Supplementary Figures

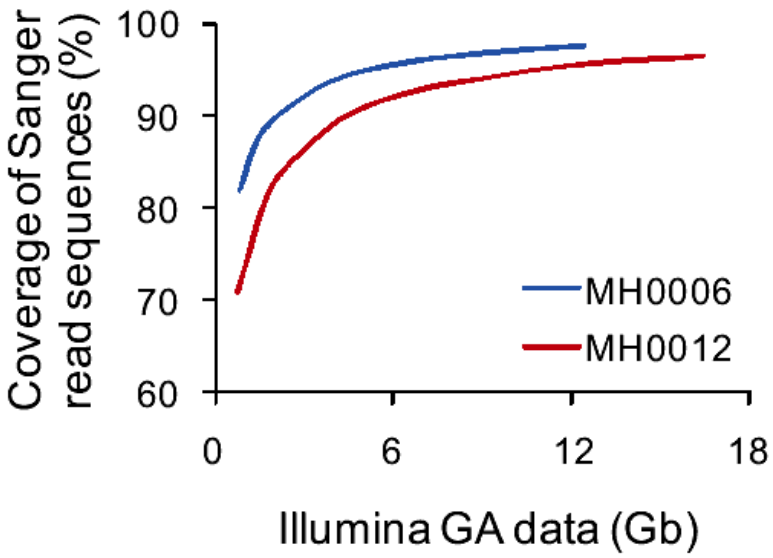

Figure 1 | Coverage of Sanger sequencing reads by Illumina GA reads. Two DNA samples were analyzed by both Sanger and Illumina sequencing. The Illumina GA reads were aligned against the Sanger reads using SOAP. The 50 bp region at the end of the Sanger reads was not taken into account since they may have relatively lower sequence quality and are less likely to be mapped by Illumina GA reads. Subset of Illumina GA reads were extracted from the whole dataset randomly to calculate the coverage of the Sanger read sequence at a given Illumina GA sequencing depth.

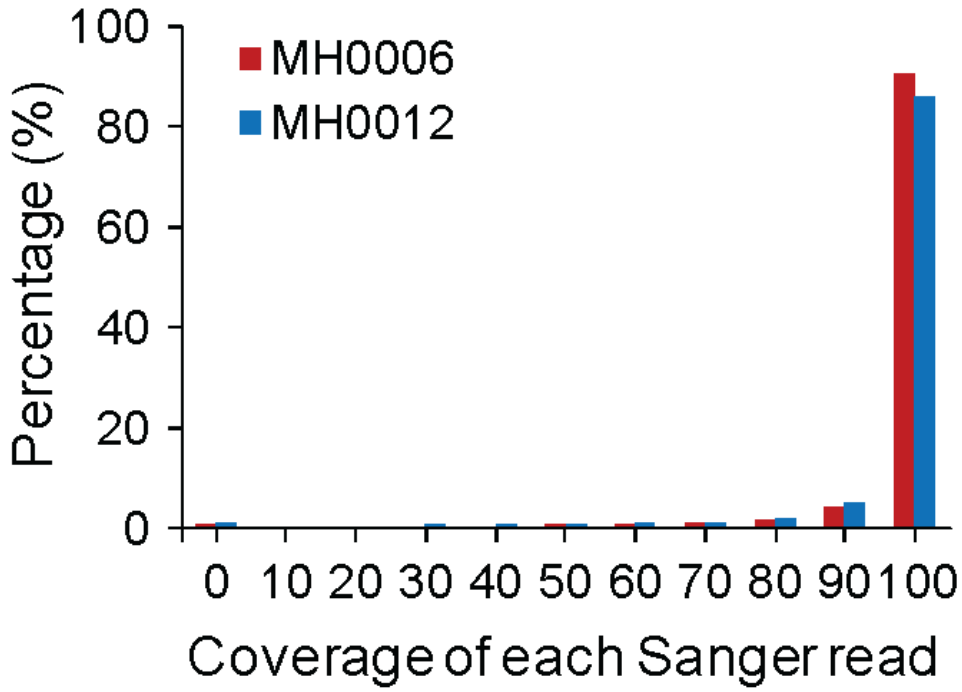

Figure 2 | Distribution of Illumina GA sequencing read coverage of each Sanger read. The Illumina GA reads were aligned against the Sanger reads using SOAP, the coverage of each Sanger read was calculated and the percentage in each coverage interval was determined.

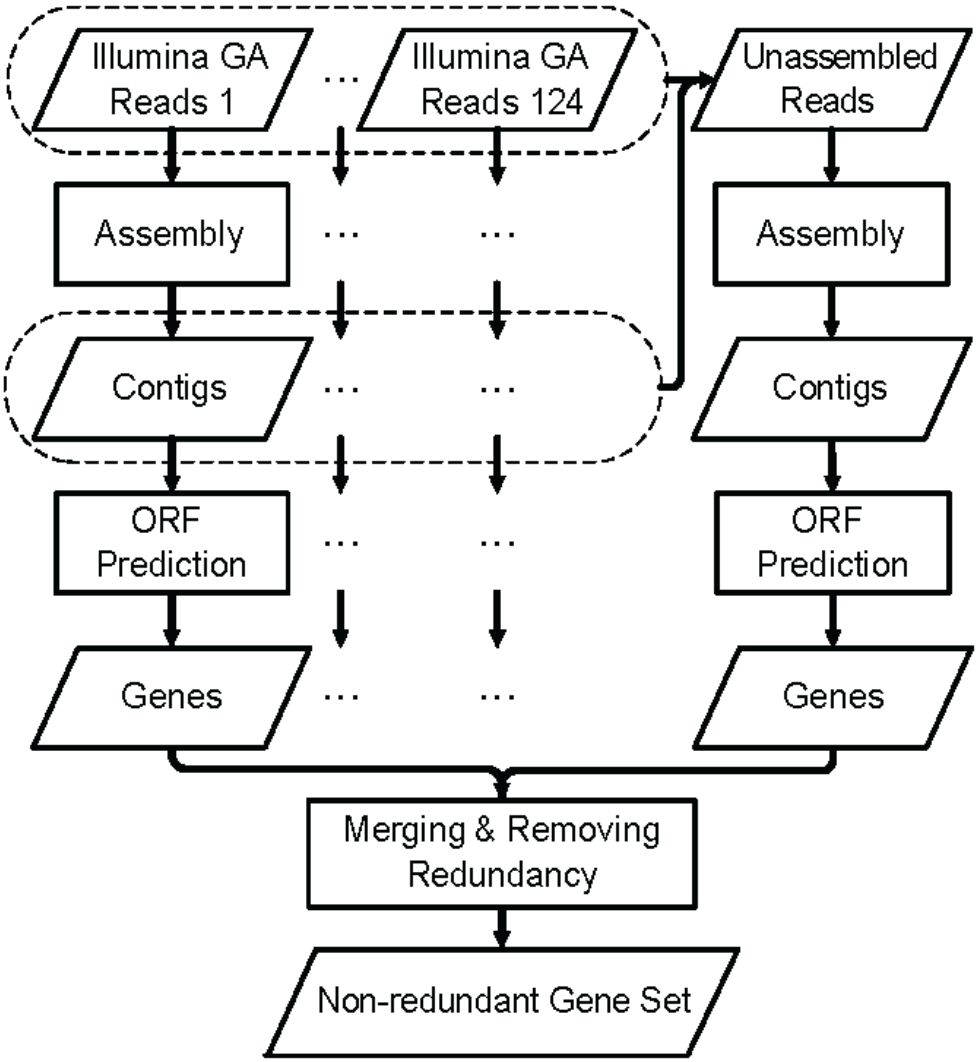

Figure 3 | Flowchart of human gut microbiome data analysis process. We performed *de novo* short reads assembly for each sample independently, then all the unassembled reads were pooled for another round of assembly. ORFs were predicted in each of the contig set, and were merged by removing redundancy. The non-redundant gene set was used in all further analysis.

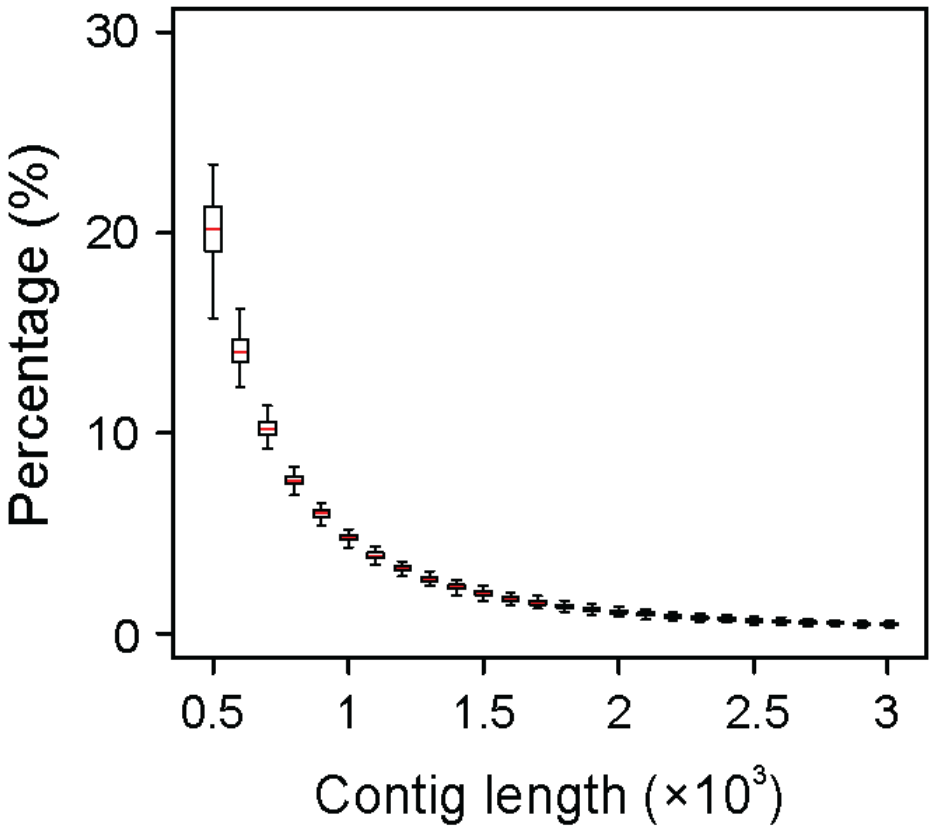

Figure 4 | Length distribution of assembled contigs. The number of contigs in different length bins for each individual was computed, and the data from all 124 individuals were pooled. Boxes denote 25% and 75% percentiles, the red line corresponds to the median, and the "whiskers" indicate interquartile range from either or both ends of the box

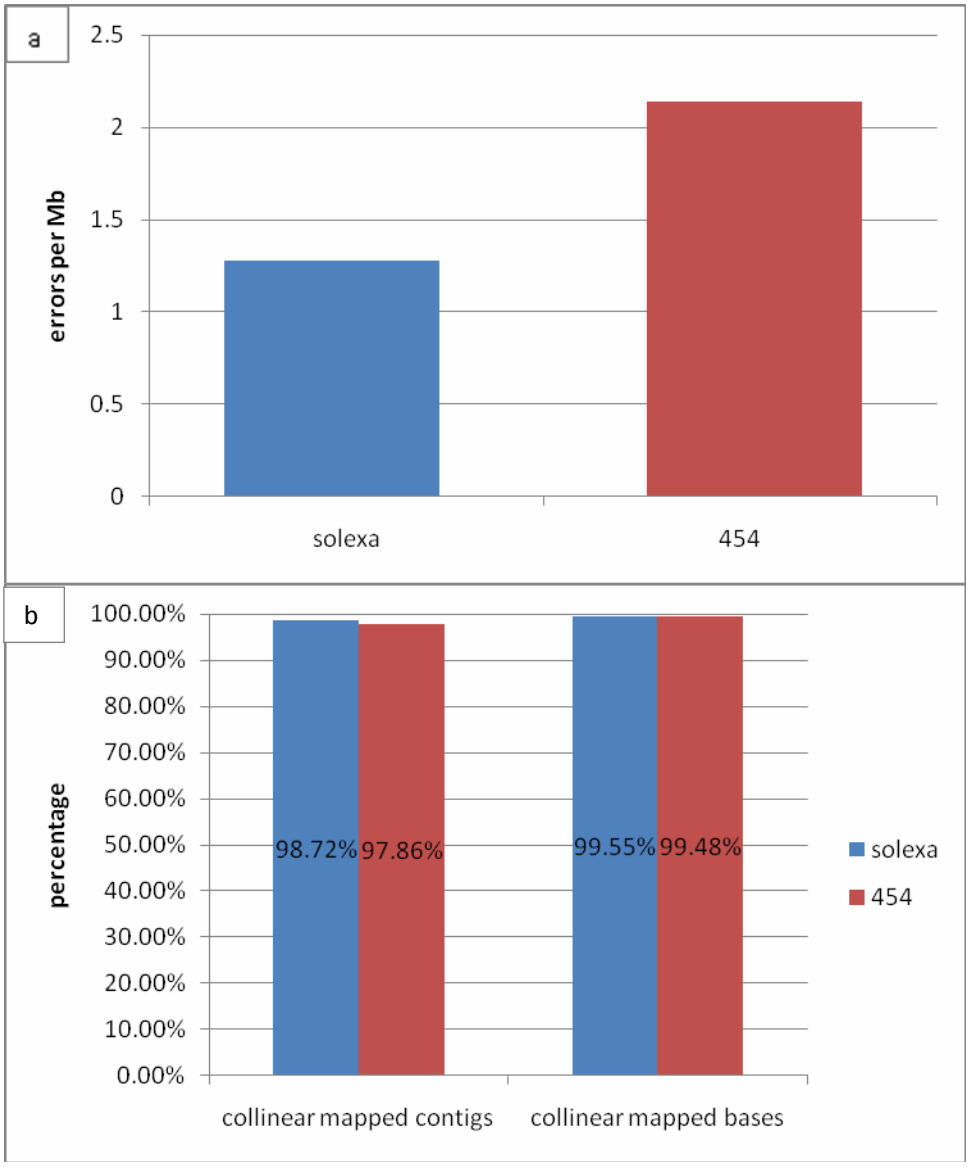

Figure 5 | Validating Illumina contigs using Sanger reads. Illumina/454 contigs from samples MH0006 and MH0012 were mapped to Sanger reads from the same samples. Aligned regions were scanned for breakage of collinearity, and each unique break is counted as an error. a. number of errors per Mb of Illumina/454 contigs mapped to Sanger reads. b. percentage of collinear Illumina/454 contigs and collinear basepairs in those contigs.

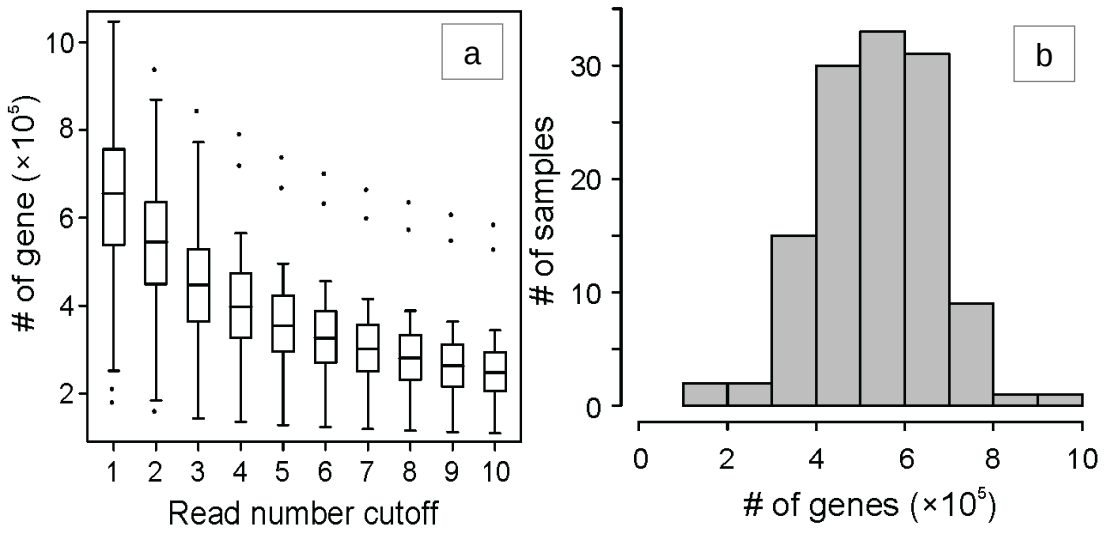

Figure 6 | a, Number of unique genes identified by read number cutoffs. At each cutoff, the minimum, median, and maximum number of genes identified over all the 124 samples are shown. **b, Histogram of number of unique genes in different samples.** Short reads were aligned on the non-redundant gene set, and at least 2 matched reads were required to define a gene's presence in the sample.

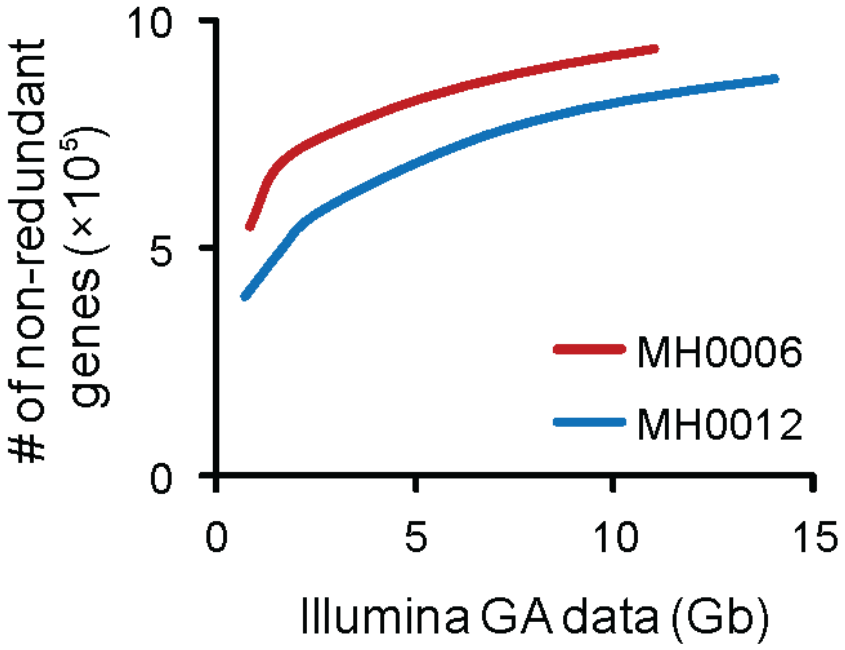

Figure 7 | Number of unique genes identified with increase of sequencing depth in sample MH0006 and MH0012.

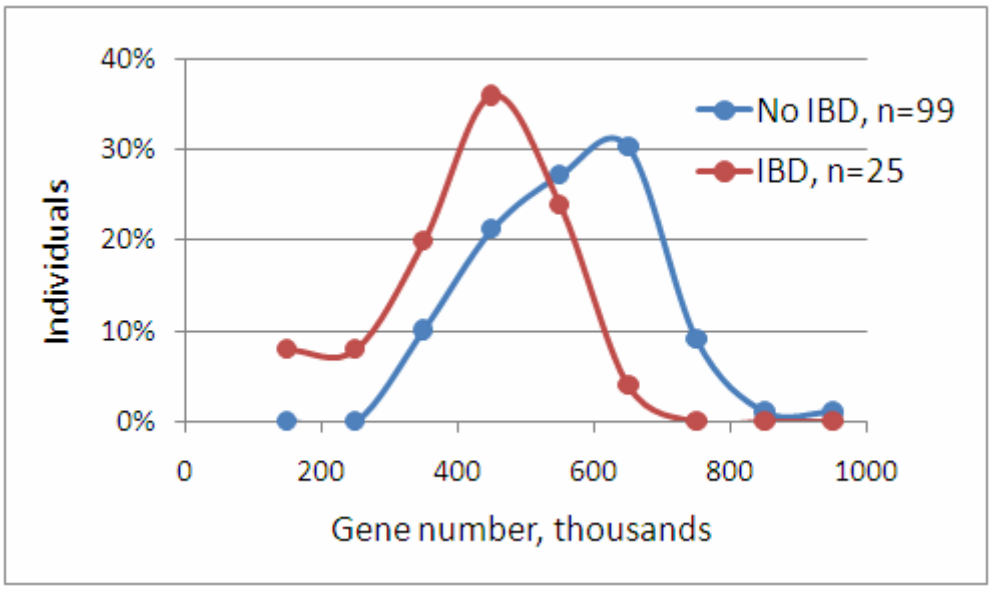

Figure 8 | Distribution of nonredundant bacterial genes in IBD patients and healthy controls. The proportion of individuals having a given number of genes (classes of 100 thousand genes were used) is shown.The average gene number for IBD patients and individuals not suffering from IBD was425,397 $\pm$ 126,685 (s.d.; n=25) and 564,070 $\pm$ 121,962 (s.d.; n=99), respectively; $p<10^{-6}$ (one-tailed Student t test).

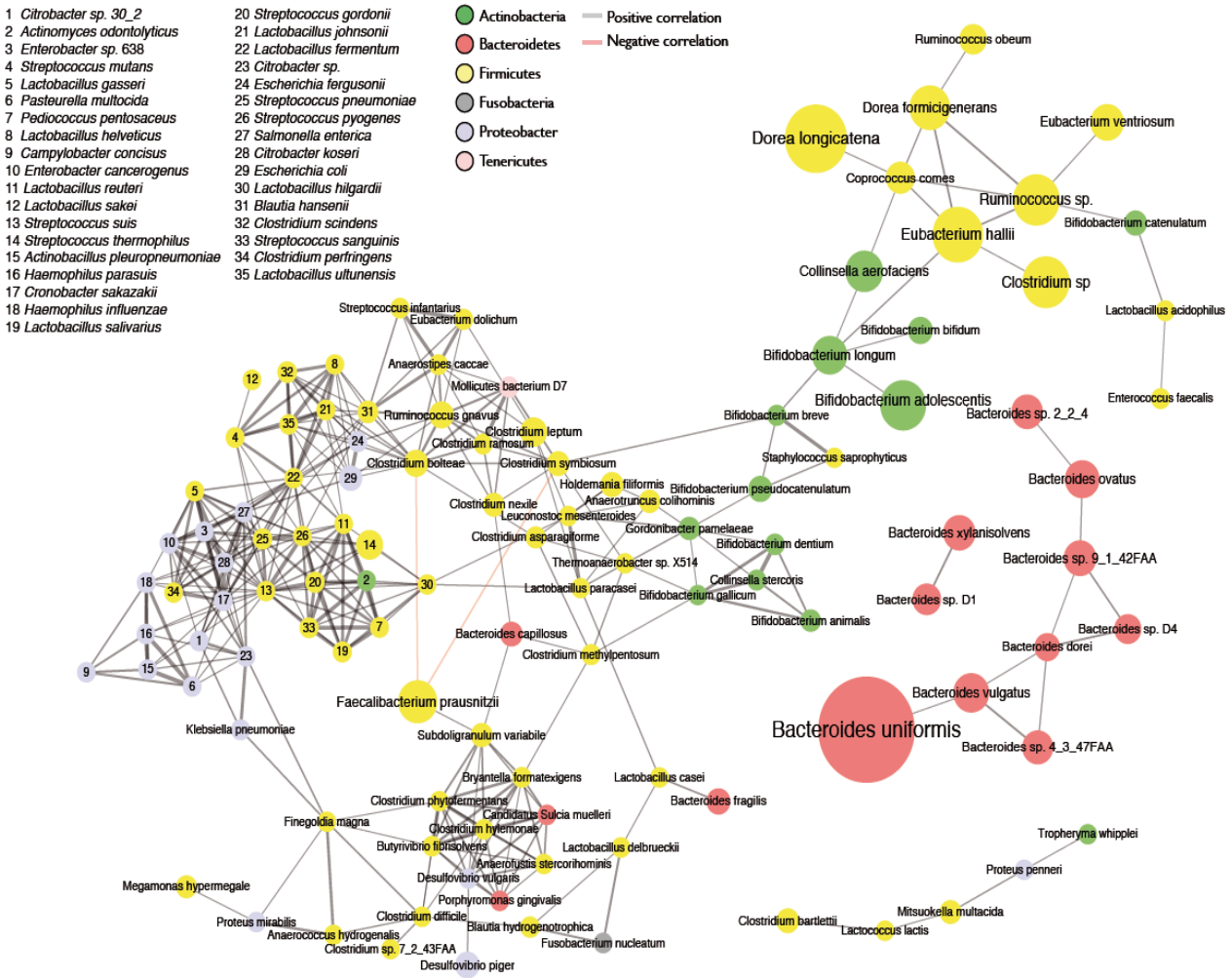

Figure 9 | Relations between the most abundant bacterial species. The network was deduced from the analysis of 155 bacterial species present in at least 1 individual at a genome coverage of ≥1%. Size of the nodes (circles) indicates species abundance over the cohort, width of the edges (lines connecting the circles) indicates the value of the Pearson correlation coefficient (only the 342 values above 0.4 or below -0.4 out of a total of 11,935 were used for the network).

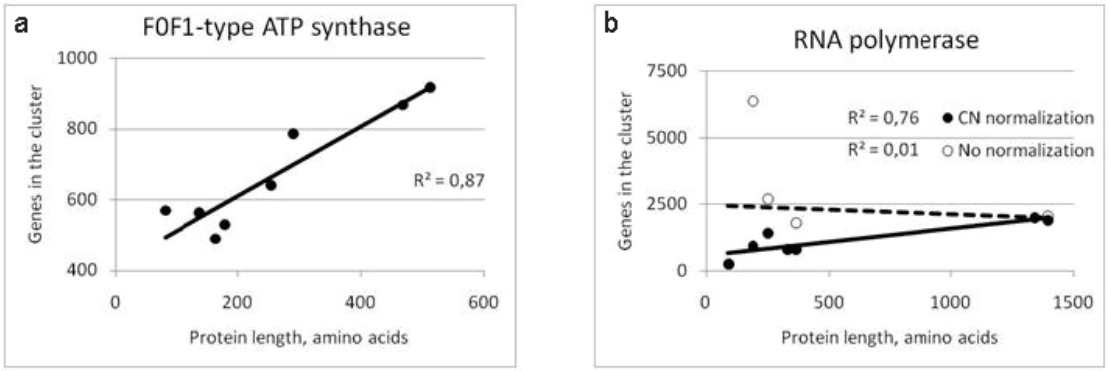

Figure 10 | a, The number of genes assigned to different clusters is correlated with the protein length. b, The effect of copy number (CN) normalization to a single copy is illustrated for RNA polymerase. Protein length is the median of the proteins in the bacterial genomes used to compose eggNOG clusters, the CN is the average number of a function in a genome that encodes it.

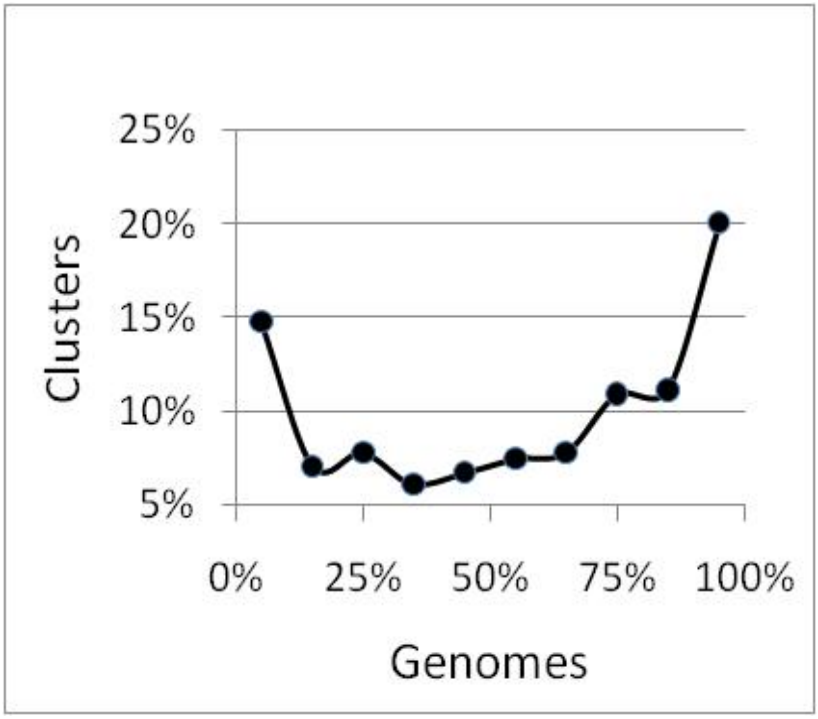

Figure 11 | Distribution of the range clusters across the eggNOG genomes.

Supplementary Tables

Table 1 | DNA sample information.

All Danish individuals in the present subsample were originally recruited from a larger population-based sample of middle-aged people living in the northern part of Copenhagen region and sampled from the centralized personal number register. At the original recruitment the individuals included in the present study had normal fasting plasma glucose and normal 2 hour plasma glucose following an oral glucose tolerance test. At the time of fecal sampling all were examined in the fasting state and had non-diabetic fasting plasma glucose levels below 7,0 mmol/l. All of the IBD patients were in clinical remission at the time of fecal sampling. N refers to no IBD, CD & UC to Crohn's disease and ulcerative colitis, respectively.

Sample Name	Country	Gender	Age	BMI	IBD
MH0001	Denmark	female	49	25.55	N
MH0002	Denmark	female	59	27.28	N
MH0003	Denmark	male	69	33.19	N
MH0004	Denmark	male	59	31.18	N
MH0005	Denmark	male	64	21.68	N
MH0006	Denmark	female	59	22.38	N
MH0007	Denmark	male	69	33.60	N
MH0008	Denmark	male	59	24.35	N
MH0009	Denmark	male	64	29.04	N
MH0010	Denmark	male	64	33.27	N
MH0011	Denmark	female	0	22.31	N
MH0012	Denmark	female	42	32.10	N
MH0013	Denmark	male	54	20.46	N
MH0014	Denmark	female	54	38.49	N
MH0015	Denmark	male	59	25.47	N
MH0016	Denmark	female	49	30.50	N
MH0017	Denmark	male	64	21.81	N
MH0018	Denmark	male	49	31.37	N
MH0019	Denmark	female	44	20.01	N
MH0020	Denmark	female	63	33.23	N
MH0021	Denmark	female	49	25.42	N
MH0022	Denmark	male	64	24.42	N
MH0023	Denmark	male	69	31.74	N
MH0024	Denmark	female	59	22.72	N
MH0025	Denmark	female	49	34.20	N
MH0026	Denmark	female	49	37.32	N
MH0027	Denmark	female	59	23.07	N
MH0028	Denmark	female	44	22.70	N
MH0030	Denmark	male	59	35.21	N
MH0031	Denmark	male	69	22.34	N
MH0032	Denmark	male	69	35.28	N
MH0033	Denmark	female	59	31.95	N

MH0034	Denmark	male	54	39.97	N
MH0035	Denmark	male	49	22.66	N
MH0036	Denmark	male	64	30.74	N
MH0037	Denmark	male	44	24.02	N
MH0038	Denmark	female	54	21.97	N
MH0039	Denmark	male	58	23.07	N
MH0040	Denmark	female	67	20.87	N
MH0041	Denmark	male	59	23.17	N
MH0042	Denmark	male	49	24.46	N
MH0043	Denmark	male	69	23.72	N
MH0044	Denmark	male	64	24.48	N
MH0045	Denmark	male	59	25.11	N
MH0046	Denmark	male	54	23.74	N
MH0047	Denmark	female	69	30.40	N
MH0048	Denmark	female	54	19.40	N
MH0049	Denmark	female	44	35.52	N
MH0050	Denmark	male	49	25.08	N
MH0051	Denmark	female	69	23.15	N
MH0052	Denmark	female	49	33.18	N
MH0053	Denmark	female	49	32.70	N
MH0054	Denmark	male	49	20.31	N
MH0055	Denmark	male	59	30.29	N
MH0056	Denmark	male	54	25.35	N
MH0057	Denmark	female	54	32.98	N
MH0058	Denmark	female	54	22.04	N
MH0059	Denmark	male	59	33.27	N
MH0060	Denmark	male	54	23.52	N
MH0061	Denmark	female	69	30.12	N
MH0062	Denmark	female	49	37.54	N
MH0063	Denmark	male	59	30.23	N
MH0064	Denmark	female	54	23.18	N
MH0065	Denmark	male	59	28.23	N
MH0066	Denmark	female	44	20.79	N
MH0067	Denmark	male	54	21.07	N
MH0068	Denmark	female	54	28.97	N
MH0069	Denmark	female	59	36.71	N
MH0070	Denmark	male	49	22.69	N
MH0071	Denmark	female	44	25.37	N
MH0072	Denmark	female	64	40.21	N
MH0073	Denmark	male	54	32.49	N
MH0074	Denmark	female	49	20.46	N
MH0075	Denmark	male	64	30.55	N
MH0076	Denmark	female	69	34.78	N
MH0077	Denmark	female	49	24.92	N
MH0078	Denmark	female	49	36.90	N
MH0079	Denmark	female	64	19.97	N
MH0080	Denmark	female	59	18.59	N
MH0081	Denmark	female	49	37.95	N
MH0082	Denmark	female	59	22.56	N
MH0083	Denmark	female	54	30.59	N

MH0084	Denmark	male	64	31.67	N
MH0085	Denmark	female	59	36.46	N
MH0086	Denmark	female	59	21.59	N
O2.UC-1	Spain	male	37	31.02	Y
O2.UC-11	Spain	female	34	18.68	Y
O2.UC-12	Spain	male	43	21.60	Y
O2.UC-13	Spain	female	68	23.38	Y
O2.UC-14	Spain	male	31	32.65	Y
O2.UC-16	Spain	male	47	26.42	Y
O2.UC-17	Spain	male	56	21.87	Y
O2.UC-18	Spain	male	48	25.72	Y
O2.UC-19	Spain	male	42	24.15	Y
O2.UC-20	Spain	female	51	24.03	Y
O2.UC-21	Spain	female	49	30.46	Y
O2.UC-22	Spain	male	44	25.39	Y
O2.UC-23	Spain	female	44	28.16	Y
O2.UC-24	Spain	female	55	28.76	Y
O2.UC-4	Spain	female	57	28.53	Y
V1.CD-1	Spain	female	25	17.93	Y
V1.CD-11	Spain	female	62	35.46	N
V1.CD-12	Spain	female	41	20.20	Y
V1.CD-13	Spain	male	68	25.69	N
V1.CD-14	Spain	female	41	23.12	N
V1.CD-15	Spain	female	34	19.00	Y
V1.CD-2	Spain	male	49	27.76	N
V1.CD-3	Spain	female	18	21.51	N
V1.CD-4	Spain	female	46	29.69	N
V1.CD-6	Spain	female	36	18.52	Y
V1.CD-8	Spain	male	51	29.38	N
V1.CD-9	Spain	female	48	27.55	N
V1.UC-10	Spain	male	45	27.31	Y
V1.UC-13	Spain	female	51	28.51	Y
V1.UC-14	Spain	female	53	20.25	Y
V1.UC-15	Spain	female	25	22.77	Y
V1.UC-17	Spain	female	41	24.46	Y
V1.UC-18	Spain	female	63	28.67	N
V1.UC-19	Spain	female	37	21.19	N
V1.UC-21	Spain	male	62	25.21	Y
V1.UC-6	Spain	female	38	23.18	N
V1.UC-7	Spain	female	19	23.05	N
V1.UC-8	Spain	male	22	25.40	N
V1.UC-9	Spain	male	32	30.37	N

Table 2 | Summary of Sanger reads. The reads were sequenced by 3730xl. Low-quality sequences at both ends with phred score less than 20 were trimmed. Very short reads with length less than 100 bp were filtered.

Sample ID	# Sanger reads	Average length (bp)	Total length (bp)
MH0006	237,567	660.65	156,949,306
MH0012	230,768	670.26	154,675,458

Table 3 | Summary of Illumina GA reads. We constructed libraries with three different insert sizes of about 135 bp, 200 bp, and 400 bp. The insert sizes of each library were estimated by re-aligning the paired-end reads on the assembled contigs.

Sample ID	Paired-end insert size (bp)	Read length (bp)	# of reads	Data (Gb)	human reads, %	# of high quality reads
MH0047	136/378	75	35,355,400	2.65	0.18	26,932,064
MH0021	134/354	75	36,454,400	2.73	0.12	26,258,326
MH0079	135/360	75	38,011,600	2.85	0.40	27,418,899
MH0078	146/373	75	38,038,200	2.85	1.56	26,051,537
MH0052	141/367	75	39,538,000	2.97	0.08	28,575,036
MH0049	134/343	75	40,444,200	3.03	0.06	30,654,842
MH0076	134/409	75	40,697,000	3.05	0.42	30,650,106
MH0051	143/374	75	41,911,800	3.14	0.32	25,963,104
MH0048	143/349	75	42,923,600	3.22	0.26	26,972,970
O2.UC-14	141/355	75	43,343,000	3.25	0.06	26,942,750
MH0015	235	44	44,671,400	1.97	0.04	33,014,675
MH0018	233	44	45,081,400	1.98	2.14	36,609,695
MH0027	238	44	45,190,000	1.99	0.09	32,377,390
MH0017	223	44	45,557,200	2.00	0.04	36,154,362
MH0022	256	44	46,415,000	2.04	0.21	37,112,508
MH0023	237	44	48,598,400	2.14	0.04	37,782,998
MH0019	249	44	49,229,400	2.17	0.06	38,856,780
MH0026	156/398	75	49,812,000	3.74	0.05	37,484,066
MH0013	238	44	50,257,200	2.21	1.63	40,028,120
MH0005	237	44	50,704,800	2.23	0.23	39,407,333
MH0007	195	44	50,719,800	2.23	0.31	36,956,284
MH0008	219	44	51,411,000	2.26	0.10	38,156,496
V1.UC-7	141/356	75	51,911,400	3.89	14.67	36,788,540
MH0010	220	44	52,218,200	2.30	0.08	39,169,850
V1.CD-12	148/361	75	53,519,400	4.01	0.02	40,609,134
O2.UC-20	141/362	75	53,637,200	4.02	0.03	38,376,747
V1.CD-15	143/351	75	53,938,600	4.05	2.85	40,560,446
O2.UC-19	133/352	75	54,537,600	4.09	0.01	38,459,550
MH0004	218	44	55,829,800	2.46	0.95	40,288,492
MH0062	144/357	75	57,128,400	4.28	14.32	36,809,224
MH0066	147/429	75	57,234,200	4.29	0.05	36,114,997
O2.UC-21	142/362	75	57,856,000	4.34	0.03	34,832,308
V1.CD-13	139/352	75	58,145,800	4.36	0.04	42,560,831
MH0080	140/376	75	58,220,800	4.37	0.13	46,590,749
V1.UC-13	131/352	75	58,381,400	4.38	9.77	38,553,580
MH0032	142/370	75	58,822,400	3.93	0.39	50,110,067
O2.UC-12	153/384	75	58,927,800	4.42	0.12	36,908,526
MH0001	214	44	59,239,200	2.61	0.06	45,016,612
V1.UC-6	142/376	75	59,270,800	4.45	0.41	43,150,856
MH0060	142/367	75	60,156,000	4.51	0.07	41,112,227
MH0053	137/416	75	60,788,600	4.56	0.07	43,283,564
MH0002	139/370	75	61,077,000	4.58	0.15	46,570,095
O2.UC-11	142/377	75	61,253,800	4.59	0.14	38,507,042
MH0059	136/370	75	61,574,600	4.62	0.18	41,025,606

V1.UC-8	131/346	75	61,781,000	4.63	0.54	46,669,762
O2.UC-1	142/363	75	61,879,000	4.64	5.78	42,189,776
MH0054	137/401	75	62,086,000	4.66	0.15	42,262,748
MH0030	142/354	75	62,148,800	4.18	0.57	51,622,803
MH0046	140/372	75	62,539,800	4.69	0.06	35,798,104
MH0040	136/373	75	62,661,200	4.70	0.05	44,249,732
V1.UC-9	132/362	75	62,758,600	4.71	0.06	46,906,812
MH0025	146/356	75	62,948,000	4.22	0.06	55,604,591
V1.UC-10	141/356	75	63,137,000	4.74	9.54	41,725,594
O2.UC-22	140/363	75	63,220,000	4.74	0.04	44,245,030
V1.CD-14	143/356	75	63,574,000	4.77	0.06	44,023,096
O2.UC-17	135/350	75	63,583,000	4.77	0.02	40,304,612
MH0028	138/368	75	63,652,000	4.25	0.04	55,152,519
MH0057	140/379	75	64,021,000	4.80	0.09	43,050,028
MH0024	124/353	75	64,097,000	4.30	0.10	54,915,985
V1.CD-9	137/348	75	64,146,600	4.81	0.05	47,137,595
MH0041	141/364	75	64,246,800	4.82	0.44	44,382,480
MH0037	145/377	75	64,247,400	4.82	0.05	42,034,099
O2.UC-24	142/372	75	64,629,200	4.85	0.25	41,658,584
O2.UC-16	134/352	75	64,811,000	4.86	0.01	42,268,974
O2.UC-23	141/363	75	64,898,800	4.87	21.33	36,784,558
MH0055	137/412	75	65,444,600	4.91	0.09	47,112,307
V1.UC-21	150/372	75	65,652,400	4.92	0.41	46,336,135
MH0071	137/388	75	65,684,000	4.93	0.23	46,242,273
MH0050	141/339	75	65,725,200	4.93	0.05	47,792,033
MH0020	148/375	75	66,964,400	5.02	0.61	45,813,646
MH0081	128/365	75	66,988,600	5.02	0.04	47,851,897
V1.UC-18	146/374	75	67,064,200	5.03	0.14	48,379,265
MH0085	136/379	75	67,083,400	5.03	0.08	49,434,491
O2.UC-18	138/360	75	67,094,400	5.03	0.08	41,907,663
MH0044	139/383	75	67,124,400	5.03	0.05	44,864,080
MH0074	142/394	75	67,143,600	5.04	0.03	46,030,505
MH0068	145/366	75	67,160,800	5.04	0.09	43,867,534
MH0031	136/381	75	67,460,600	4.54	0.72	55,540,263
MH0045	130/380	75	67,520,800	5.06	0.13	49,434,301
V1.CD-3	144/374	75	67,524,800	5.06	0.38	45,352,742
MH0033	148/377	75	67,543,400	5.07	0.05	44,996,133
MH0064	142/375	75	67,747,800	5.08	0.13	44,732,297
MH0083	133/378	75	68,074,800	5.11	0.03	47,903,806
MH0082	134/359	75	68,130,400	5.11	0.19	48,347,954
V1.UC-15	139/367	75	68,192,000	5.11	1.95	46,595,929
MH0058	145/369	75	68,286,800	5.12	0.13	49,822,998
V1.UC-19	141/354	75	68,295,000	5.12	0.12	49,233,576
O2.UC-13	143/364	75	68,563,400	5.14	0.08	43,619,828
V1.UC-14	142/363	75	68,566,800	5.14	0.68	49,611,323
V1.CD-8	145/356	75	68,642,000	5.15	0.11	49,908,306
MH0067	135/369	75	68,699,200	5.15	0.02	44,484,637
O2.UC-4	140/361	75	68,735,200	5.16	0.15	43,263,975
MH0065	144/420	75	68,945,200	5.17	0.14	44,579,795
MH0056	135/368	75	68,969,600	5.17	0.78	49,219,471

MH0003	143/357	75	69,136,400	5.19	0.07	50,384,869
MH0075	136/382	75	69,195,400	5.19	0.03	47,065,103
MH0061	134/362	75	69,261,600	5.19	2.08	44,515,482
V1.CD-6	136/358	75	69,437,200	5.21	2.65	42,483,316
V1.CD-1	143/366	75	69,669,800	5.23	0.11	47,131,739
MH0070	137/363	75	69,670,400	5.23	0.17	47,921,594
V1.CD-4	142/375	75	70,111,400	5.26	0.04	42,939,132
V1.CD-2	142/374	75	70,150,600	5.26	0.36	46,703,072
MH0063	141/371	75	70,215,000	5.27	0.08	45,961,029
MH0077	135/377	75	70,381,000	5.28	0.27	49,660,515
MH0039	144/370	75	70,466,000	5.28	0.04	42,717,184
MH0073	139/373	75	71,179,400	5.34	0.15	47,234,201
MH0036	142/368	75	71,764,000	5.38	0.14	47,706,635
V1.UC-17	138/356	75	72,222,200	5.42	26.13	34,437,937
MH0035	138/355	75	72,306,000	5.42	0.80	49,289,239
MH0042	141/416	75	72,568,200	5.44	0.36	43,424,052
MH0038	144/372	75	74,210,800	5.57	0.25	45,475,391
MH0043	143/366	75	74,479,600	5.59	0.82	49,362,253
MH0084	140/365	75	75,047,200	5.63	0.07	53,565,815
V1.CD-11	138/348	75	75,344,800	5.65	0.02	51,389,665
MH0016	131/374	75	76,178,600	5.71	0.08	54,396,021
MH0014	141/390	75	77,468,200	5.81	0.21	56,304,766
MH0086	146/432	75	78,495,800	5.89	0.04	55,433,667
MH0011	146/368	75	79,208,000	5.94	0.27	58,503,358
MH0009	146/373	75	80,011,000	6.00	0.14	58,452,504
MH0072	147/435	75	89,101,800	6.68	0.19	49,286,672
MH0034	147/373	75	94,636,800	7.10	0.01	37,400,897
MH0069	152/430	75	97,568,600	7.32	0.06	69,722,220
MH0006	233	44/75	188,437,600	12.60	0.39	161,184,217
MH0012	134/378	75	221,338,200	16.60	0.05	186,062,388
Total			**8,037,389,400**	**576.72**	**Median**	**5,637,140,962**
Average			**62,521,423**	**4.49**	**0.12**	**43,359,790**

Table 4 | Summary of de novo assembly results. Assembled sequences with length below 500 bp were excluded from the contig set.

Sample ID	# of contigs	Contig N50 (bp)	Total length (Mb)	% reads assembled	Unassembled reads (Gb)
MH0001	14,301	1,618	19.69	46.34	1.06
MH0002	65,392	1,680	88.77	45.31	1.91
MH0003	68,658	2,640	119.59	54.40	1.72
MH0004	23,793	1,681	31.92	41.54	1.05
MH0005	14,339	1,684	19.62	40.22	1.04
MH0006	144,440	2,025	217.77	52.39	5.24
MH0007	28,108	1,270	32.00	29.15	1.16
MH0008	26,506	1,768	37.24	43.53	0.95
MH0009	70,014	2,440	112.96	44.14	2.45
MH0010	25,674	1,815	36.52	48.77	0.88
MH0011	86,201	2,158	134.25	46.09	2.37
MH0012	140,991	2,478	237.58	42.77	7.99
MH0013	20,495	2,332	32.20	41.22	1.05
MH0014	66,724	2,957	120.54	50.90	2.08
MH0015	25,933	1,645	34.46	35.53	0.94
MH0016	64,124	2,915	114.03	53.89	1.88
MH0017	24,948	1,679	34.06	39.57	0.96
MH0018	13,247	1,619	17.73	35.23	1.07
MH0019	28,786	1,977	41.95	46.76	0.91
MH0020	44,930	4,708	98.78	56.81	1.49
MH0021	54,101	1,608	70.67	46.49	1.06
MH0022	21,872	1,773	30.00	35.04	1.06
MH0023	16,214	2,100	25.57	35.80	1.07
MH0024	43,145	1,512	54.45	33.21	2.41
MH0025	76,287	1,968	111.48	42.24	2.11
MH0026	33,408	3,769	69.38	43.97	1.58
MH0027	20,369	985	19.72	19.96	1.14
MH0028	61,004	2,630	104.54	49.65	1.80
MH0030	39,267	2,828	66.60	36.73	2.15
MH0031	53,292	1,878	75.84	36.41	2.33
MH0032	37,287	1,921	54.46	20.72	2.60
MH0033	61,782	2,616	102.04	49.87	1.69
MH0034	24,508	2,107	37.63	14.53	2.40
MH0035	68,287	2,075	102.94	48.22	1.93
MH0036	58,690	2,330	94.35	49.44	1.81
MH0037	48,356	2,526	80.44	48.21	1.63
MH0038	50,381	2,921	90.35	47.75	1.79
MH0039	66,509	2,087	104.17	47.66	1.68
MH0040	73,068	2,225	115.15	49.53	1.68

MH0041	55,146	2,531	91.83	52.02	1.60
MH0042	60,378	2,046	90.65	34.38	2.14
MH0043	74,846	1,987	111.11	47.58	1.96
MH0044	49,190	3,590	96.43	50.68	1.66
MH0045	65,434	1,898	94.43	49.18	1.89
MH0046	50,836	2,290	81.81	43.04	1.53
MH0047	36,612	1,374	44.15	41.16	1.19
MH0048	48,416	1,460	60.16	37.77	1.26
MH0049	29,468	2,514	47.75	42.90	1.31
MH0050	71,587	2,096	109.04	49.10	1.83
MH0051	48,208	1,690	65.49	43.82	1.10
MH0052	53,204	1,642	70.93	41.89	1.25
MH0053	52,816	2,396	86.76	42.96	1.85
MH0054	59,223	2,669	102.44	54.54	1.44
MH0055	61,507	2,511	104.18	50.22	1.76
MH0056	71,029	1,678	95.95	32.90	2.80
MH0057	40,554	2,691	72.17	49.88	1.62
MH0058	75,310	2,136	116.26	46.07	2.02
MH0059	77,573	2,176	124.22	48.44	1.59
MH0060	63,059	1,966	92.86	47.05	1.63
MH0061	37,149	3,619	73.17	48.68	1.75
MH0062	59,196	2,134	90.62	40.38	1.92
MH0063	56,634	2,049	84.50	39.93	2.07
MH0064	54,725	2,187	85.58	32.20	2.28
MH0065	69,122	2,083	104.22	44.04	1.87
MH0066	56,860	2,466	92.91	42.80	1.55
MH0067	41,271	4,494	95.51	62.76	1.24
MH0068	45,417	3,249	83.23	49.89	1.65
MH0069	80,769	1,908	117.50	39.35	3.17
MH0070	72,871	2,208	112.98	46.83	1.91
MH0071	53,046	2,810	92.40	37.73	2.16
MH0072	52,323	1,784	74.17	36.22	2.36
MH0073	56,858	2,125	87.06	53.49	1.65
MH0074	44,523	3,990	95.83	56.27	1.51
MH0075	68,045	2,089	103.76	45.96	1.91
MH0076	57,502	1,812	80.56	47.18	1.22
MH0077	78,357	1,744	108.27	47.23	1.97
MH0078	22,164	1,658	30.01	25.41	1.48
MH0079	48,717	1,622	64.50	48.42	1.06
MH0080	76,021	2,074	116.51	54.64	1.59
MH0081	73,686	1,930	108.94	51.34	1.75
MH0082	73,890	2,211	115.72	49.36	1.84
MH0083	66,302	2,337	107.84	50.17	1.79

MH0084	32,542	2,233	50.16	19.55	3.23
MH0085	47,522	2,935	86.29	43.61	2.09
MH0086	88,752	2,381	148.47	53.70	1.93
O2.UC-1	51,221	1,682	69.20	35.00	2.18
O2.UC-11	36,782	1,670	50.37	26.39	2.13
O2.UC-12	47,451	2,264	75.28	40.28	1.66
O2.UC-13	54,055	2,853	95.05	48.36	1.69
O2.UC-14	26,263	2,135	39.80	27.46	1.47
O2.UC-16	25,434	1,561	33.39	14.31	2.72
O2.UC-17	31,446	1,486	40.55	20.05	2.42
O2.UC-18	36,747	3,824	74.01	39.66	1.90
O2.UC-19	59,079	2,311	92.86	47.36	1.52
O2.UC-20	57,728	1,836	82.85	46.40	1.54
O2.UC-21	37,854	1,478	47.55	23.14	2.01
O2.UC-22	66,174	1,989	100.24	45.25	1.82
O2.UC-23	46,329	1,959	66.70	27.21	2.60
O2.UC-24	48,257	2,350	76.25	31.12	2.16
O2.UC-4	52,210	2,293	82.94	44.75	1.80
V1.CD-1	30,435	2,912	54.83	33.38	2.36
V1.CD-11	43,837	3,775	83.58	47.70	2.02
V1.CD-12	36,815	2,283	59.20	50.33	1.51
V1.CD-13	63,979	2,207	99.94	48.08	1.66
V1.CD-14	76,242	2,060	116.12	48.96	1.69
V1.CD-15	30,840	3,185	56.10	46.47	1.68
V1.CD-2	16,741	715	12.26	2.34	3.43
V1.CD-3	49,312	1,986	74.88	39.79	2.06
V1.CD-4	64,855	2,172	100.99	49.31	1.63
V1.CD-6	52,405	1,912	76.79	45.10	1.80
V1.CD-8	76,704	2,125	116.87	44.60	2.08
V1.CD-9	51,597	3,444	99.57	53.10	1.66
V1.UC-10	70,208	2,290	112.96	43.16	2.00
V1.UC-13	54,707	2,788	94.12	47.55	1.70
V1.UC-14	60,809	2,443	99.23	49.47	1.89
V1.UC-15	59,391	2,731	104.22	50.02	1.78
V1.UC-17	51,062	2,180	79.04	31.53	2.35
V1.UC-18	54,194	2,721	94.28	54.98	1.64
V1.UC-19	65,470	2,825	113.78	50.30	1.84
V1.UC-21	55,907	2,226	89.43	49.60	1.76
V1.UC-6	88,651	1,766	125.25	49.50	1.64
V1.UC-7	35,227	2,405	56.97	28.50	2.31
V1.UC-8	76,786	2,197	121.81	52.32	1.68
V1.UC-9	69,741	1,842	99.85	43.12	2.00

Table 5 | List of 194 public human gut bacterial genomes. "HMP" stands for strains which were sequenced by the Human Microbiome Project of NIH ((http://nihroadmap.nih.gov/hmp/); "MetaHIT" stands for strains which were sequenced by the MetaHIT Project of EU (http://www.metahit.eu/); "GenBank" stands for strains which were collected in GenBank database (http://www.ncbi.nlm.nih.gov/).

Organism	**Source**
Acidaminococcus sp. D21	HMP
Actinomyces odontolyticus ATCC 17982	HMP
Alistipes putredinis DSM 17216	HMP
Anaerobaculum hydrogeniformans ATCC BAA-1850	HMP
Anaerococcus hydrogenalis DSM 7454	HMP
Anaerofustis stercorihominis DSM 17244	HMP
Anaerostipes caccae DSM 14662	HMP
Anaerotruncus colihominis DSM 17241	HMP
Bacteroides caccae ATCC 43185	HMP
Bacteroides capillosus ATCC 29799	HMP
Bacteroides cellulosilyticus DSM 14838	HMP
Bacteroides coprocola DSM 17136	HMP
Bacteroides coprophilus DSM 18228	HMP
Bacteroides dorei DSM 17855	HMP
Bacteroides eggerthii DSM 20697	HMP
Bacteroides finegoldii DSM 17565	HMP
Bacteroides fragilis 3_1_12	HMP
Bacteroides intestinalis DSM 17393	HMP
Bacteroides ovatus ATCC 8483	HMP
Bacteroides pectinophilus ATCC 43243	HMP
Bacteroides plebeius DSM 17135	HMP
Bacteroides sp. 1_1_6	HMP
Bacteroides sp. 2_1_7	HMP
Bacteroides sp. 2_2_4	HMP
Bacteroides sp. 3_2_5	HMP
Bacteroides sp. 4_3_47FAA	HMP
Bacteroides sp. 9_1_42FAA	HMP
Bacteroides sp. D1	HMP
Bacteroides sp. D2	HMP
Bacteroides sp. D4	HMP
Bacteroides stercoris ATCC 43183	HMP
Bacteroides uniformis ATCC 8492	HMP
Bifidobacterium adolescentis L2-32	HMP
Bifidobacterium angulatum DSM 20098	HMP
Bifidobacterium bifidum NCIMB 41171	HMP

Bifidobacterium breve DSM 20213	HMP
Bifidobacterium catenulatum DSM 16992	HMP
Bifidobacterium dentium ATCC 27678	HMP
Bifidobacterium gallicum DSM 20093	HMP
Bifidobacterium longum subsp. infantis ATCC 55813	HMP
Bifidobacterium longum subsp. infantis CCUG 52486	HMP
Bifidobacterium pseudocatenulatum DSM 20438	HMP
Blautia hansenii DSM 20583	HMP
Blautia hydrogenotrophica DSM 10507	HMP
Bryantella formatexigens DSM 14469	HMP
Butyrivibrio crossotus DSM 2876	HMP
Catenibacterium mitsuokai DSM 15897	HMP
Citrobacter sp. 30_2	HMP
Citrobacter youngae ATCC 29220	HMP
Clostridiales bacterium 1_7_47FAA	HMP
Clostridium asparagiforme DSM 15981	HMP
Clostridium bartlettii DSM 16795	HMP
Clostridium bolteae ATCC BAA-613	HMP
Clostridium hathewayi DSM 13479	HMP
Clostridium hiranonis DSM 13275	HMP
Clostridium hylemonae DSM 15053	HMP
Clostridium leptum DSM 753	HMP
Clostridium methylpentosum DSM 5476	HMP
Clostridium nexile DSM 1787	HMP
Clostridium ramosum DSM 1402	HMP
Clostridium scindens ATCC 35704	HMP
Clostridium sp. 7_2_43FAA	HMP
Clostridium sp. L2-50	HMP
Clostridium sp. M62/1	HMP
Clostridium sp. SS2/1	HMP
Clostridium spiroforme DSM 1552	HMP
Clostridium sporogenes ATCC 15579	HMP
Clostridium symbiosum ATCC 14940	HMP
Collinsella aerofaciens ATCC 25986	HMP
Collinsella intestinalis DSM 13280	HMP
Collinsella stercoris DSM 13279	HMP
Coprococcus comes ATCC 27758	HMP
Coprococcus eutactus ATCC 27759	HMP
Desulfovibrio piger ATCC 29098	HMP
Dorea formicigenerans ATCC 27755	HMP
Dorea longicatena DSM 13814	HMP

Enterobacter cancerogenus ATCC 35316	HMP
Enterococcus faecalis TX0104	HMP
Enterococcus faecalis TX1322	HMP
Enterococcus faecium TX1330	HMP
Escherichia coli SE11	HMP
Escherichia sp. 1_1_43	HMP
Escherichia sp. 3_2_53FAA	HMP
Escherichia sp. 4_1_40B	HMP
Eubacterium biforme DSM 3989	HMP
Eubacterium dolichum DSM 3991	HMP
Eubacterium hallii DSM 3353	HMP
Eubacterium siraeum DSM 15702	HMP
Eubacterium ventriosum ATCC 27560	HMP
Faecalibacterium prausnitzii A2-165	HMP
Faecalibacterium prausnitzii M21/2	HMP
Fusobacterium gonidiaformans ATCC 25563	HMP
Fusobacterium mortiferum ATCC 9817	HMP
Fusobacterium sp. 2_1_31	HMP
Fusobacterium sp. 4_1_13	HMP
Fusobacterium sp. 7_1	HMP
Fusobacterium sp. D11	HMP
Fusobacterium sp. D12	HMP
Fusobacterium ulcerans ATCC 49185	HMP
Fusobacterium varium ATCC 27725	HMP
Helicobacter bilis ATCC 43879	HMP
Helicobacter canadensis MIT 98-5491	HMP
Helicobacter cinaedi CCUG 18818	HMP
Helicobacter pullorum MIT 98-5489	HMP
Helicobacter winghamensis ATCC BAA-430	HMP
Holdemania filiformis DSM 12042	HMP
Lactobacillus acidophilus ATCC 4796	HMP
Lactobacillus antri DSM 16041	HMP
Lactobacillus brevis subsp. gravesensis ATCC 27305	HMP
Lactobacillus buchneri ATCC 11577	HMP
Lactobacillus fermentum ATCC 14931	HMP
Lactobacillus helveticus DSM 20075	HMP
Lactobacillus hilgardii ATCC 8290	HMP
Lactobacillus paracasei subsp. paracasei 8700:2	HMP
Lactobacillus paracasei subsp. paracasei ATCC 25302	HMP
Lactobacillus plantarum subsp. plantarum ATCC 14917	HMP
Lactobacillus reuteri CF48-3A	HMP

Lactobacillus reuteri JCM 1112	HMP
Lactobacillus reuteri MM2-3	HMP
Lactobacillus reuteri MM4-1A	HMP
Lactobacillus reuteri SD2112	HMP
Lactobacillus rhamnosus LMS2-1	HMP
Lactobacillus ruminis ATCC 25644	HMP
Lactobacillus salivarius ATCC 11741	HMP
Lactobacillus ultunensis DSM 16047	HMP
Leuconostoc mesenteroides subsp. cremoris ATCC 19254	HMP
Listeria grayi DSM 20601	HMP
Methanobrevibacter smithii DSM 2374	HMP
Methanobrevibacter smithii DSM 2375	HMP
Mitsuokella multacida DSM 20544	HMP
Mollicutes bacterium D7	HMP
Oxalobacter formigenes HOxBLS	HMP
Oxalobacter formigenes OXCC13	HMP
Parabacteroides johnsonii DSM 18315	HMP
Parabacteroides merdae ATCC 43184	HMP
Parvimonas micra ATCC 33270	HMP
Prevotella copri DSM 18205	HMP
Proteus penneri ATCC 35198	HMP
Providencia alcalifaciens DSM 30120	HMP
Providencia rettgeri DSM 1131	HMP
Providencia rustigianii DSM 4541	HMP
Providencia stuartii ATCC 25827	HMP
Roseburia intestinalis L1-82	HMP
Roseburia inulinivorans DSM 16841	HMP
Ruminococcus gnavus ATCC 29149	HMP
Ruminococcus lactaris ATCC 29176	HMP
Ruminococcus obeum ATCC 29174	HMP
Ruminococcus sp. 5_1_39BFAA	HMP
Ruminococcus torques ATCC 27756	HMP
Streptococcus infantarius subsp. infantarius ATCC BAA-102	HMP
Subdoligranulum variabile DSM 15176	HMP
Weissella paramesenteroides ATCC 33313	HMP
Bacteroides xylanisolvens XB1A	MetaHIT
Butyrivibrio fibrisolvens 16/4	MetaHIT
Coprococcus comes SL7/1	MetaHIT
Enterococcus sp. 7L76	MetaHIT
Eubacterium rectale DSM 17629	MetaHIT
Eubacterium rectale M104/1	MetaHIT

Eubacterium siraeum 70/3	MetaHIT
Faecalibacterium prausnitzii SL3/3	MetaHIT
Gordonibacter pamelaeae 7-10-1-b	MetaHIT
Megamonas hypermegale ART12/1	MetaHIT
Roseburia intestinalis M50/1	MetaHIT
Ruminococcus bromii L2-63	MetaHIT
Ruminococcus obeum A2-162	MetaHIT
Ruminococcus sp. 18P13	MetaHIT
Ruminococcus sp. SR1/5	MetaHIT
Ruminococcus torques L2-14	MetaHIT
unknown sp. SS3/4	MetaHIT
Bacteroides fragilis NCTC 9343	GenBank
Bacteroides fragilis YCH46	GenBank
Bacteroides thetaiotaomicron VPI-5482	GenBank
Bacteroides vulgatus ATCC 8482	GenBank
Bifidobacterium adolescentis ATCC 15703	GenBank
Bifidobacterium animalis subsp. lactis AD011	GenBank
Bifidobacterium longum DJO10A	GenBank
Bifidobacterium longum NCC2705	GenBank
Bifidobacterium longum subsp. infantis ATCC 15697	GenBank
Lactobacillus acidophilus NCFM	GenBank
Lactobacillus brevis ATCC 367	GenBank
Lactobacillus casei ATCC 334	GenBank
Lactobacillus casei BL23	GenBank
Lactobacillus delbrueckii subsp. bulgaricus ATCC 11842	GenBank
Lactobacillus delbrueckii subsp. bulgaricus ATCC BAA-365	GenBank
Lactobacillus fermentum IFO 3956	GenBank
Lactobacillus gasseri ATCC 33323	GenBank
Lactobacillus helveticus DPC 4571	GenBank
Lactobacillus johnsonii NCC 533	GenBank
Lactobacillus plantarum WCFS1	GenBank
Lactobacillus reuteri DSM 20016	GenBank
Lactobacillus sakei subsp. sakei 23K	GenBank
Lactobacillus salivarius UCC118	GenBank
Methanobrevibacter smithii ATCC 35061	GenBank
Parabacteroides distasonis ATCC 8503	GenBank

Table 6| ORF prediction in each sample. ORFs were predicted in each of the 124 samples. The ORFs with length shorter than 100 bp were disregarded.

Sample ID	# ORFs	Average length (bp)	# complete ORFs	% complete ORFs	# fragmental ORFs	% fragmental ORFs
MH0001	26,953	642	9,628	35.72	17,325	64.28
MH0002	125,165	613	43,569	34.81	81,596	65.19
MH0003	148,803	699	63,829	42.89	84,974	57.11
MH0004	42,407	656	14,401	33.96	28,006	66.04
MH0005	25,817	675	9,036	35.00	16,781	65.00
MH0006	286,373	666	111,390	38.90	174,983	61.10
MH0007	48,189	583	12,977	26.93	35,212	73.07
MH0008	50,568	646	17,961	35.52	32,607	64.48
MH0009	148,232	658	61,414	41.43	86,818	58.57
MH0010	46,365	695	16,148	34.83	30,217	65.17
MH0011	176,165	662	70,159	39.83	106,006	60.17
MH0012	305,090	679	130,932	42.92	174,158	57.08
MH0013	42,940	652	17,238	40.14	25,702	59.86
MH0014	148,583	706	67,551	45.46	81,032	54.54
MH0015	46,940	646	15,567	33.16	31,373	66.84
MH0016	140,136	708	61,436	43.84	78,700	56.16
MH0017	45,651	655	15,446	33.83	30,205	66.17
MH0018	23,198	664	7,842	33.80	15,356	66.20
MH0019	54,726	675	19,984	36.52	34,742	63.48
MH0020	113,474	749	57,530	50.70	55,944	49.30
MH0021	98,576	619	31,137	31.59	67,439	68.41
MH0022	40,092	648	13,638	34.02	26,454	65.98
MH0023	32,786	683	13,750	41.94	19,036	58.06
MH0024	76,889	610	23,900	31.08	52,989	68.92
MH0025	150,162	642	55,644	37.06	94,518	62.94
MH0026	82,651	731	41,390	50.08	41,261	49.92
MH0027	30,170	575	6,342	21.02	23,828	78.98
MH0028	131,576	694	56,516	42.95	75,060	57.05
MH0030	84,874	675	37,905	44.66	46,969	55.34
MH0031	102,769	638	38,113	37.09	64,656	62.91
MH0032	74,382	630	27,259	36.65	47,123	63.35
MH0033	131,550	671	55,397	42.11	76,153	57.89
MH0034	50,355	642	19,711	39.14	30,644	60.86
MH0035	136,131	659	51,188	37.60	84,943	62.40
MH0036	121,152	678	48,173	39.76	72,979	60.24
MH0037	103,128	670	42,830	41.53	60,298	58.47

MH0038	112,073	703	49,394	44.07	62,679	55.93
MH0039	136,058	663	51,819	38.09	84,239	61.91
MH0040	148,102	678	58,437	39.46	89,665	60.54
MH0041	116,738	683	48,757	41.77	67,981	58.23
MH0042	122,426	639	46,948	38.35	75,478	61.65
MH0043	148,817	635	58,060	39.01	90,757	60.99
MH0044	116,106	719	56,018	48.25	60,088	51.75
MH0045	126,089	654	45,612	36.17	80,477	63.83
MH0046	103,499	685	40,886	39.50	62,613	60.50
MH0047	62,694	608	18,171	28.98	44,523	71.02
MH0048	85,939	605	25,996	30.25	59,943	69.75
MH0049	60,622	686	24,732	40.80	35,890	59.20
MH0050	143,875	656	54,717	38.03	89,158	61.97
MH0051	88,456	639	28,642	32.38	59,814	67.62
MH0052	97,090	633	32,176	33.14	64,914	66.86
MH0053	111,733	676	47,518	42.53	64,215	57.47
MH0054	128,550	697	55,568	43.23	72,982	56.77
MH0055	131,643	685	55,739	42.34	75,904	57.66
MH0056	133,176	624	43,738	32.84	89,438	67.16
MH0057	88,567	697	37,964	42.86	50,603	57.14
MH0058	151,512	663	57,544	37.98	93,968	62.02
MH0059	160,668	671	63,106	39.28	97,562	60.72
MH0060	123,934	653	45,801	36.96	78,133	63.04
MH0061	88,369	712	42,577	48.18	45,792	51.82
MH0062	118,425	666	45,459	38.39	72,966	61.61
MH0063	113,469	640	45,151	39.79	68,318	60.21
MH0064	112,407	661	44,667	39.74	67,740	60.26
MH0065	140,752	647	55,470	39.41	85,282	60.59
MH0066	119,279	679	49,353	41.38	69,926	58.62
MH0067	109,111	761	57,553	52.75	51,558	47.25
MH0068	101,942	706	46,641	45.75	55,301	54.25
MH0069	159,490	640	60,084	37.67	99,406	62.33
MH0070	149,652	656	59,730	39.91	89,922	60.09
MH0071	116,441	681	50,735	43.57	65,706	56.43
MH0072	100,344	634	35,870	35.75	64,474	64.25
MH0073	113,509	666	43,560	38.38	69,949	61.62
MH0074	111,740	745	56,905	50.93	54,835	49.07
MH0075	136,521	663	52,612	38.54	83,909	61.46
MH0076	109,477	643	38,622	35.28	70,855	64.72
MH0077	151,491	620	53,714	35.46	97,777	64.54
MH0078	41,852	615	14,845	35.47	27,007	64.53
MH0079	91,926	603	31,462	34.23	60,464	65.77
MH0080	153,127	660	58,112	37.95	95,015	62.05

MH0081	147,079	642	55,132	37.48	91,947	62.52
MH0082	153,163	656	60,893	39.76	92,270	60.24
MH0083	139,100	674	56,663	40.74	82,437	59.26
MH0084	65,841	657	25,706	39.04	40,135	60.96
MH0085	103,668	721	45,313	43.71	58,355	56.29
MH0086	189,088	685	79,299	41.94	109,789	58.06
O2.UC-1	96,512	623	34,228	35.47	62,284	64.53
O2.UC-11	71,353	600	25,604	35.88	45,749	64.12
O2.UC-12	98,964	654	40,906	41.33	58,058	58.67
O2.UC-13	120,394	681	53,386	44.34	67,008	55.66
O2.UC-14	54,131	627	21,850	40.37	32,281	59.63
O2.UC-16	46,037	629	16,064	34.89	29,973	65.11
O2.UC-17	57,242	612	18,924	33.06	38,318	66.94
O2.UC-18	89,842	713	43,701	48.64	46,141	51.36
O2.UC-19	119,324	674	47,238	39.59	72,086	60.41
O2.UC-20	111,230	652	40,383	36.31	70,847	63.69
O2.UC-21	67,129	618	20,422	30.42	46,707	69.58
O2.UC-22	129,995	669	48,544	37.34	81,451	62.66
O2.UC-23	88,756	655	32,357	36.46	56,399	63.54
O2.UC-24	98,619	672	40,044	40.60	58,575	59.40
O2.UC-4	110,442	653	44,431	40.23	66,011	59.77
V1.CD-1	66,424	713	29,247	44.03	37,177	55.97
V1.CD-11	103,511	700	50,067	48.37	53,444	51.63
V1.CD-12	76,160	671	29,483	38.71	46,677	61.29
V1.CD-13	131,976	661	53,521	40.55	78,455	59.45
V1.CD-14	152,955	662	59,530	38.92	93,425	61.08
V1.CD-15	67,076	728	30,003	44.73	37,073	55.27
V1.CD-2	24,797	347	7,562	30.50	17,235	69.50
V1.CD-3	98,039	664	37,143	37.89	60,896	62.11
V1.CD-4	132,437	665	51,773	39.09	80,664	60.91
V1.CD-6	101,507	656	37,491	36.93	64,016	63.07
V1.CD-8	154,183	659	59,856	38.82	94,327	61.18
V1.CD-9	121,137	709	57,170	47.19	63,967	52.81
V1.UC-10	145,180	679	58,711	40.44	86,469	59.56
V1.UC-13	114,043	720	48,733	42.73	65,310	57.27
V1.UC-14	128,427	665	54,155	42.17	74,272	57.83
V1.UC-15	129,972	693	56,686	43.61	73,286	56.39
V1.UC-17	102,865	668	40,669	39.54	62,196	60.46
V1.UC-18	117,084	697	49,826	42.56	67,258	57.44
V1.UC-19	143,362	689	63,221	44.10	80,141	55.90
V1.UC-21	116,580	669	47,051	40.36	69,529	59.64
V1.UC-6	168,200	648	58,195	34.60	110,005	65.40
V1.UC-7	68,798	725	28,091	40.83	40,707	59.17

V1.UC-8	159,651	660	64,601	40.46	95,050	59.54
V1.UC-9	136,210	634	49,492	36.34	86,718	63.66
Mixed assembly	598,853	537	114,955	19.20	483,898	80.80
Sum	**14,048,045**	**661**	**5,465,716**	**38.91**	**8,582,329**	**61.09**

Table 7 | Non-redundant genes. Genes were compared at 95 % identity cut-off. Those that were overlapped over 90% length were considered redundant and removed. Common and rare genes were present in >50% and < 20% of individuals, respectively.

	# of genes	Total length (bp)	Mean length (bp)
Non-redundant gene set	3,299,822	2,323,171,095	704.03
Common	294,110	292,960,308	996.09
Rare	2,375,655	1,510,527,924	635.84

Table 8 | Common species in human gut. 75 non-redundant species with >1% genome coverage by Illumina reads in > 50 % of the cohort individuals.

Species	Number of individuals	
	Genome coverage >1%	Genome coverage >10%
Faecalibacterium prausnitzii SL3 3	124	121
Roseburia intestinalis M50 1	124	120
Dorea formicigenerans	123	118
Bacteroides vulgatus ATCC 8482	124	117
Clostridium sp SS2-1	121	117
Bacteroides uniformis	123	116
Eubacterium hallii	122	116
Bacteroides sp. 9_1_42FAA	124	115
unknown sp SS3 4	123	115
Coprococcus comes SL7 1	124	114
Eubacterium rectale M104 1	123	114
Ruminococcus sp SR1 5	124	114
Dorea longicatena	123	113
Bacteriodes xylanisolvens XB1A	124	111
Bacteroides sp. 2_1_7	124	111
Bacteroides sp. 2_2_4	124	110
Ruminococcus torques L2-14	124	110
Bacteroides sp. 4_3_47FAA	123	108
Alistipes putredinis	123	106
Collinsella aerofaciens	113	104
Parabacteroides distasonis ATCC 8503	123	103
Eubacterium siraeum 70 3	119	101
Bacteroides ovatus	123	99
Bacteroides sp. D4	124	99
Bacteroides sp. D1	123	98
Bacteroides thetaiotaomicron VPI-5482	123	98
Bacteroides dorei	124	95
Parabacteroides merdae	116	91
Bifidobacterium longum subsp. infantis CCUG 52486	103	88
Ruminococcus obeum A2-162	124	88
Bifidobacterium adolescentis	98	86
Bacteroides caccae	120	80
Ruminococcus bromii L2-63	119	79
Ruminococcus lactaris	124	77
Eubacterium ventriosum	119	73

Coprococcus eutactus	118	57
Akkermansia muciniphila ATCC BAA-835	82	56
Bacteroides stercoris	121	56
Bacteroides cellulosilyticus	96	51
Bacteroides fragilis 3_1_12	114	50
Eubacterium biforme	64	50
Bacteroides eggerthii	122	48
Streptococcus thermophilus LMD-9	114	48
Bacteroides capillosus	124	42
Holdemania filiformis	122	42
Clostridium leptum	119	41
Prevotella copri	69	41
Clostridium sp L2-50	123	40
Bacteroides plebeius	93	36
Butyrivibrio crossotus	123	36
Bacteroides coprocola	112	33
Bacteroides finegoldii	124	32
Clostridium bartlettii	83	32
Clostridium sp M62 1	124	31
Escherichia coli O157:H7 str. EC4115	68	31
Parabacteroides johnsonii	119	31
Subdoligranulum variabile	123	31
Bacteroides intestinalis	116	27
Catenibacterium mitsuokai	65	26
Clostridium bolteae	104	26
Bifidobacterium pseudocatenulatum	65	24
Anaerotruncus colihominis	123	20
Bifidobacterium catenulatum	75	18
Ruminococcus gnavus	123	18
Bacteroides coprophilus	111	15
Bacteroides pectinophilus	122	13
Gordonibacter pamelaeae gen nov sp Nov	89	13
Clostridium asparagiforme	118	10
Clostridium nexile	124	9
Blautia hansenii	116	5
Clostridium scindens	113	5
Enterococcus faecalis TX0104	114	3
Mollicutes bacterium D7	74	3
Bryantella formatexigens	78	1
Clostridium methylpentosum	103	1

Table 9 | Number of genes classified. The predicted genes were aligned to the known microbial genes, the genes in KEGG orthology database and in COG database. Blastp software was used to align genes with E-value <1E-5, and the best hit was selected. LCA-based algorithm was used to assign gene sequences to taxa. When a gene was conserved in many species, it was assigned to the lowest common ancestor (LCA). However, if the LCA is at phylum-level or below, it was considered to be "Classified" all the same. If not, it was treated as "Unclassified".

		Common genes	Rare genes	All genes
Total		**294,110**	**2,375,655**	**3,299,822**
Phylotype	Unknown	5.99%	27.64%	22.93%
	Unclassified	4.31%	3.82%	3.88%
	Classified	89.70%	68.54%	73.19%
eggNOG	Unannotated	23.88%	47.36%	42.46%
	Annotated	76.12%	52.64%	57.54%
COG	Unannotated	31.63%	54.89%	49.99%
	Annotated	68.37%	45.11%	50.01%
KEGG Orthology	Unannotated	34.73%	57.84%	52.97%
	Annotated	65.27%	42.16%	47.03%
KEGG Pathway	Unannotated	73.81%	83.19%	81.26%
	Annotated	26.19%	16.81%	18.74%

Table 10 | Range clusters. (TableS10_RangeClusters.xls)

Table 11 | Functions present in the human metagenome and genome. (TableS11_Genome&metagenome.xls)

Table 12 | 89 frequent microbial species/strains in human gut.

Species/strain	Source
Methanobrevibacter smithiiF1	HMP
Methanobrevibacter smithii ATCC 35061	GenBank
Methanobrevibacter smithii	HMP
Eubacterium hallii	HMP
Eubacterium siraeum	HMP
Eubacterium ventriosum	HMP
Mitsuokella multacida	HMP
Acidaminococcus sp. D21	HMP
Ruminococcus gnavus	HMP
Ruminococcus obeum	HMP
Ruminococcus bromii	MetaHIT
Ruminococcus torques	HMP
Faecalibacterium prausnitzii M21/2	HMP
Faecalibacterium prausnitzii	MetaHIT
butyrate-producing bacterium SR1/5	MetaHIT
Clostridium sp. SS2/1	HMP
Clostridium sp. M62/1	HMP
Roseburia intestinalis L1-82	HMP
Butyrivibrio crossotus	HMP
Coprococcus eutactus	HMP
Coprococcus comes	HMP
Dorea formicigenerans	HMP
Dorea longicatena	HMP
Enterococcus casseliflavus EC10	HMP
Enterococcus casseliflavus EC20	HMP
Enterococcus faecalis D6	HMP
Enterococcus gallinarum EG1	HMP
Enterococcus faecium Com15	HMP
Enterococcus faecium 1,230,933	HMP
Catenibacterium mitsuokai	HMP
Escherichia coli CFT073	GenBank
Escherichia coli S88	GenBank
Escherichia coli ED1a	GenBank
Escherichia coli HS	GenBank
Escherichia coli 536	GenBank
Escherichia coli IAI39	GenBank
Escherichia coli SMS-3-5	GenBank
Escherichia coli UTI89	GenBank

Escherichia coli APEC O1	GenBank
Escherichia coli O127:H6 str. E2348/69	GenBank
Escherichia coli ATCC 8739	GenBank
Escherichia sp. 3_2_53FAA	HMP
Klebsiella pneumoniae 342	GenBank
Klebsiella pneumoniae subsp. pneumoniae MGH 78578	GenBank
Citrobacter sp. 30_2	HMP
Desulfovibrio piger (ATCC 29098)	HMP
Akkermansia muciniphila ATCC BAA-835	GenBank
Alistipes putredinis	HMP
Bacteroides thetaiotaomicron VPI-5482	GenBank
Bacteroides intestinalis	HMP
Bacteroides sp. D1	HMP
Bacteroides sp. 1_1_6	HMP
Bacteroides ovatus	HMP
Bacteroides plebeius	HMP
Bacteroides sp. 4_3_47FAA	HMP
Bacteroides coprocola	HMP
Bacteroides pectinophilus	HMP
Bacteroides sp. 2_2_4	HMP
Bacteroides uniformis	HMP
Bacteroides coprophilus	HMP
Bacteroides sp. 2_1_7	HMP
Bacteroides cellulosilyticus	HMP
Bacteroides sp. 9_1_42FAA	HMP
Bacteroides fragilis NCTC 9343	GenBank
Bacteroides fragilis YCH46	GenBank
Bacteroides caccae	HMP
Bacteroides vulgatus ATCC 8482	GenBank
Bacteroides stercoris	HMP
Bacteroides sp. 3_2_5	HMP
Bacteroides dorei	HMP
Bacteroides sp. XB1A	MetaHIT
Bacteroides eggerthii	HMP
Bacteroides sp. D4	HMP
Parabacteroides distasonis ATCC 8503	GenBank
Parabacteroides merdae	HMP
Prevotella copri	HMP
Bifidobacterium bifidum NCIMB 41171	HMP
Bifidobacterium dentium	HMP
Bifidobacterium pseudocatenulatum	HMP
Bifidobacterium animalis subsp. lactis AD011	GenBank
Bifidobacterium catenulatum	HMP

Bifidobacterium breve	HMP
Bifidobacterium longum NCC2705	GenBank
Bifidobacterium longum bv. infantis CCUG 52486	HMP
Bifidobacterium longum subsp. infantis ATCC 15697	GenBank
Bifidobacterium longum DJO10A	GenBank
Bifidobacterium adolescentis ATCC 15703	GenBank
Bifidobacterium adolescentis L2-32	HMP
Collinsella aerofaciens	HMP