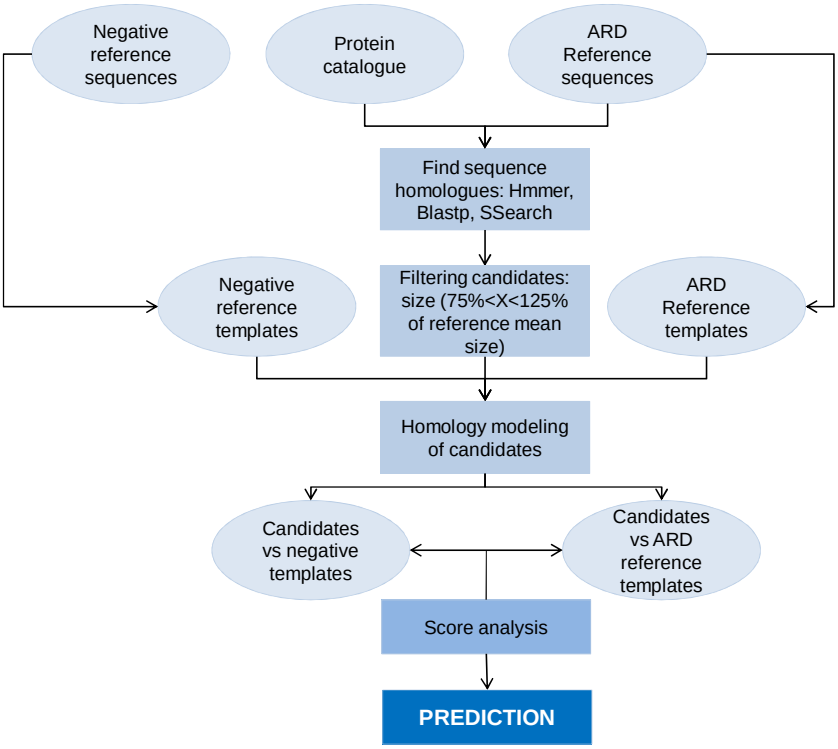
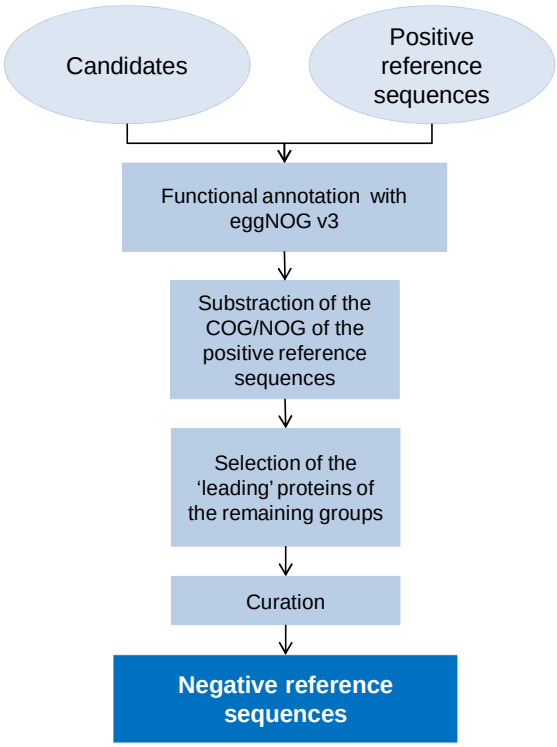
Prediction of the intestinal resistome by a novel 3D-based method
Supplementary material

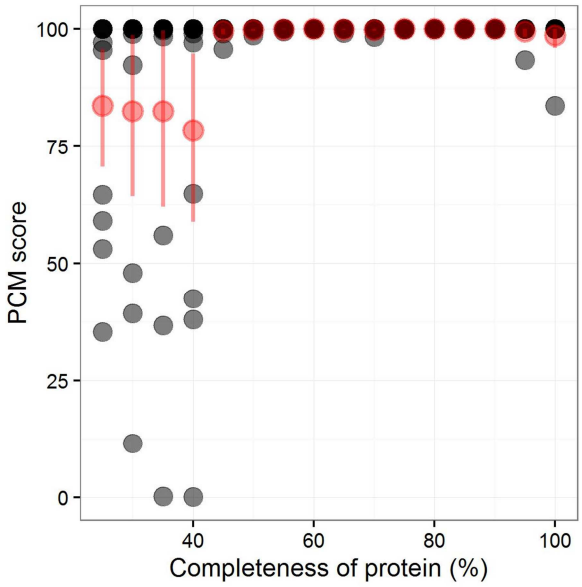
Extended Data Fig. 1: Illustration of the concept of pairwise comparative modelling. ARD: antibiotic resistance determinant.



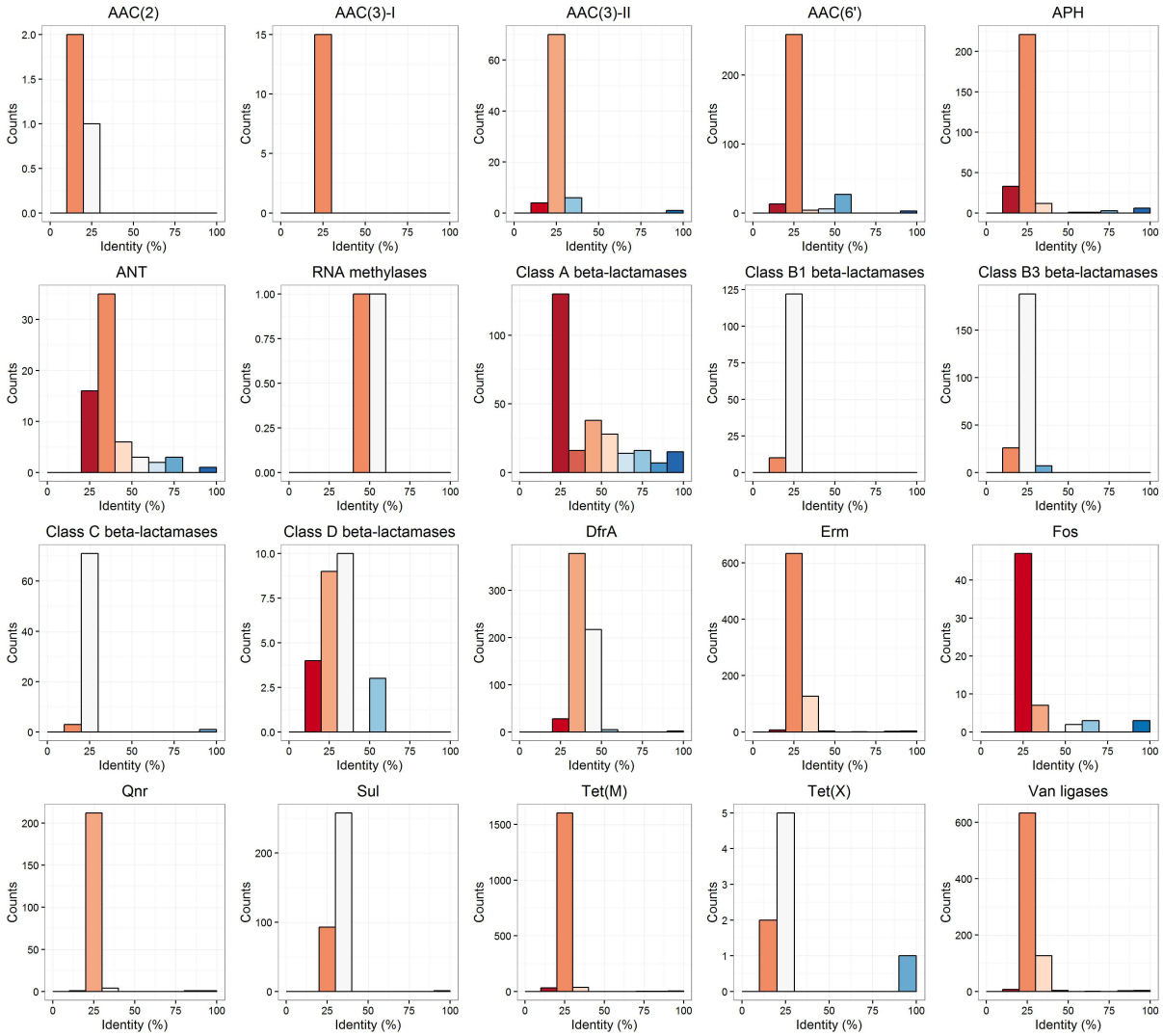
Extended Data Fig. 2: Illustration of the process used for the choice of negative templates.



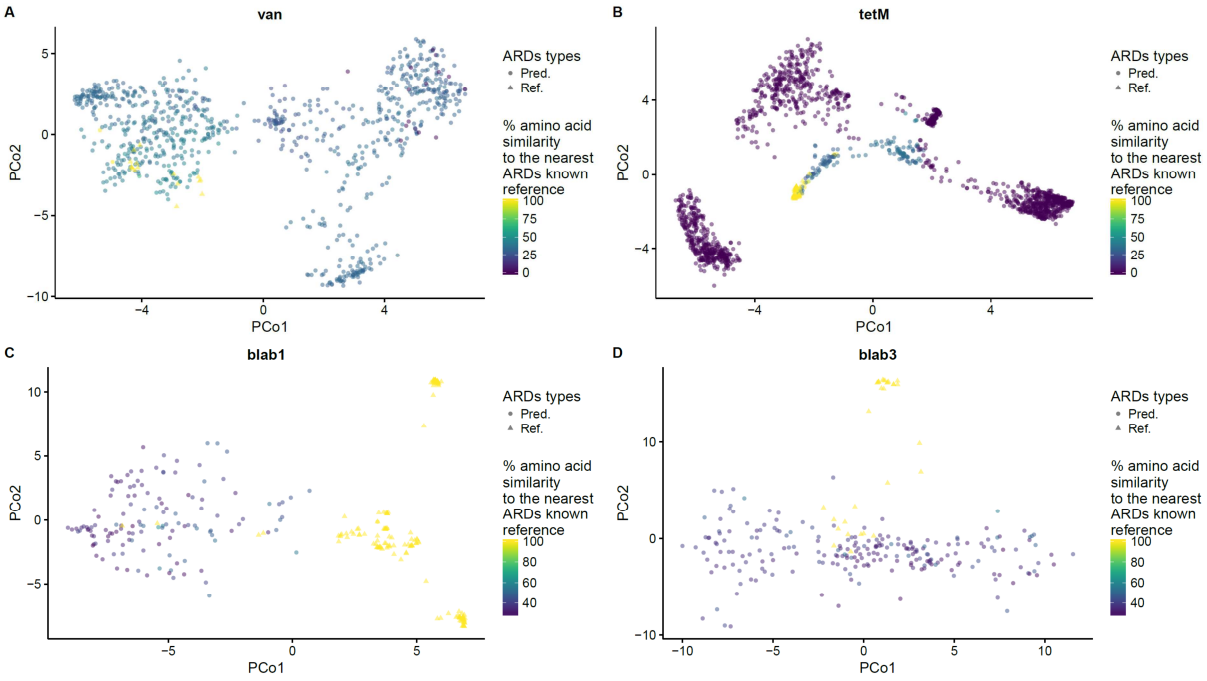
Extended Data Fig. 3: performance of PCM with incomplete genes/proteins. Each grey dot represents a PCM experiment. Red dots depict the mean and 95% confidence interval for each completeness values of the protein. Briefly, the 3.9M gene catalogue harbours 41.4% of genes that are predicted to be incomplete either on the 5', the 3' or both extremities[1]. We tested to which extent the prediction of chopped ARDs could remain correct. We selected 12 reference class A beta-lactamases (BlaZ, CblA-1, CepA-29, CfxA2, CfxA6, CTX-M-8, KPC-10, OXY-1, PER-1, SHV-100, TEM-101 and VEB-1) for which we iteratively removed 5% of the amino acid sequence at both edges in order to obtain 16 chopped candidates (from 100% to 25%) per reference ARD.



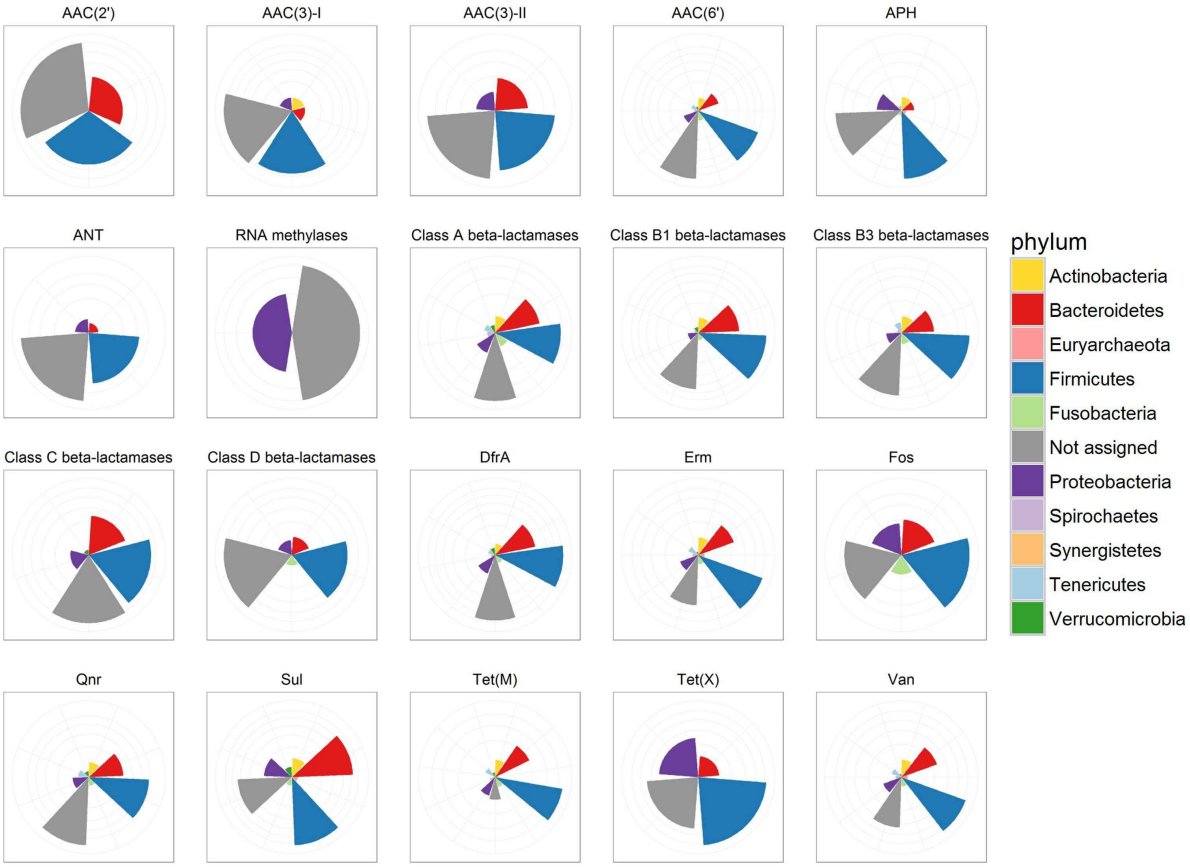
Extended Data Fig. 4: Number of predictions of antibiotic resistance determinants (ARDs) from a 3.9M gene catalogue of the intestinal microbiota[1] using PCM, in connection with the maximal identity observed with a reference ARD. The colours depict the percentage of identity (from red [low identity] to blue [high identity]).



Extended Data Fig. 5: Principal component analysis built from the amino acid similarity matrix between predicted and reference antibiotic resistance determinants. Dots and triangles account for predicted and references ARDs, respectively. The colour gradient accounts for the amino-acid identity between a predicted ARDs and the nearest known reference ARDs. (A) Van ligases (van); (B) Tet(M) (tetM) (C) Class B1 beta-lactamases (blab1); (D) Class B3 beta-lactamases (blab3).

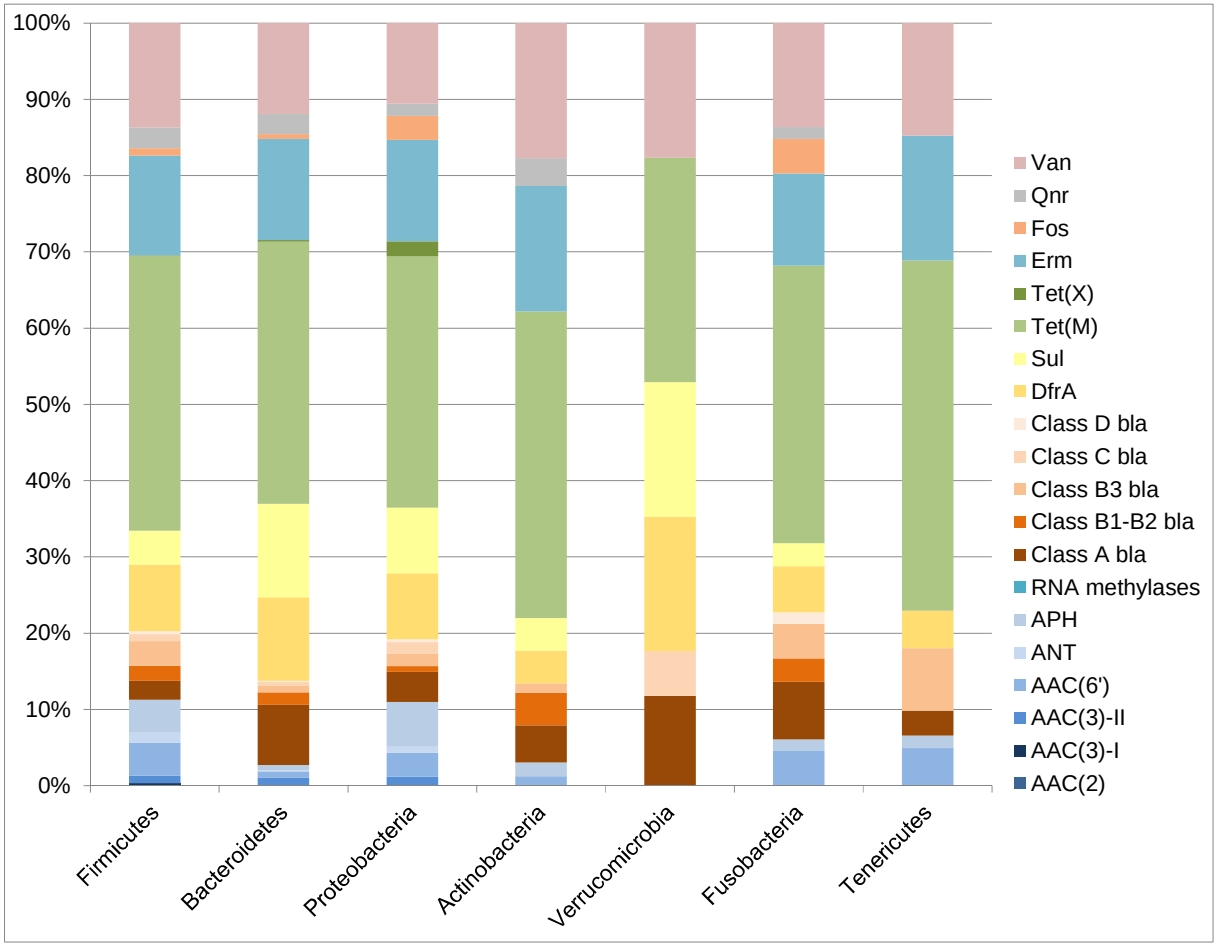


Extended Data Fig. 6: Pie chart of the distribution of phyla among the families of predicted ARDs



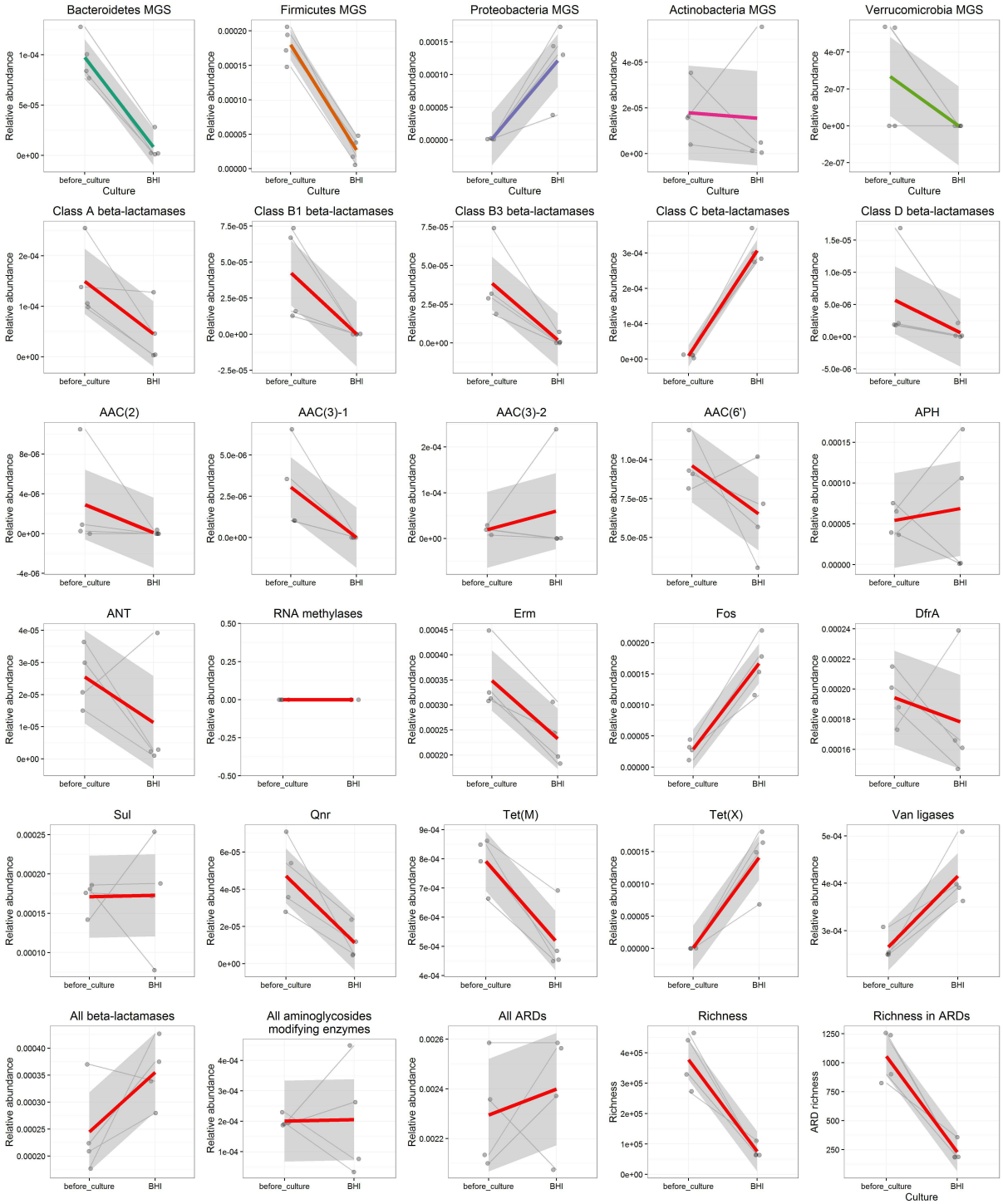
AAC: aminoglycoside acetylase; ANT: aminoglycoside nucleotidyltransferase; APH: aminoglycoside phosphotransferase; DfrA: type A dihydrofolate reductase; Sul: dihydropteroate reductase; Erm: erythromycin ribosome methylase; Qnr: quinolone resistance; Fos: fosfomycin resistance (Fos); Van: D-Ala – D-Lac/Ser ligase (vancomycin resistance).

Extended Data Fig. 7: Bar plot of the distribution of pdARDs among the main phyla



Van: D-Ala – D-Lac/Ser ligase (vancomycin resistance); Qnr: quinolone resistance; Fos: fosfomycin resistance (Fos); Erm: erythromycin ribosome methylase; Sul: dihydropteroate reductase; DfrA: type A dihydrofolate reductase; bla: beta-lactamase; APH: aminoglycoside phosphotransferase; ANT: aminoglycoside nucleotidyl transferase; AAC: aminoglycoside acetylase.

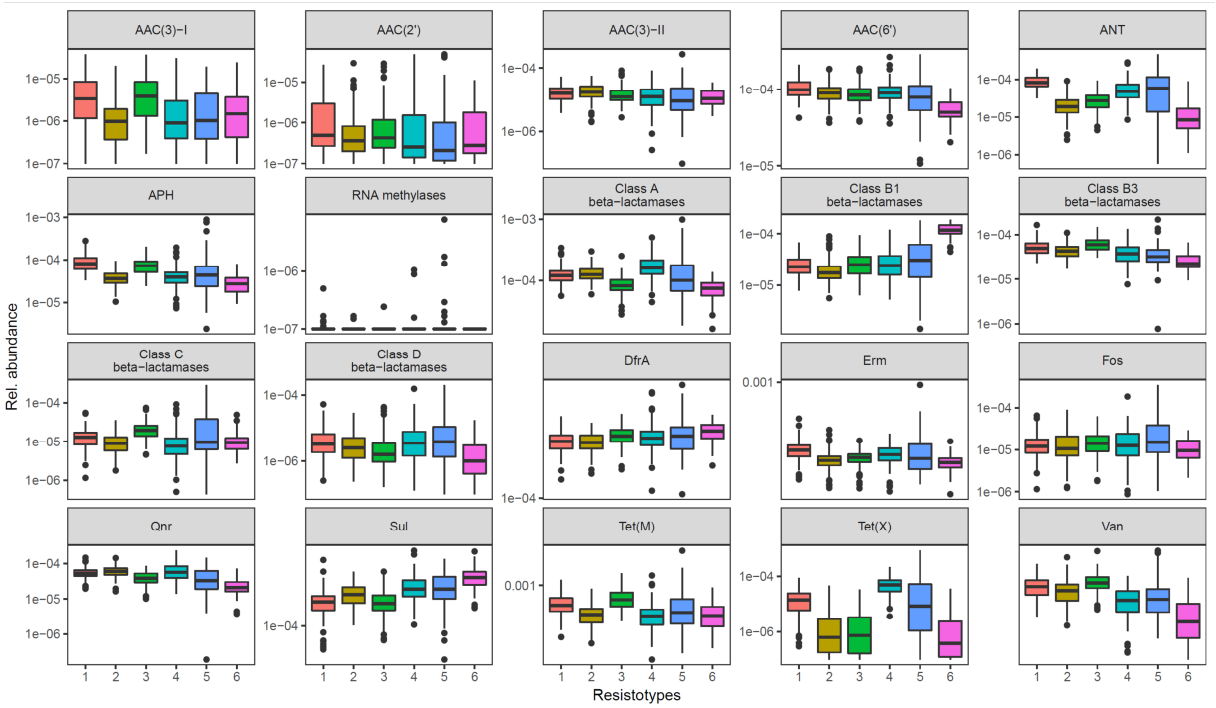
Extended Data Fig. 8: Connected dot-plots of the variations of the relative abundances of the main metagenomics species phyla and pdARDs in faecal samples before and after an aerobic culture in brain-heart infusion (non-specific broth). The shaded grey area depicts the 95% confidence interval around the red, linear regression line.



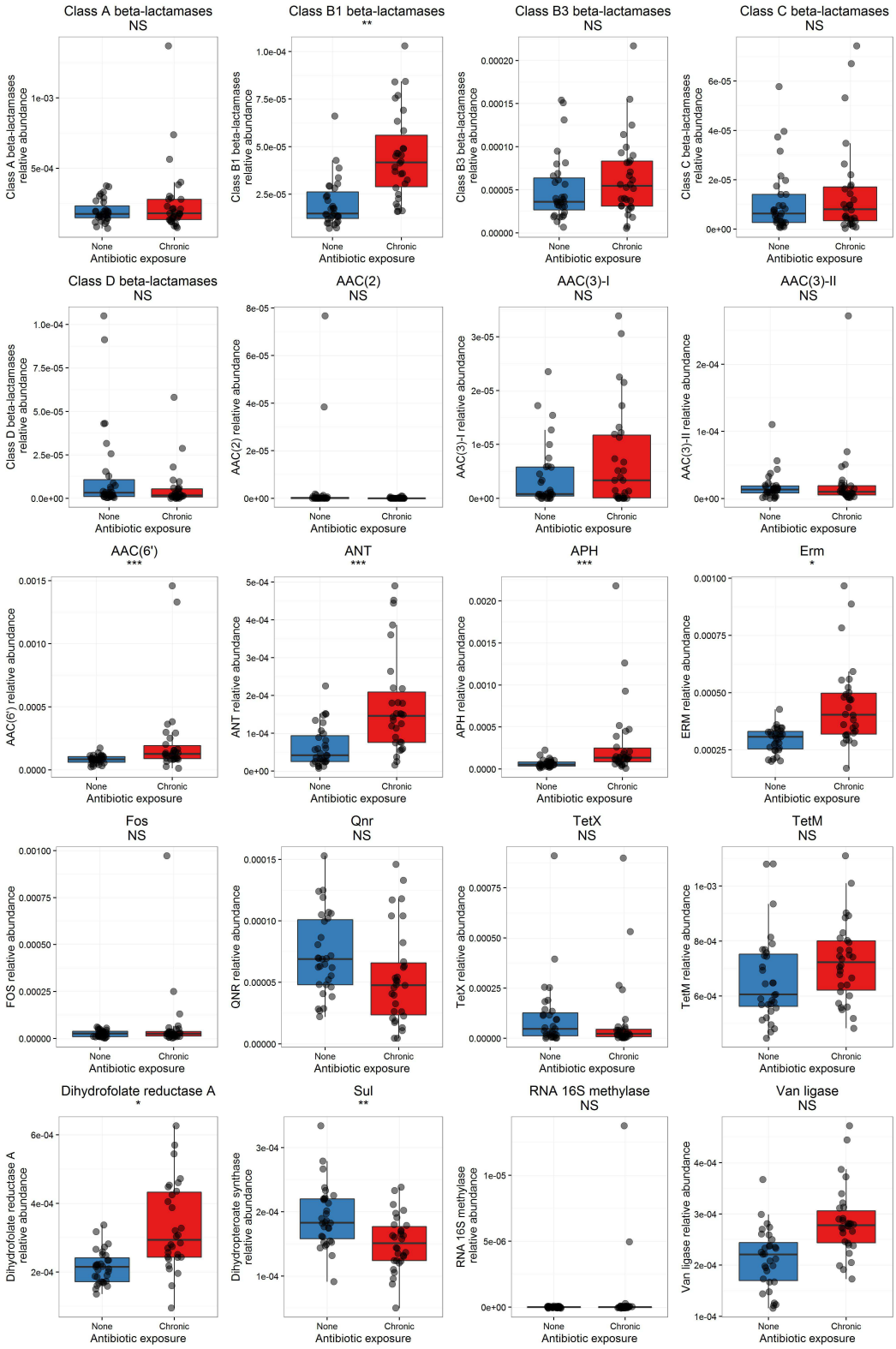
AAC: aminoglycoside acetylase; ANT: aminoglycoside nucleotidyltransferase; APH: aminoglycoside phosphotransferase; DfrA: type A dihydrofolate reductase; Sul: dihydropteroate reductase; Erm:

erythromycin ribosome methylase; Qnr: quinolone resistance; Fos: fosfomycin resistance (Fos); Van: D-Ala – D-Lac/Ser ligase (vancomycin resistance).

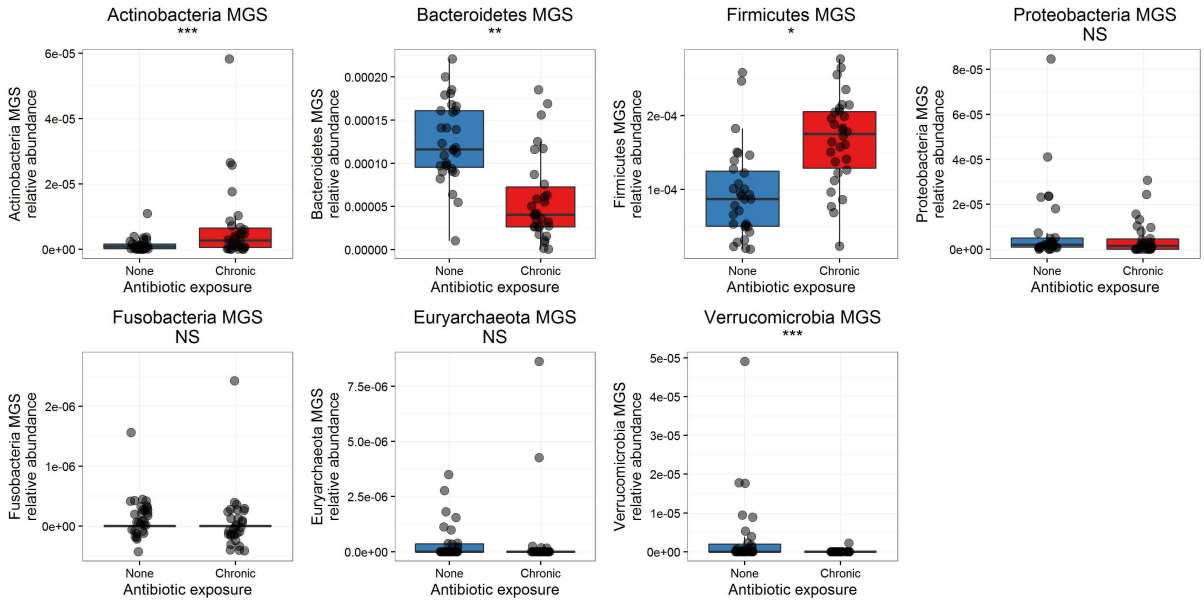
Extended Data Fig. 9: Distribution of the relative abundances of pdARDs according to their family and the resistotypes (x-axis).



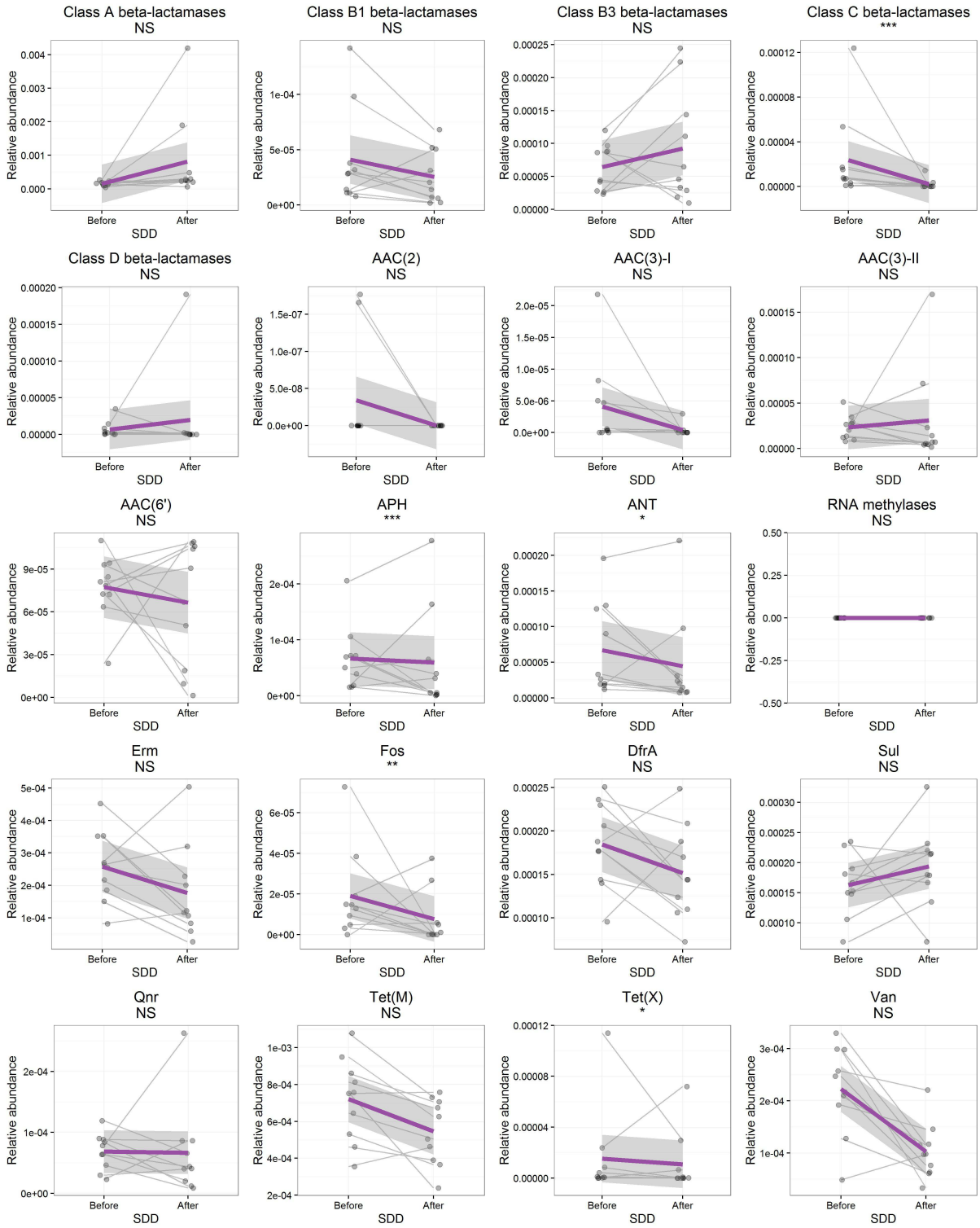
Extended Data Fig. 10: Boxplots and dot-plots of the relative abundance of predicted antibiotic resistance determinants (pdARDs) in subjects with chronic exposure to antibiotics compared to subjects with no recent exposure to antibiotics. The unpaired Wilcoxon test was used. ***: p<0.001, ** p<0.01, *: p<0.05, NS: not significant.



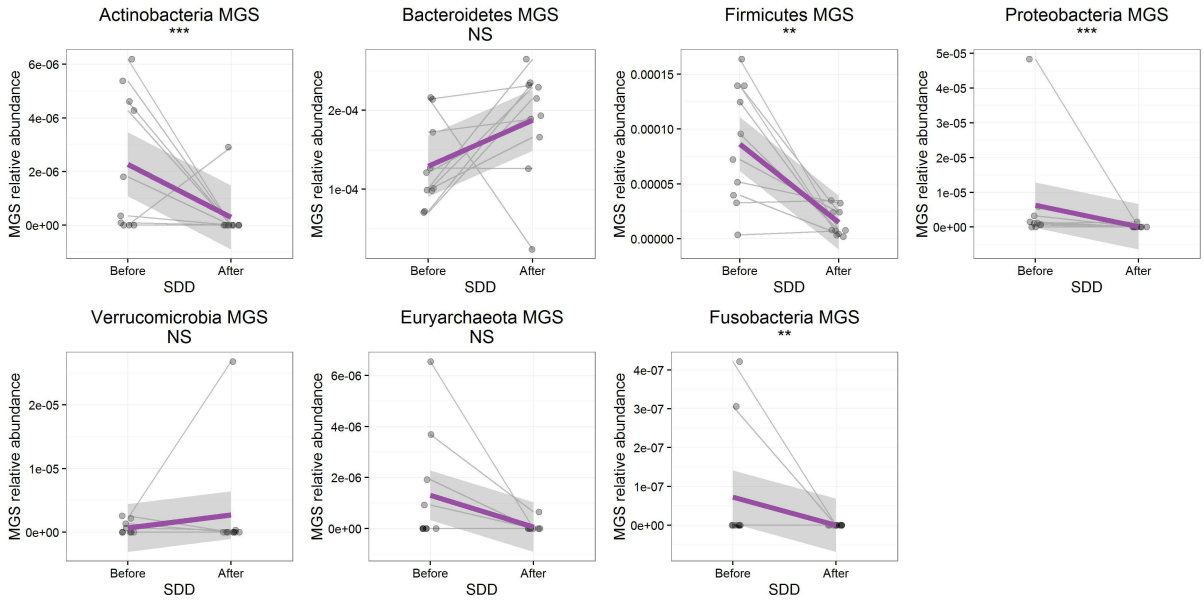
Extended Data Fig. 11: Boxplots and dot-plots of the relative abundance of metagenomic species (MGS) at the phylum level in subjects with chronic exposure to antibiotics compared to subjects with no recent exposure to antibiotics. The unpaired Wilcoxon test was used. ***: p<0.001, ** p<0.01, *: p<0.05, NS: not significant.



Extended Data Fig. 12: Connected dot-plots of the dynamics of the variations of relative abundance of pdARDs before and after selective digestive decontamination (SDD). ***: p<0.001, ** p<0.01, *: p<0.05, NS: not significant. The shaded grey area depicts the 95% confidence interval around the purple, linear regression line.



Extended Data Fig. 13: Connected dot-plots of the relative abundance of metagenomic species (MGS) at the phylum level in subjects before and after selective digestive decontamination (SDD). ***: p<0.001, ** p<0.01, NS: not significant. The shaded grey area depicts the 95% confidence interval around the purple, linear regression line.



Supplementary Tables legends

Supplementary Table 1: The 6,095 predicted antibiotic resistance determinants (pdARDs) that were found in the 3.9M protein catalogue. PCM score missing values means that the candidate could not be modelled with the negative template, so that the PCM score was considered to be over 50%.

Supplementary Table 2: Description of the 16 candidates sharing at least 40% amino acid identity with a reference ARD but being not predicted as an ARD by PCM. ARD: antibiotic resistance determinant. PCM: pairwise comparative modelling. The TM-score represents the degree of correct alignment of the structure generated by PCM and a reference structure (the highest score being 1).

Supplementary Table 3: Details on the 74 metagenomic species (MGS) that were found to be differentially abundant between subjects with no recent antibiotic exposure to antibiotics and subjects under chronic exposure to antibiotics. ARD: antibiotic resistance determinant. Padj: adjusted p-value.

Supplementary Table 4: Details on the 133 metagenomic species (MGS) that were found to be differentially abundant between subjects before and after selective digestive decontamination (SDD). ARD: antibiotic resistance determinant. Padj: adjusted p-value.

Supplementary Table 5: Predictions of antibiotic resistance determinants in the functional metagenomics dataset from soils[5] by the pairwise comparative modelling (PCM) method.