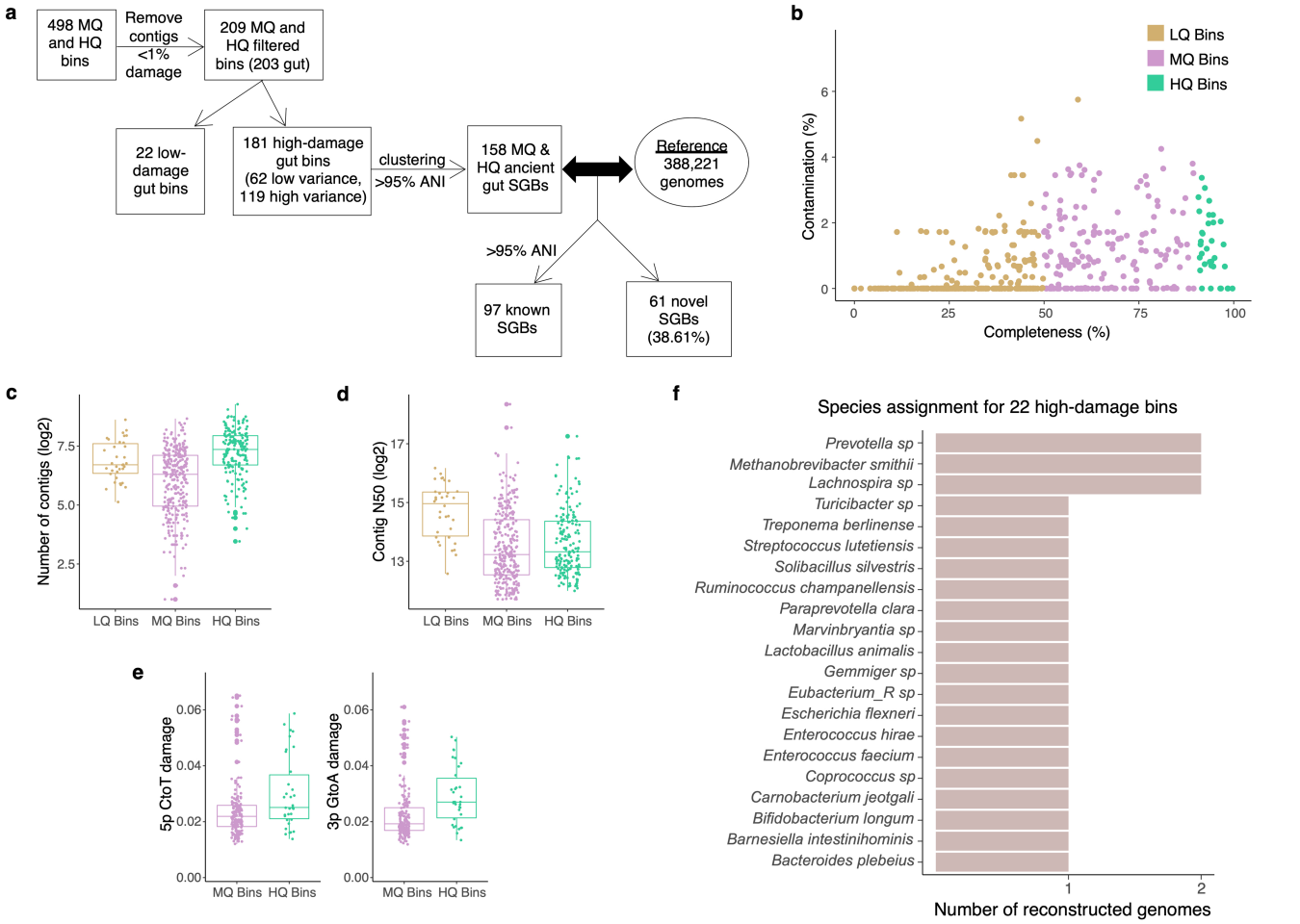
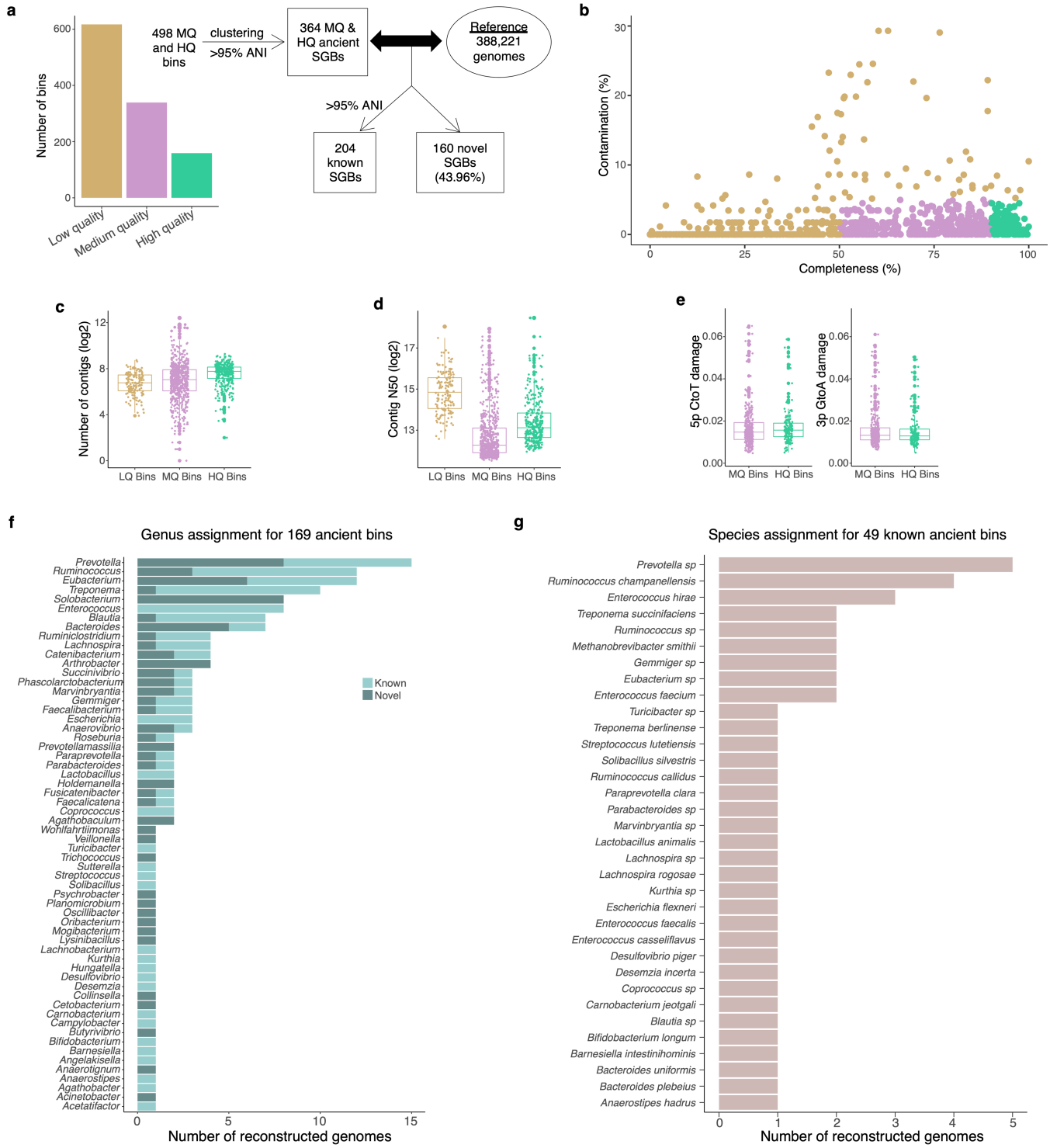
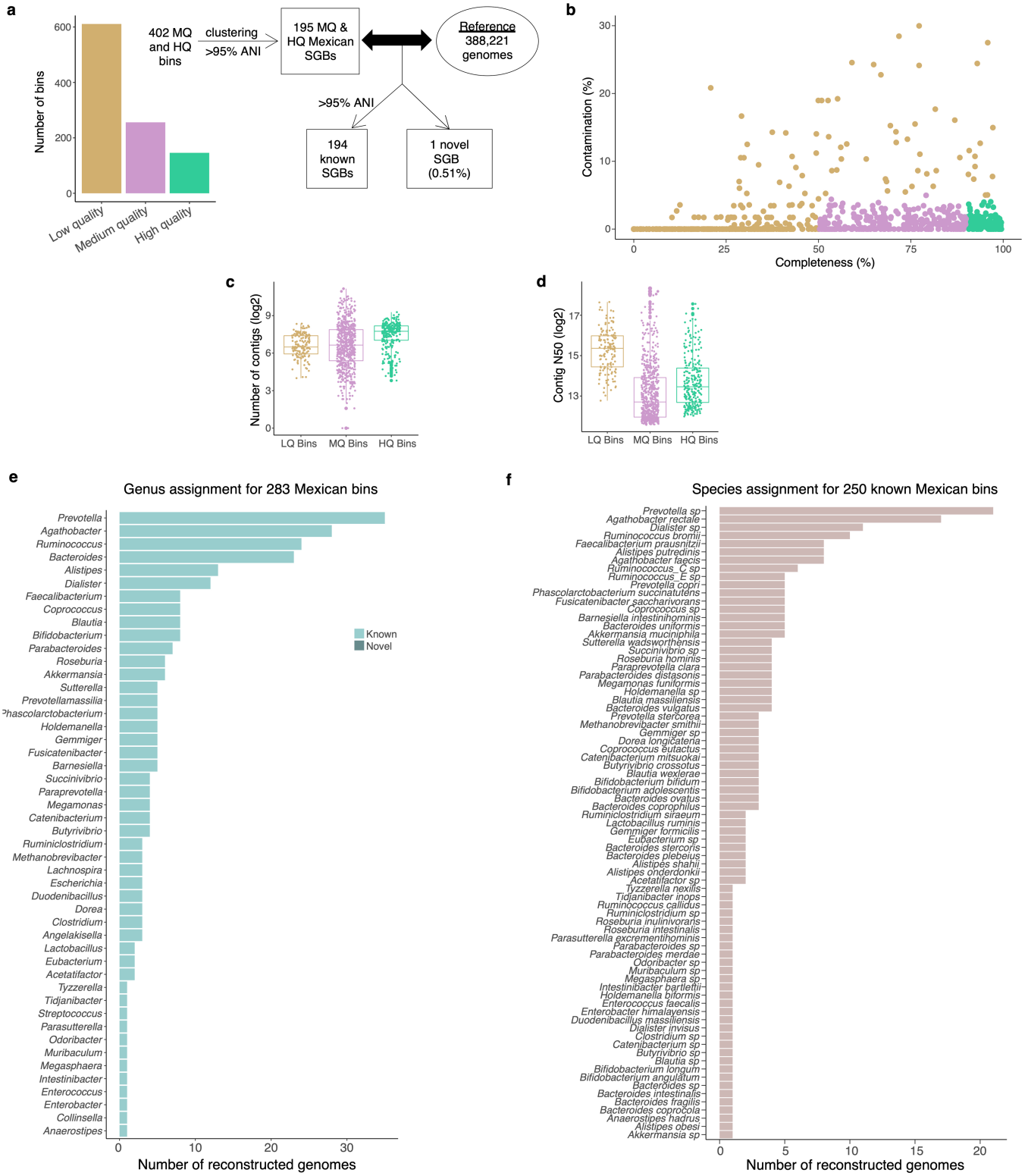


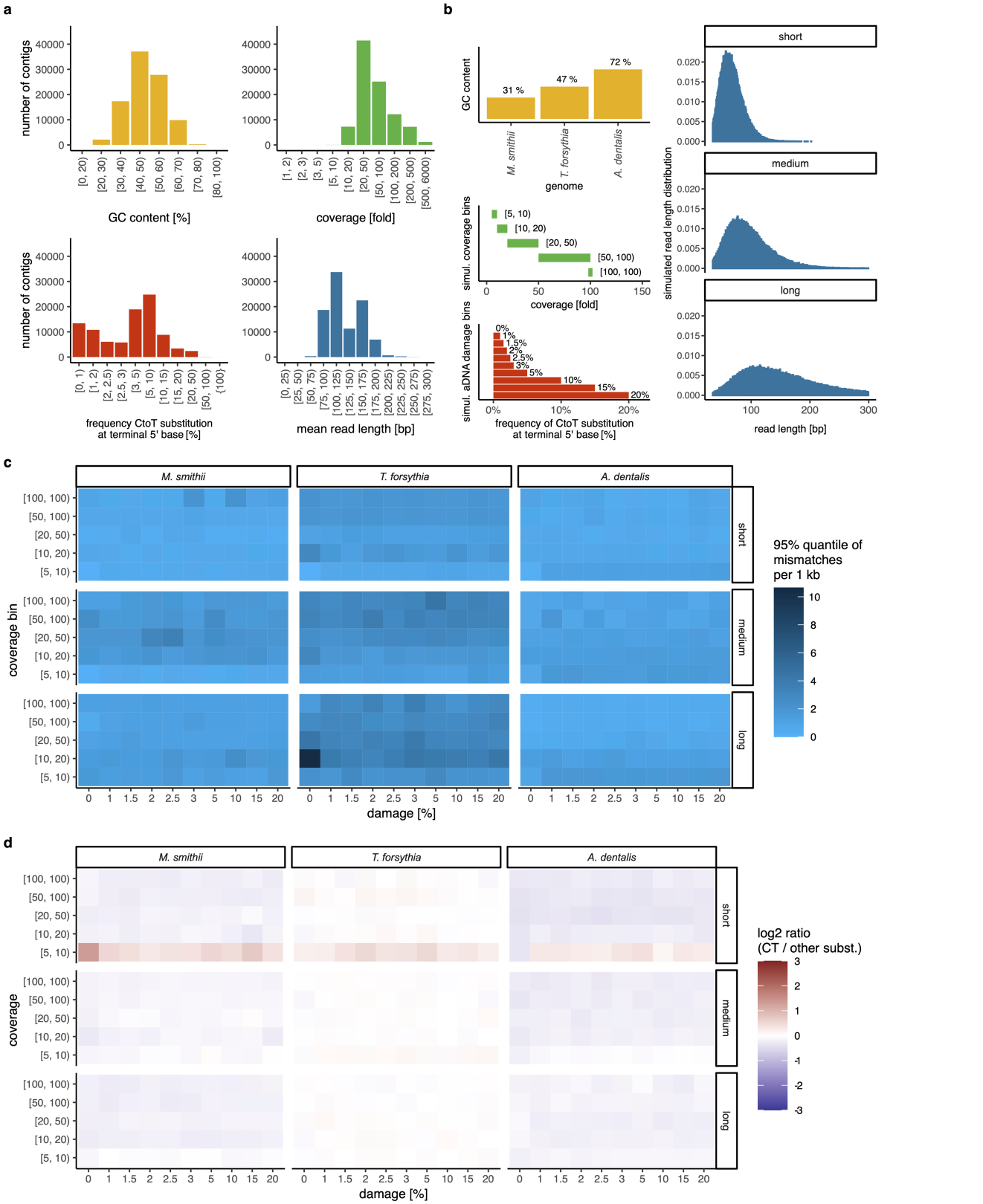
Extended Data Fig. 6 | De novo genome reconstruction from palaeofaeces recovers 181 authenticated ancient gut microbial genomes, 39% of which are novel SGBs. Related to Fig. 2. **a**–**d**, CheckM[79] quality estimation for de novo reconstructed microbial genomes for the 209 filtered bins (low-quality bins, $n = 285$; medium-quality bins, $n = 175$; high-quality bins, $n = 34$). Genomes were classified as low quality (LQ; completeness ≤ 50% or contamination > 5%), medium quality (MQ; 90% ≥ completeness > 50%, contamination < 5%) or high quality (HQ; completeness > 90% and contamination < 5%). **a**, Filtering steps, number of bins that belong to each of the quality categories and classification of novel SGBs. **b**, Contamination and completeness distribution for the filtered bins. **c**, Distribution of the number of contigs for each of the quality categories. **d**, Distribution of contig N50 values for each of the quality categories. **e**, Damage levels, specifically C-to-T substitutions at the 5′ end and G-to-A substitutions at the 3′ end of the reads, for each ancient bin as estimated by DamageProfiler[88] (medium-quality bins, $n = 175$; high-quality bins, $n = 34$). **f**, GTDB-Tk[23] species assignment for the known species. In **c**–**e**, data are presented as box plots (middle line, median; lower hinge, first quartile; upper hinge, third quartile; upper whisker extends from the hinge to the largest value no further than 1.5× the interquartile range from the hinge; lower whisker extends from the hinge to the smallest value at most 1.5× the interquartile range from the hinge; data beyond the end of the whiskers are individually plotted outlying points).



Extended Data Fig. 7 | De novo genome reconstruction from palaeofaeces recovers 498 medium- and high-quality microbial genomes, 44% of which are novel SGBs. Related to Fig. 2. **a**–**d**, CheckM[79] quality estimation of all 498 de novo reconstructed microbial genomes (low-quality bins, $n = 617$; medium-quality bins, $n = 339$; high-quality bins, $n = 159$). Genomes were classified as low quality (completeness ≤ 50% or contamination > 5%), medium quality (90% ≥ completeness > 50% and contamination < 5%) or high quality (completeness > 90% and contamination < 5%). **a**, Number of bins that belong to each of the quality categories and classification of novel SGBs. **b**, Contamination and completeness distribution for the reconstructed genomes. **c**, Distribution of the number of contigs for each of the quality categories. **d**, Distribution of contig N50 values for each of the quality categories. **e**, Damage levels, specifically C-to-T substitutions at the 5′ end and G-to-A substitutions at the 3′ end of the reads, for each bin as estimated by DamageProfiler[88] (medium-quality bins, $n = 339$; high-quality bins, $n = 159$). **f**, GTDB-Tk[23] genus estimation for members of both the novel and known SGBs. **g**, GTDB-Tk[23] species assignment for members of the known SGBs. In **c**–**e**, data are presented as box plots (middle line, median; lower hinge, first quartile; upper hinge, third quartile; upper whisker extends from the hinge to the largest value no further than 1.5× the interquartile range from the hinge; lower whisker extends from the hinge to the smallest value at most 1.5× the interquartile range from the hinge; data beyond the end of the whiskers are individually plotted outlying points).



Extended Data Fig. 8 | De novo genome reconstruction from present-day individuals of Mexican ancestry recovers 402 medium- and high-quality genomes, only 1 of which is a novel SGB. Related to Fig. 2. **a**–**d**, CheckM[79] quality estimation of all de novo reconstructed microbial genomes (low-quality bins, $n = 611$; medium-quality bins, $n = 256$; high-quality bins, $n = 146$). Genomes were classified as low quality (completeness ≤ 50% or contamination > 5%), medium quality (90% ≥ completeness > 50% and contamination < 5%) or high quality (completeness > 90% and contamination < 5%). **a**, The number of bins that belong to each of the quality categories and classification of novel SGBs. **b**, Contamination and completeness distribution for the reconstructed genomes. **c**, Distribution of the number of contigs for each of the quality categories. **d**, Distribution of contig N50 values for each of the quality categories. **e**, GTDB-Tk[23] genus estimation for members of both the novel and the known Mexican SGBs. **f**, GTDB-Tk[23] species assignment for members of the known Mexican SGBs. In **c**, **d**, data are presented as box plots (middle line, median; lower hinge, first quartile; upper hinge, third quartile; upper whisker extends from the hinge to the largest value no further than 1.5× the interquartile range from the hinge; lower whisker extends from the hinge to the smallest value at most 1.5× the interquartile range from the hinge; data beyond the end of the whiskers are individually plotted outlying points).



Extended Data Fig. 9 | Effect of aDNA damage on the assembly of short-read data. Related to Fig. 2, see Supplementary Information section 6. **a**, Distribution of the values of four sequencing data variables that may have an effect on the assembly of short-read data and were observed in the 498 medium-quality and high-quality MAGs assembled in this study. **b**, Overview of the parameter space of the variables GC content, sequencing depth, observed aDNA damage and read length that was used for simulating short-read sequencing using gargammel[107]. **c**, Number of mismatches per 1 kb of alignable contig sequence with respect to the reference genome as observed at the 95% quantile for all combinations of reference genome, read length distribution, simulated aDNA damage and coverage averaged across the five replicates. **d**, The $\log_2$-transformed ratio of C-to-T substitutions to the average number of all other substitutions per 1 kb of alignable contig sequence for all combinations of reference genome, read length distribution, simulated aDNA damage and coverage averaged across the five replicates. Positive values indicate an excess of C-to-T substitutions.