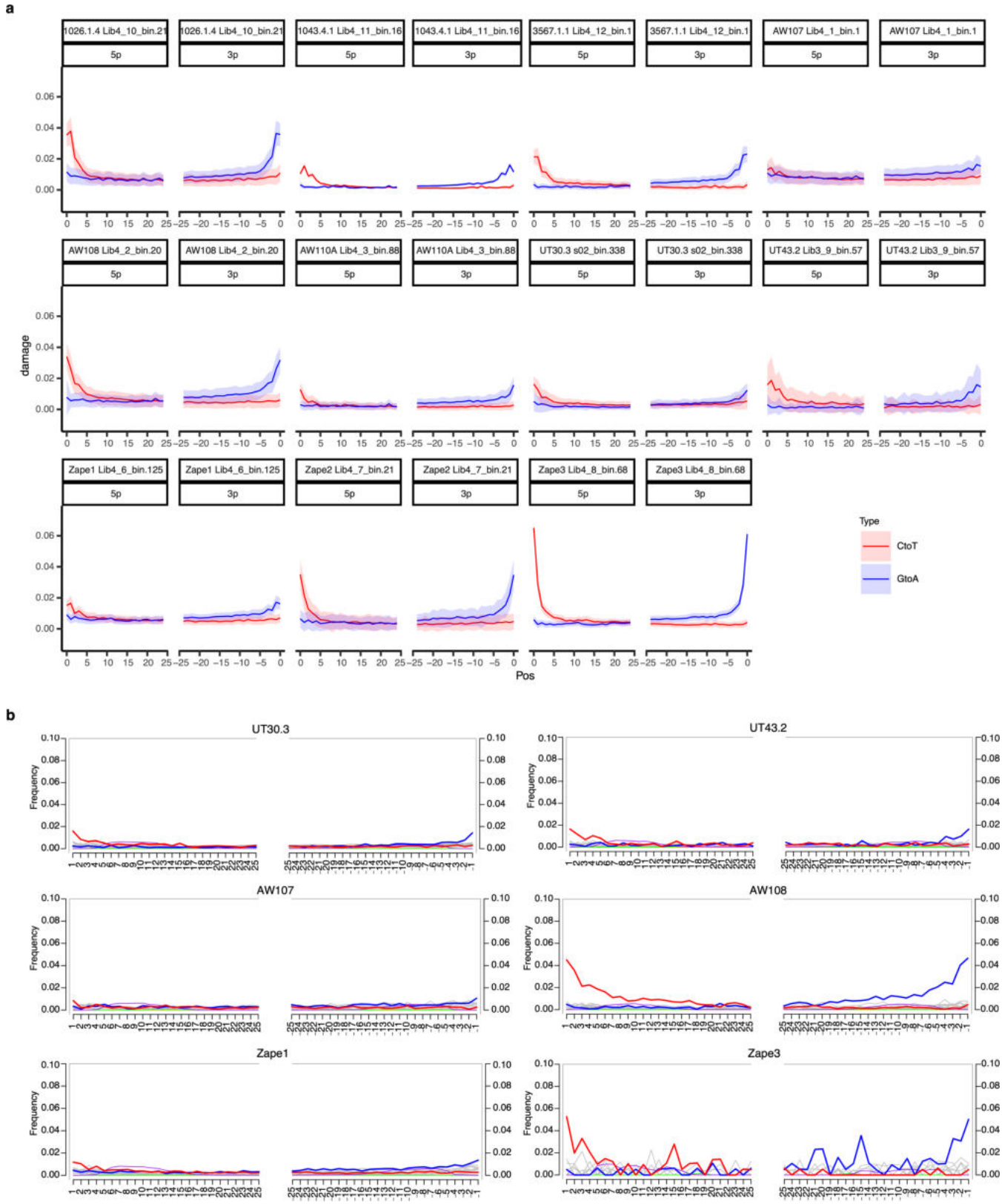
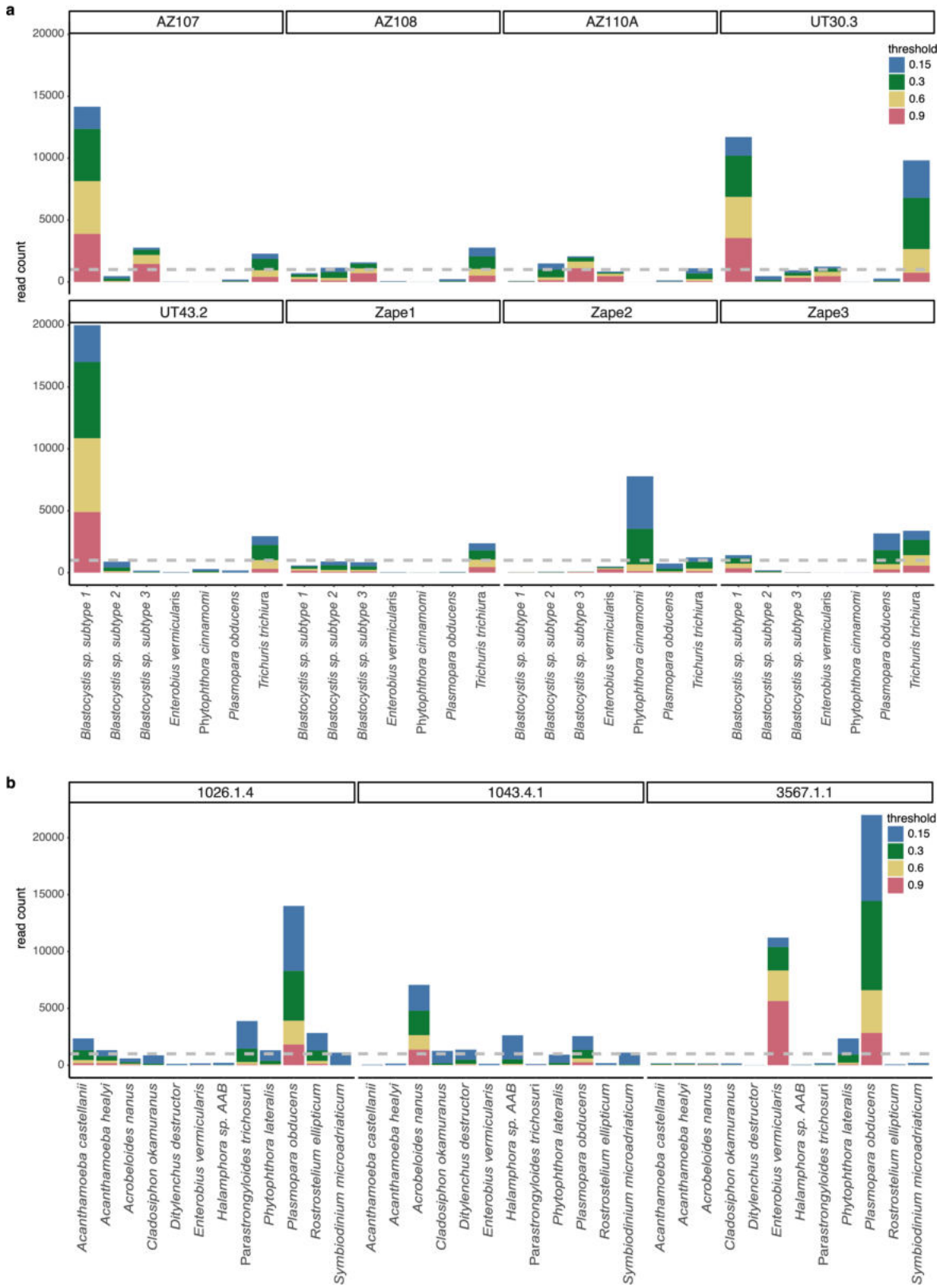
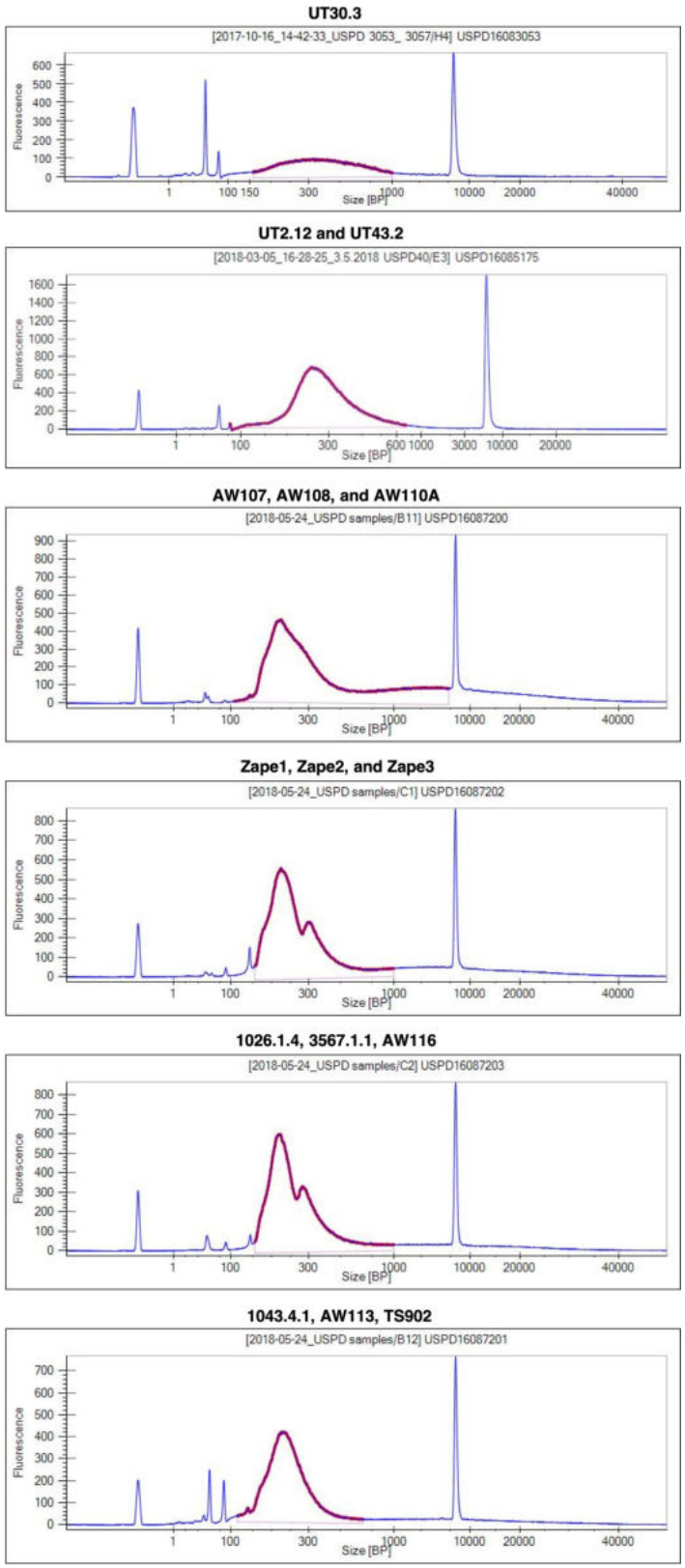



**Extended Data Fig. 2** | See next page for caption.

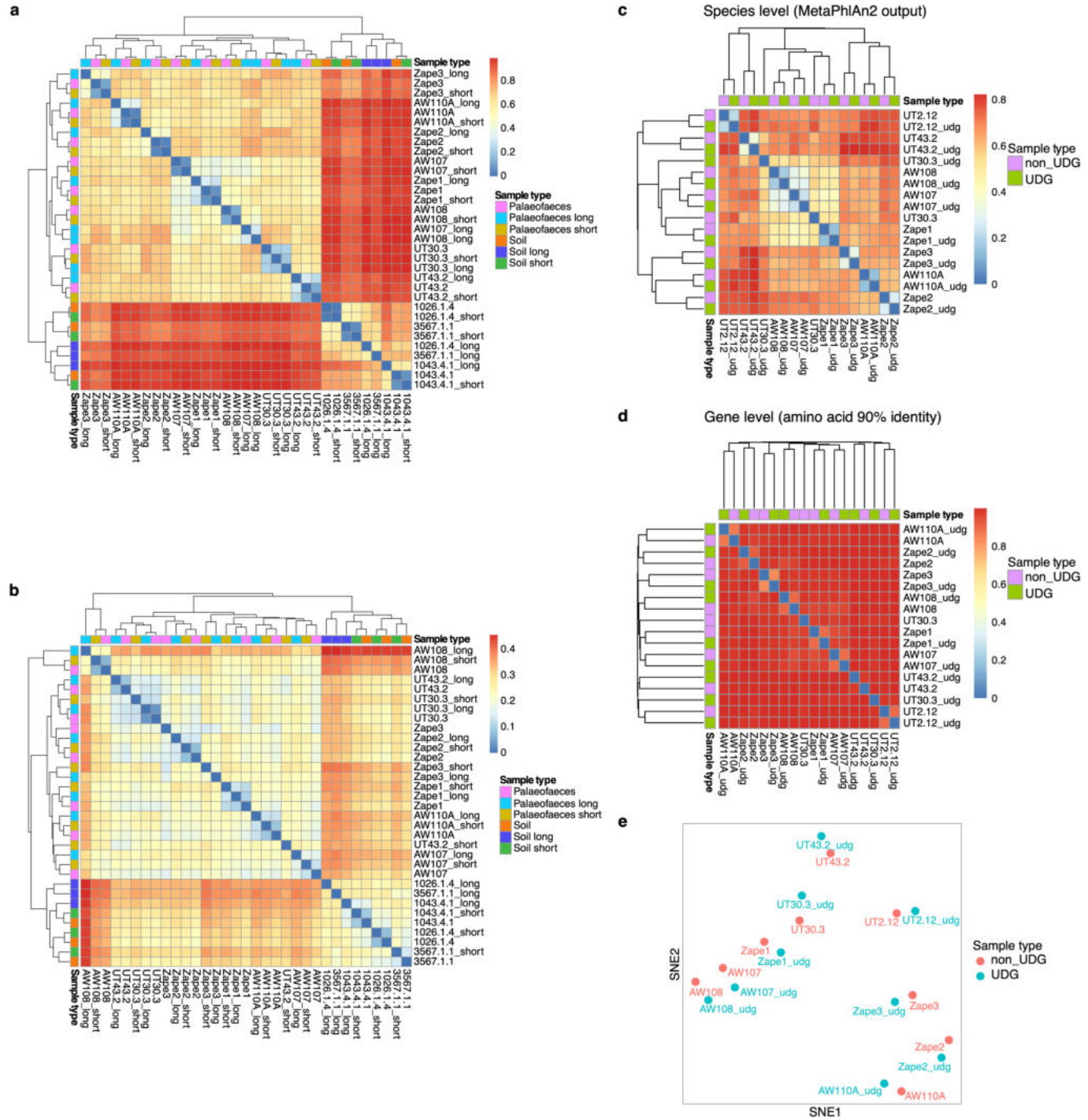
**Extended Data Fig. 2 | Microbial DNA and mtDNA damage patterns.** **a**, Microbial damage patterns of the palaeofaeces and the Boomerang soil samples as identified by DamageProfiler[88]. A medium-quality or high-quality reconstructed genome was used as reference for its respective sample. All MAGs used as reference genomes for the palaeofaeces are of known gut microbial species. The red line indicates the average frequency of C-to-T substitutions across all contigs per bin and the blue line indicates the average frequency of G-to-A substitutions across all contigs per bin. The shaded areas show the s.d. (1026.1.4 Lib4_10_bin.21, $n = 488$ contigs; 1043.4.1 Lib4_11_bin.16, $n = 133$ contigs; 3567.1.1 Lib4_12_bin.1, $n = 278$ contigs; AW107 Lib4_1_bin.1, $n = 208$ contigs; AW108 Lib4_2_bin.20, $n = 337$ contigs; AW110A Lib4_3_bin.88, $n = 210$ contigs; UT30.3 s02_bin.338, $n = 74$ contigs; UT43.2 Lib3_9_bin.57, $n = 174$ contigs; Zape1 Lib4_6_bin.125, $n = 212$ contigs; Zape2 Lib4_7_bin.21, $n = 241$ contigs; Zape3 Lib4_8_bin.68, $n = 324$ contigs). Contigs with fewer than 1,000 reads aligned were removed from the analysis. **b**, mtDNA damage patterns of the palaeofaeces as identified by mapDamage2.0[87]. Human mtDNA (rCRS) was used as reference. The red line indicates the average frequency of C-to-T substitutions and the blue line indicates the average frequency of G-to-A substitutions. Samples AW110A and Zape2 did not have enough mtDNA reads for mtDNA damage assessment.




**Extended Data Fig. 3 | Parasites in the palaeofaeces and the soil samples classified using Kraken 2.** The bars represent the reads assigned with a Kraken[74] confidence threshold between 0.15 and 0.9. The value specifies the fraction of *k*-mers needed for the specific classification level. The grey dotted line indicates the 1,000 reads cut-off. The displayed parasites were detected above the cut-off in at least one sample. **a**, Parasites in the palaeofaeces. In six out of eight palaeofaeces samples, *Blastocystis* is above the cut-off. Subtype 1 is the dominant subtype in samples AW107, UT30.3, UT43.2 and Zape3, whereas subtype 3 is the dominant subtype in AW108 and AW110A. Other parasites do not meet the cut-off requirements described in the Methods. **b**, Parasites in the soil samples include *Acanthamoeba* (a parasite frequently found in soil) in sample 1026.1.4 and *Enterobius vermicularis* (human pinworm) in sample 3567.1.1.




**Extended Data Fig. 4 | BioAnalyzer results showing the length distribution of DNA fragments per library.** The libraries contain 120-bp adapters.




**Extended Data Fig. 5 | Species and gene content of long versus short DNA fragments and UDG-treated versus non-UDG-treated samples. a**, **b**, Pairwise comparison between whole samples, only subsets containing short reads and only subsets with long reads. **a**, Heat map of species-level pairwise Jaccard distances for whole samples, short-read subsets (reads ≤ 145 bp) and long-read subsets (reads > 145 bp). Species were identified by MetaPhlAn2[20]. The groups cluster together by sample. **b**, Heat map of gene-level pairwise Jaccard distances for whole samples, short-read subsets and long-read subsets. Genes were identified by PROKKA[38] and a count matrix was built from PROKKA output files. Groups from the same sample cluster together. **c**–**e**, Species and gene content comparison between UDG-treated libraries and non-UDG-treated libraries (Supplementary Information section 12 and Supplementary Table 10). **c**, Heat map of species-level pairwise Jaccard distances between each pair of all UDG-treated and non-UDG-treated samples. Species were identified by MetaPhlAn2[20]. Each UDG-treated library clusters with non-UDG-treated library from the same sample. **d**, Heat map of gene-level pairwise Jaccard distances between each pair of all UDG-treated and non-UDG-treated samples. Genes were identified by PROKKA[38] and non-redundant gene catalogues were generated by collapsing genes within 10% amino acid identity distance. Each UDG-treated library clusters with non-UDG-treated library from the same sample. **e**, *t*-Distributed stochastic neighbour embedding (*t*-SNE) analysis at the species level shows clustering of each UDG-treated library with the non-UDG-treated library from the same sample.