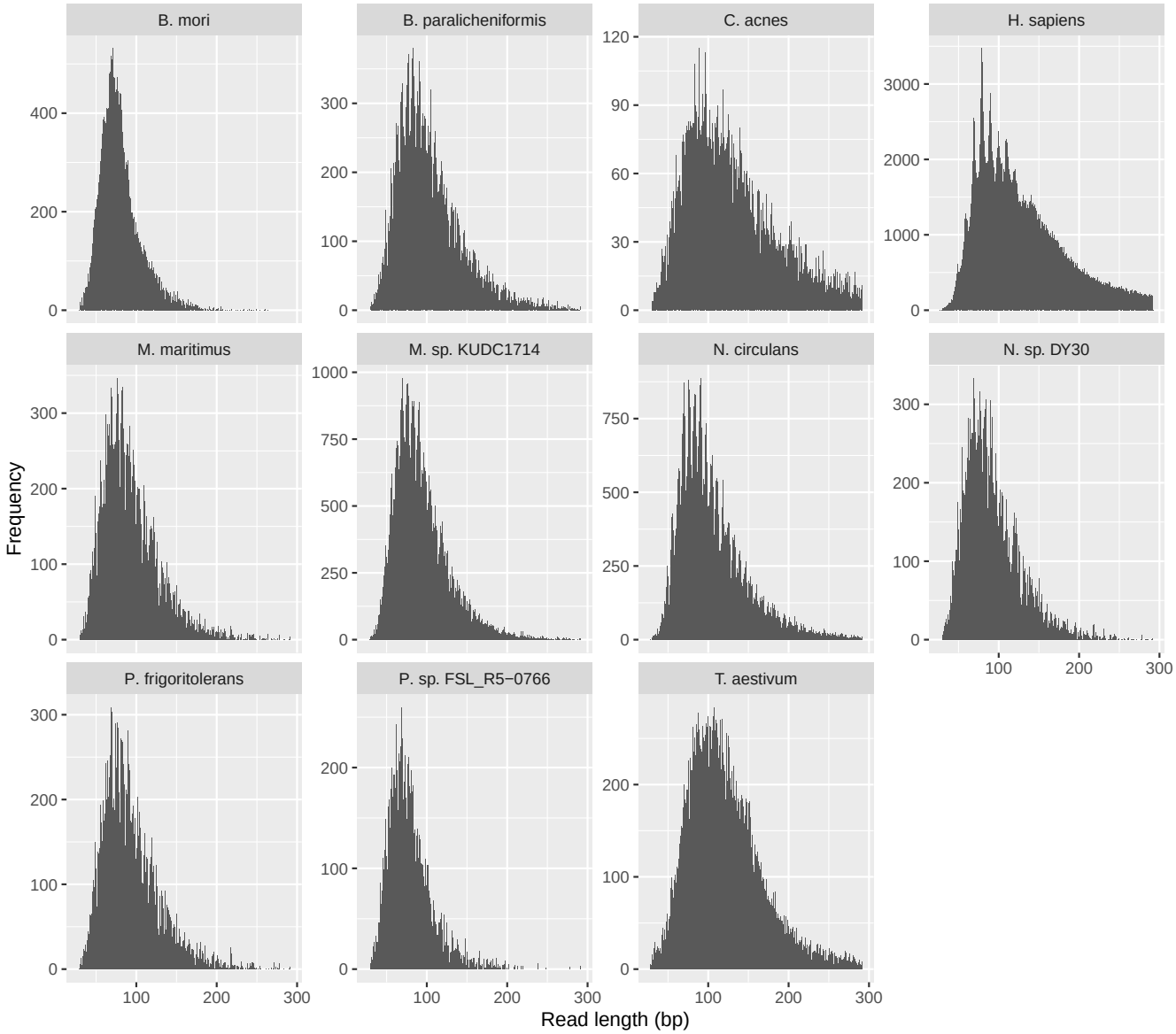

Figure 5. Length distributions of the reads aligned on selected species genomes.

Discussion

To the best of our knowledge, no results from the study of the cereal mummy metagenome have been reported to date. Therefore, it is impossible to compare our results with those of others. The key question in interpreting our study's results is whether our reads match certain genomes that are archaic or derived from modern contamination. Due to the unique nature of the sample, DNA extraction has more pitfalls and significantly lower yields than even the usual archaic samples. The size of the contigs generated from our reads is short, and their number is small compared to the number of reads. The degradation plots based on the reads realigned onto the contigs of Sample 2 show, although not perfectly, patterns that might support a historical origin. An argument favoring this could be the lack of decay signatures in Sample 1. The trend of the damage plot shows that the G>A transformation increases towards the 3p end of the sequences. At the 5p end, although not as clearly, the C>T transformation also shows an increase. The low number of reads aligned to the contigs (Sample 2: 101,330 reads) may explain the noisy nature of the curves. The increase in G>A at the 5p end and C>T at the 3p end is similar to Skoglund et al.[39], which is not USER-treated. The abrupt decrease in the increasing frequency of C>T at the 5p end and G>A at the 3p end in the last positions resembles the pattern resulting from the internal barcode ligation bias.[40] The pattern is similar when reads were aligned to selected genomes. It is reasonable to assume that the corn mummy contains one or more Poales species, and it is very unlikely that the sample is contaminated with such fragmented modern genomes of such species. Furthermore, phylogenetic analysis

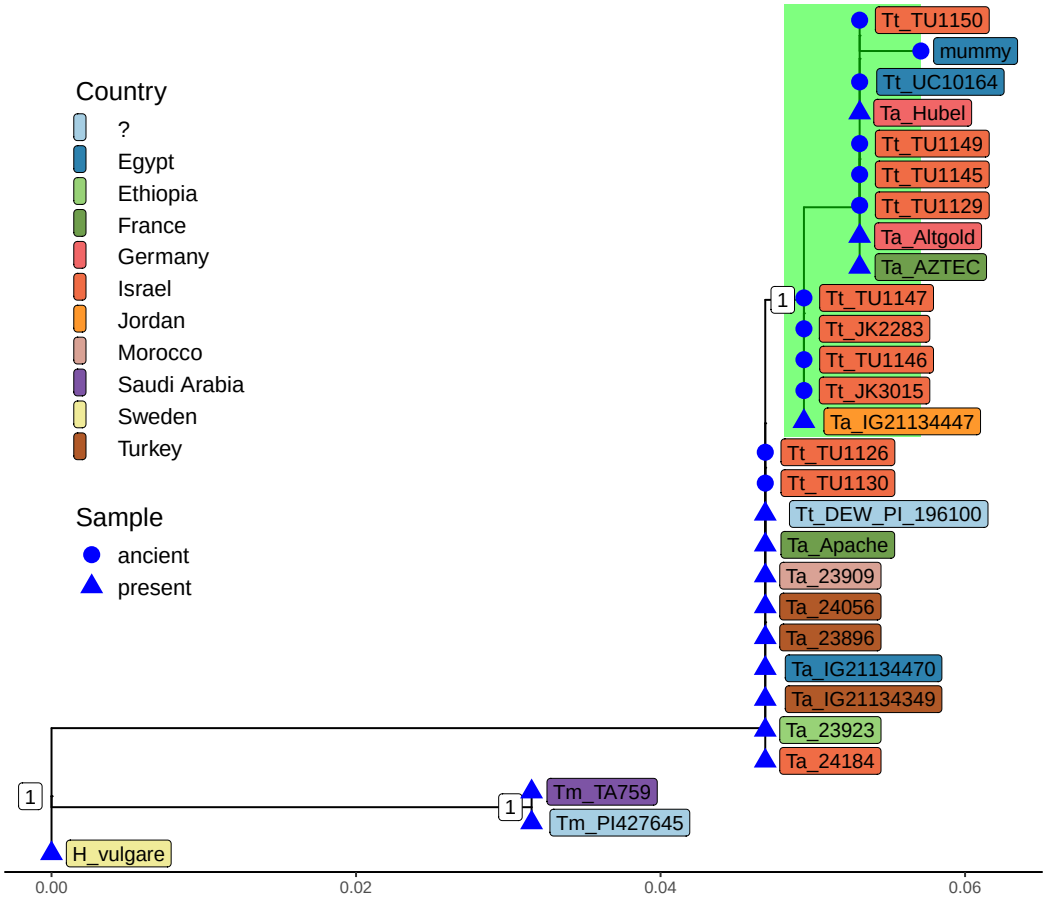

Figure 6. Gene-tree of *Triticum* and *Hordeum* samples. In the labels, the H_vulgare is *H. vulgare* (outer group), Ta is *T. aestivum*, Tm is *T. monococcum*, Tt is *T. turgidum* ssp. *dicoccum*, mummy is the longest wheat origin contig (PV405787) from the corn mummy sample. The green band indicates 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 1.

indicates that the sequence from our sample belongs to a clade in which archaic sequences represent 70%, and contains 80% of all historical *Triticum* samples. Based on these results, the degradation plot pattern of *T. aestivum* may provide a guide for interpreting the age of the genome fragments of other species. If this is accepted, the archaicity of the *H. sapiens* sequences can be reasonably assumed, as in the case of *B. mori*. The same can be said for the bacterial species, except *C. acnes*, which is more likely to be a modern contaminant.

For the bacteria found in Sample 2, for which DNA damage could be tested and showed a historical pattern of *B. paralicheniformis*,[41] *Metabacillus* sp. KUDC1714,[42] *N. circulans*,[43] their occurrence in the soil is supported by the literature. Although no literature data is available on the occurrence of *Neobacillus* sp. DY30, the origin of the genome we used is given as rhizosphere soil in the NCBI database. The presence of *P. frigoritolerans* in the rhizosphere has a beneficial effect on plant growth,[44], which has been explicitly shown for *T. aestivum*.[45] Although no data are available on *Paenibacillus* sp. FSL_R5-0766 in soil, it is generally acknowledged that numerous *Paenibacillus* species benefit plant growth in soil.[46] The only bacteria for which no literature data were found on the occurrence in the soil, rhizosphere was *M. maritimus*. This species is characteristically found in marine sediment[47] but other species of the genus occur in sandy soil.[48]

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[49] commonly used for *Triticum* phylogenetic analysis, based on the longest contig (PV405787), our sample most closely resembles ancient Egyptian and southern Levantine samples.

The interpretation of the *Bombyx* results might be linked to the *Niallia* findings, as *N. circulans* is known to be present in the

gut of *B. mori*, among other environments.[50] 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 the corn mummy. The other one is that the larva was placed on the open surface of the corn mummy before 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 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),[51] 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*. The NCBI NT database contains complete genomes from these species 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 became 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 Bombycidae species, lived in Africa.[52]

The haplogrouping of the human sequences found in Sample 2 suggests that both the mtDNA and the Y chromosome might be of African origin. These haplogroups may have been present in both modern and ancient Egypt. Within the L3 mitochondrial lineage, we found that the L3b, L3d, L3e, and L3f subclades are the most common in Central and West Africa.[53]. L3c is very rare in Soares et al., who report one sample from East Africa and one from the Middle East.[53] Rishishwar and King[54], 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.[55], 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 eight 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.[55] While the E-haplogroup identified by yHaploval indicates primarily African origin, the type J1a2a1a1a identified by pathPhynder overlaps only partially with this region.[56] On the other hand, the J1a2a1a haplotype is also found in Egypt, in addition to a higher frequency in the Arabian Peninsula, southern Levant, and southern Iraq.[56] Keita's review of the diversity of Y chromosome variants supports our findings.[57]

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 a few millimeters, it is difficult to imagine that human DNA would be present due to human touch. But, it is possible that some human DNA present in any human DNA-containing fluid may have infiltrated the mummy and been found. For example, this could have been the case if the crypt in which the mummy was kept had been contaminated with human sewage, which could have soaked the mummy. As this case 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 chance 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 shed epithelial cells 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.[58] 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. While certain bacterial sequences appear to be of modern origin, sequences from an ancient wheat cultivar, a silkmoth-related insect with its gut bacteria, and an African 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. We are also grateful to Shomarka Omar Yahya Keita for his help clarifying our wording and putting the results into context.

Author contribution statement

NS takes responsibility for the data's integrity and the data analysis's accuracy. DM, IC, and NS conceived the concept of the study. AGT, FJK, KK, and SÁN got the samples from the corn mummy. BP and GM did the DNA extraction and NGS sequencing. NS performed the bioinformatic analysis and visualization. AGT, DM, NS, and SÁN participated in the drafting of the manuscript. AGT, BP, DM, FJK, GM, IC, KK, NS, and SÁN critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

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 MILAB Artificial Intelligence National Laboratory framework.

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

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

17. 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).

18. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104 (2017).

19. 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).

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

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

22. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).

23. 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).

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

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

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

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

28. 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).

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

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

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

32. 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).

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

34. 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).

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

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

39. Skoglund, P., Ersmark, E., Palkopoulou, E. & Dalén, L. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Curr. Biol.* **25**, 1515–1519 (2015).

40. The SPAAM Community. *The Little Book of Smiley Plots* (2025). https://www.spaam-community.org/little-book-of-smiley-plots.

41. Sun, Y. *et al.* Bacillus paralicheniformis SYN-191 isolated from ginger rhizosphere soil and its growth-promoting effects in ginger farming. *BMC Microbiol.* **25**, 1–14 (2025).

42. 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).

43. 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).

44. Świątczak, J., Kalwasińska, A. & Brzezinska, M. S. Plant growth–promoting rhizobacteria: Peribacillus frigoritolerans 2RO30 and Pseudomonas Sivasensis 2RO45 for their effect on canola growth under controlled as well as natural conditions. *Front. Plant Sci.* **14**, 1233237 (2024).

45. Tara, N., Saharan, B. S. *et al.* Plant growth promoting traits shown by bacteria Brevibacterium frigrotolerans SMA23 isolated from Aloe vera rhizosphere. *Agric. Sci. Dig. Res. J.* **37**, 226–231 (2017).

46. Patowary, R. & Deka, H. Paenibacillus. In *Beneficial Microbes in Agro-Ecology*, 339–361 (Elsevier, 2020).

47. Pal, D. *et al.* Bacillus maritimus sp. nov., a novel member of the genus Bacillus isolated from marine sediment. *Int. J. Syst. Evol. Microbiol.* **67**, 60–66 (2017).

48. Zhang, G. *et al.* Mesobacillus harenae sp. nov., isolated from the sandy soil of a cold desert. *Int. J. Syst. Evol. Microbiol.* **71**, 004594 (2021).

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

50. 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).

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

52. Lin, R.-J., Lin, Y.-C., Braby, M. F., Zwick, A. & Hsu, Y.-F. Phylogenetic relationships and historical biogeography of silkmoths (lepidoptera: Bombycidae) suggest an origin in southern gondwana. *Mol. Phylogenetics Evol.* **200**, 108176 (2024).

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

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

55. 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).

56. Sahakyan, H. *et al.* Origin and diffusion of human Y chromosome haplogroup J1-M267. *Sci. Reports* **11**, 6659 (2021).

57. Keita, S. O. Y. History in the interpretation of the pattern of p49a, f TaqI RFLP Y-chromosome variation in Egypt: A consideration of multiple lines of evidence. *Am. J. Hum. Biol.* **17**, 559–567 (2005).

58. 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).