

Supplemental table 1

All bacterial genomes can be obtained from NCBI (<http://www.ncbi.nlm.nih.gov/taxonomy>).

Bacterial strain	Taxonomy_ID
Akkermansia muciniphila ATCC BAA-835	349741
Enterococcus faecalis D32	1206105
Enterococcus faecium DO	333849
Eubacterium rectale ATCC 33656	515619
Lactobacillus brevis ATCC 367	387344
Lactobacillus casei ATCC 334	321967
Methanobrevibacter smithii ATCC 35061	420247
Parabacteroides distasonis ATCC 8503	435591
Roseburia hominis A2-183	585394
Atopobium rimae ATCC 49626	553184
Bacteroides intestinalis DSM 17393	471870
Bifidobacterium angulatum DSM 20098 = JCM 7096	518635
Bifidobacterium bifidum NCIMB 41171	398513
Bifidobacterium longum subsp. longum ATCC 55813	548480
Collinsella aerofaciens ATCC 25986	411903
Collinsella intestinalis DSM 13280	521003
Dorea formicigenerans ATCC 27755	411461
Enterobacter cancerogenus ATCC 35316	500639
Escherichia coli 83972	525281
Faecalibacterium prausnitzii A2-165	411483
Fusobacterium nucleatum subsp. nucleatum ATCC 23726	525283
Fusobacterium periodonticum ATCC 33693	546275
Prevotella copri DSM 18205	537011
Ruminococcus obeum ATCC 29174	411459
Streptococcus salivarius SK126	596322
Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67	321314
Bacteroides vulgatus CL09T03C04	997891
Coprococcus comes SL7/1	717958

Supplemental figure legends

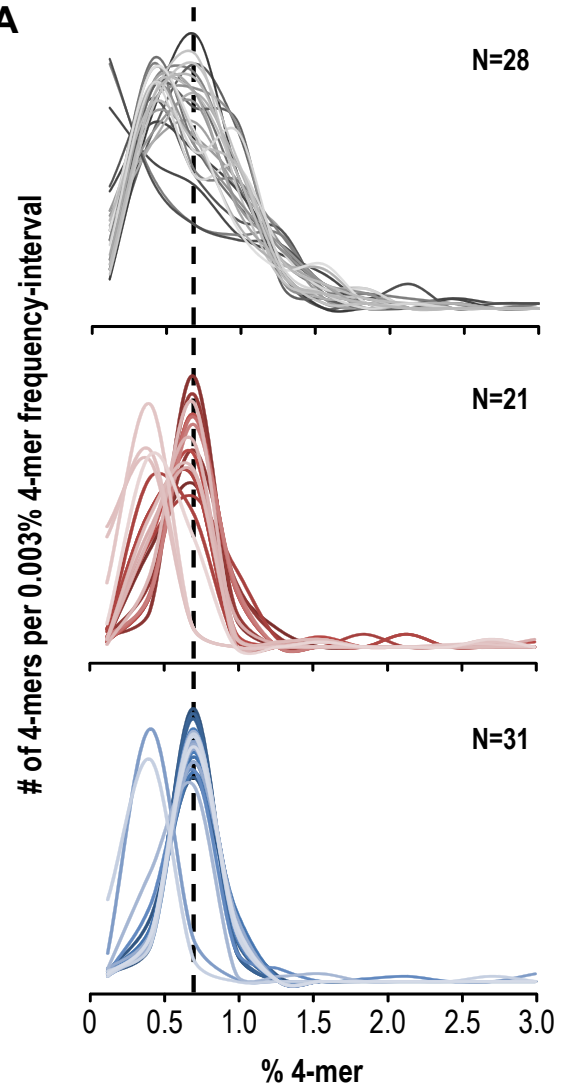
Figure S1. 4-mer distribution analysis of 26 bacterial genomes and 52 metagenomic sequences of gut microbiota from low and high bacterial content samples. **A**, 4-mer abundance distribution of 28 complete bacterial genomes (top panel) and metagenomic sequences of gut microbiota from 21 low ($<10^{10}$ bacteria) (middle panel) and 31 high ($>10^{10}$ bacteria) (bottom panel) bacterial content samples. **B**, Comparison of the median normalized Shannon Entropy of k-mer distributions from the three groups described above with whiskers depicting the 95% confidence interval. P-values are calculated using a non-parametric Mann-Whitney test.

Figure S2. 4-mer distribution analysis of barcode-cassette filtered metagenomic sequences of serially diluted gut microbiota. **A**, 4-mer abundance distribution (left panel) and individual frequency (right panel) of metagenomic sequences from colour coded dilution series metagenomics of gut microbiota from donor #1 (upper panel) and #2 (lower panel). **B**, Bar plot visualizes the normalized Shannon Entropy of 4-mer distribution for undiluted and 10-, 100-, 1000- and 10.000-fold diluted gut microbiota metagenomics from donor #1 (blue) and #2 (red). **C**, Scatter plots depict the correlation between 4-mer distributions of metagenomic sequences from undiluted gut microbiota (y-axis) and 4-mer distributions of metagenomic sequences from 10-, 100-, 1000- and 10.000-fold diluted gut microbiota (x-axis) for donor #1 (upper panel) and #2 (lower panel)

Figure S3. 4-mer distribution analysis of gene mapped metagenomic sequences of serially diluted gut microbiota. **A**, 4-mer abundance distribution (left panel) and individual frequency (right panel) of metagenomic sequences from colour coded dilution series metagenomics of gut microbiota from donor #1 (upper panel) and #2 (lower panel). **B**, Bar plot visualizes the normalized Shannon Entropy of 4-mer distribution for undiluted and 10-, 100-, 1000- and 10.000-fold diluted gut microbiota metagenomics from donor #1 (blue) and #2 (red). **C**, Scatter plots depict the correlation between 4-mer distributions of metagenomic sequences from undiluted gut microbiota (y-axis) and 4-mer distributions of metagenomic sequences from 10-, 100-, 1000- and 10.000-fold diluted gut microbiota (x-axis) for donor #1 (upper panel) and #2 (lower panel)

Figure S1

A



B

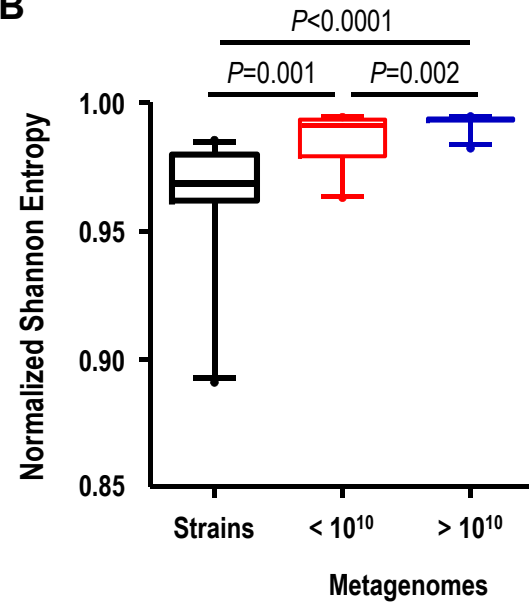


Figure S2

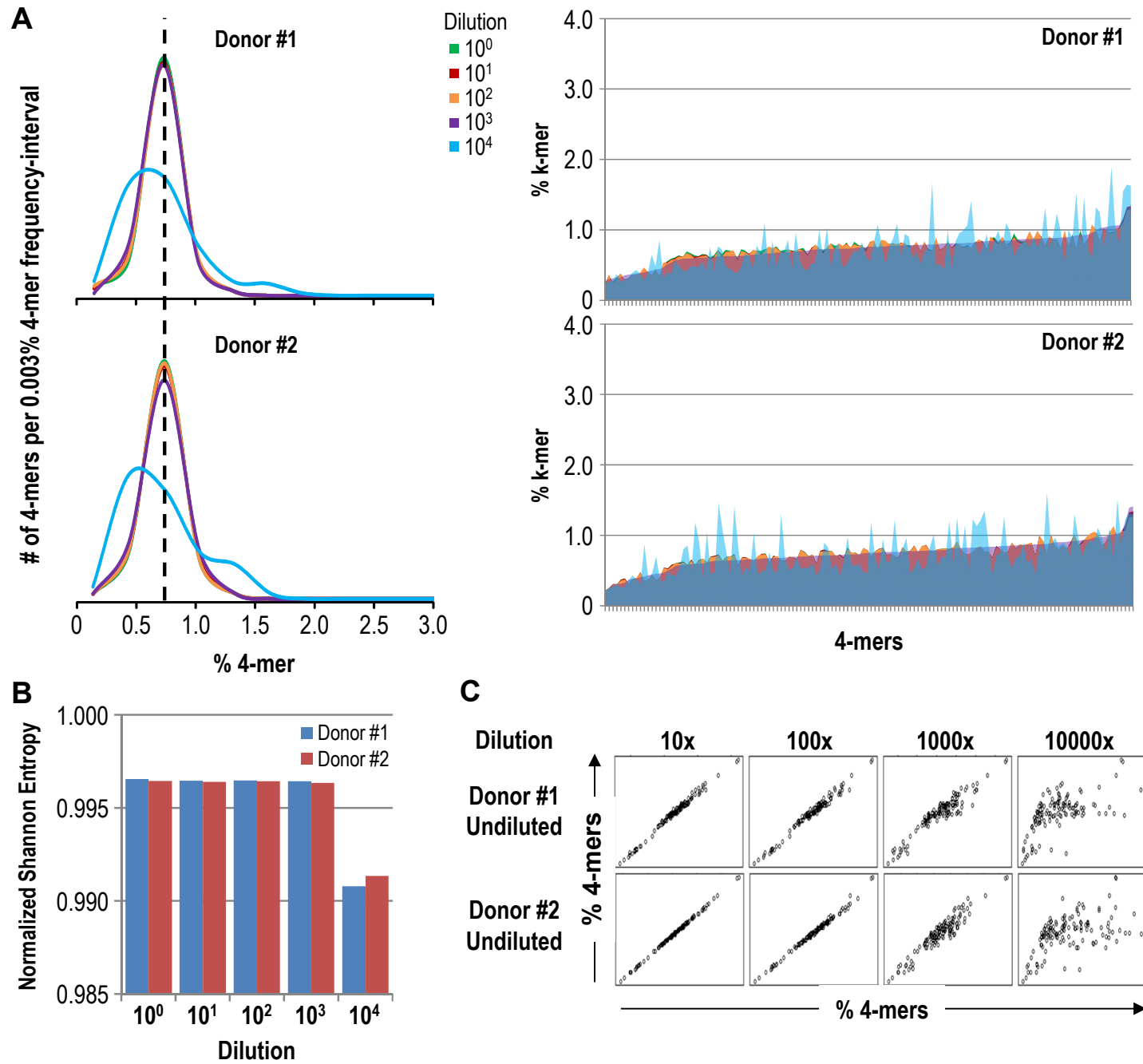


Figure S3

