

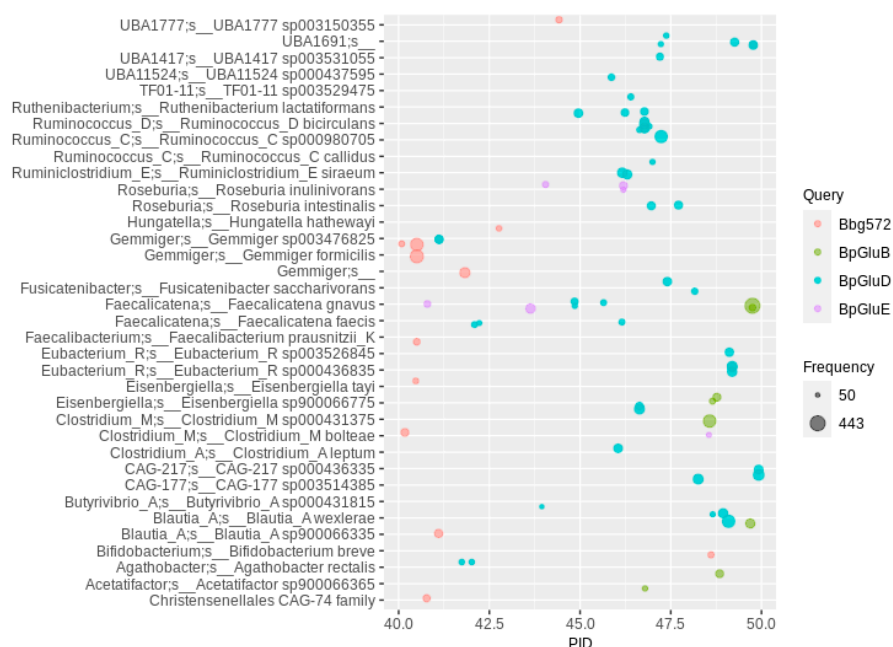
# Supplementary information to

## Flavonoid-modifying capabilities of the gut microbiome – an *in silico* study

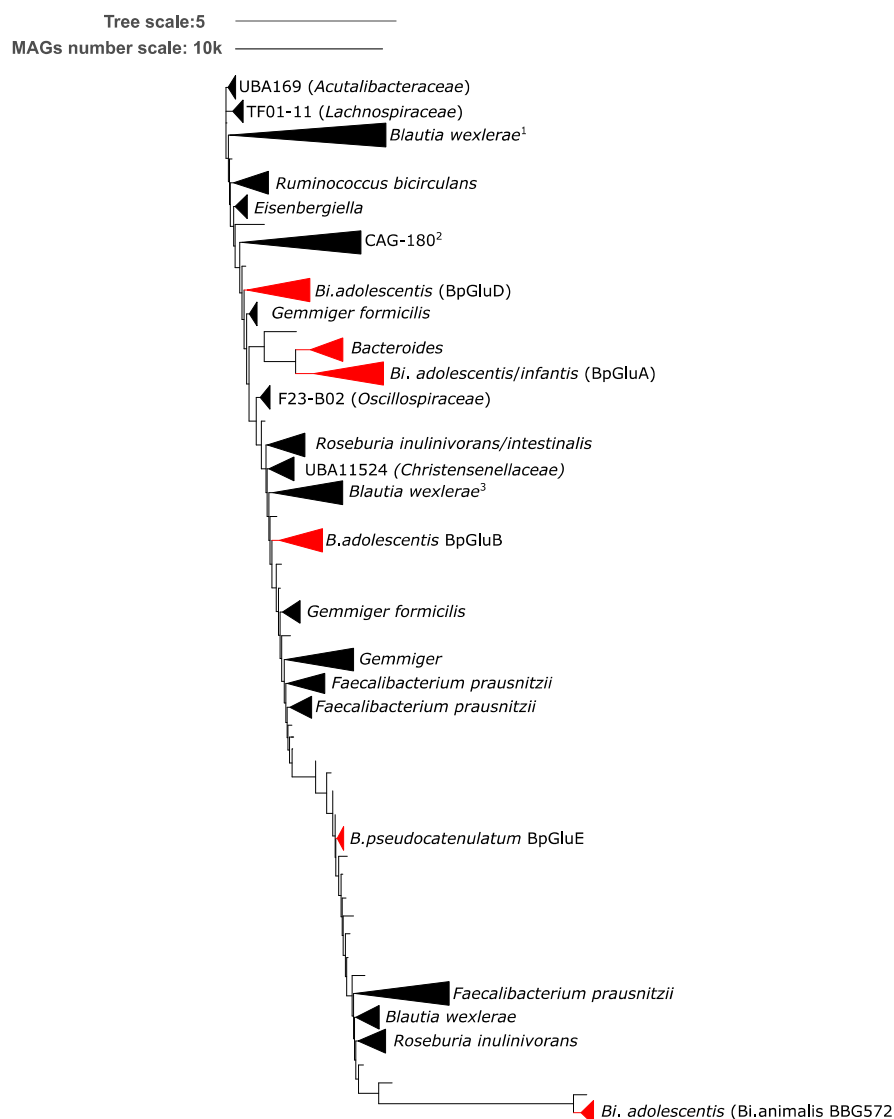
Tobias Goris<sup>1</sup>, Rafael Cuadrat<sup>2</sup>, Annett Braune<sup>1</sup>

<sup>1</sup>Department of Molecular Toxicology, Research Group Intestinal Microbiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Arthur-Scheunert-Allee 114-116, 14558 Nuthetal, Germany

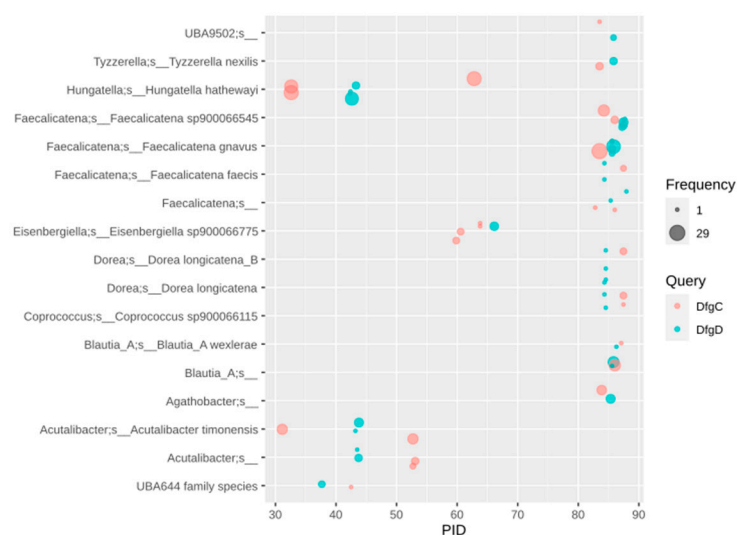
<sup>2</sup>Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Arthur-Scheunert-Allee 114-116, 14558 Nuthetal, Germany



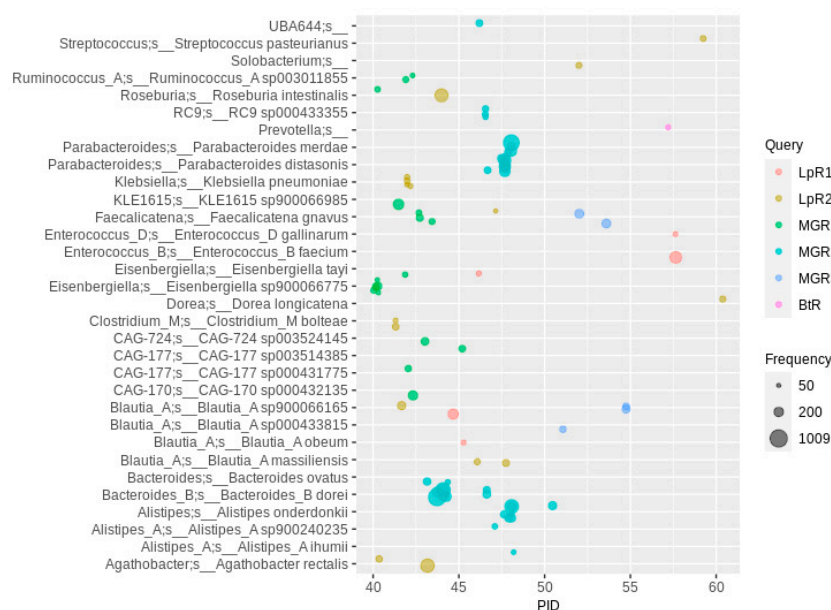
**Supplementary Figure S1:** Hits of human gut MAGs to flavonoid O-deglycosidases. A PID threshold of 40 to 50 was chosen to the queries shown in the color code (for abbreviations and details see Table 1). Hits were filtered for at least 50 occurrences, so that each bubble represents a number of redundant sequences ranging from 50 to 443.



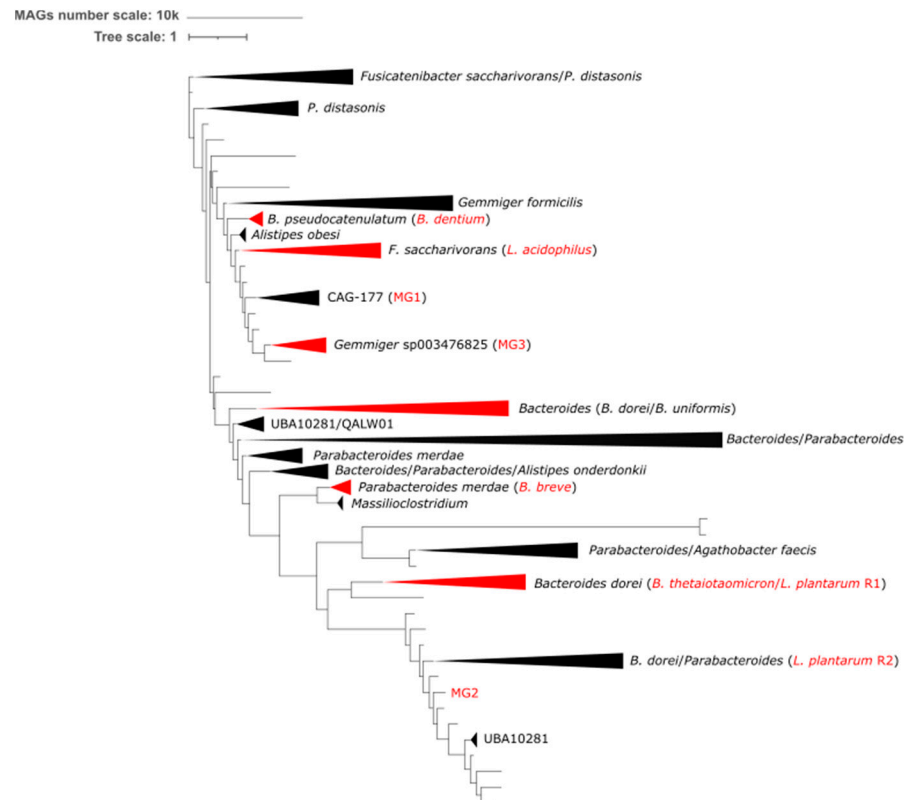
**Supplementary Figure S2:** Phylogenetic tree of O-deglycosidase sequences. Most abundant species genomes are indicated besides the triangles. Colored in red are nodes containing the six reference sequences of *Bifidobacteria* and *Bacteroides* species (in brackets, for abbreviations see Table 1). Only nodes with an overall number of at least 500 genomes (including isolate genomes) were included as triangles, the lengths of the triangles represent the number of genomes as indicated by the scale. <sup>1</sup>: Heterogenous group, also *Ruminoclostridium siraeum* and *Ruminococcus* species highly abundant. <sup>2</sup>: Heterogenous group, also high abundance of CAG-177 and *Eu-bacterium* genomes. <sup>3</sup>: Heterogenous group, also high abundance of *Faecalicatena gnavus* genomes.



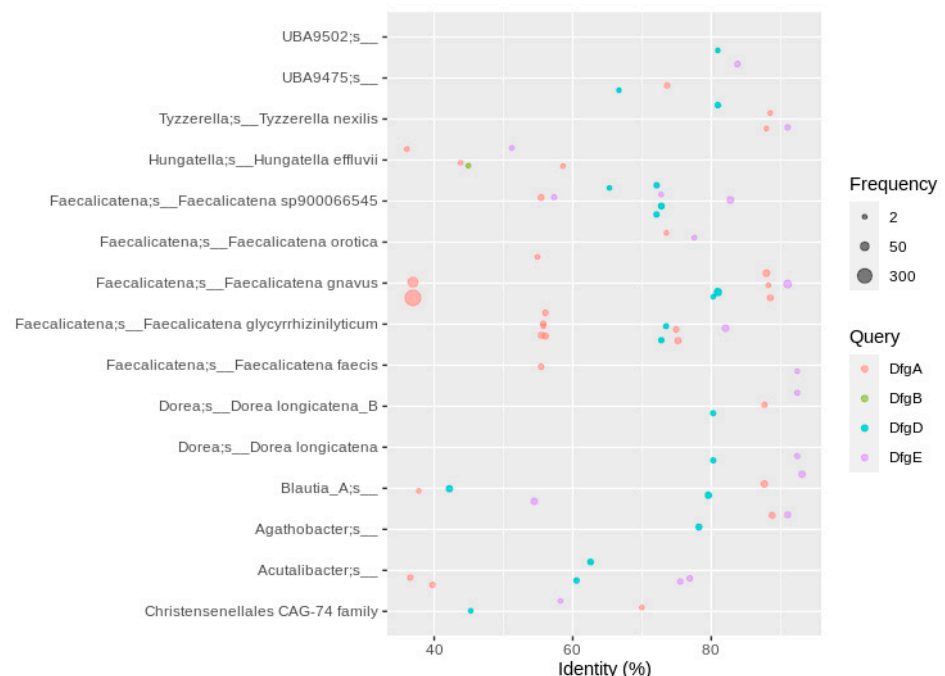
**Supplementary Figure S3:** Hits to sequences of the flavonoid O-deglycosidase DfgCD from *E. ramulus*. Hits were filtered for MAGs in which both genes were co-occurring.



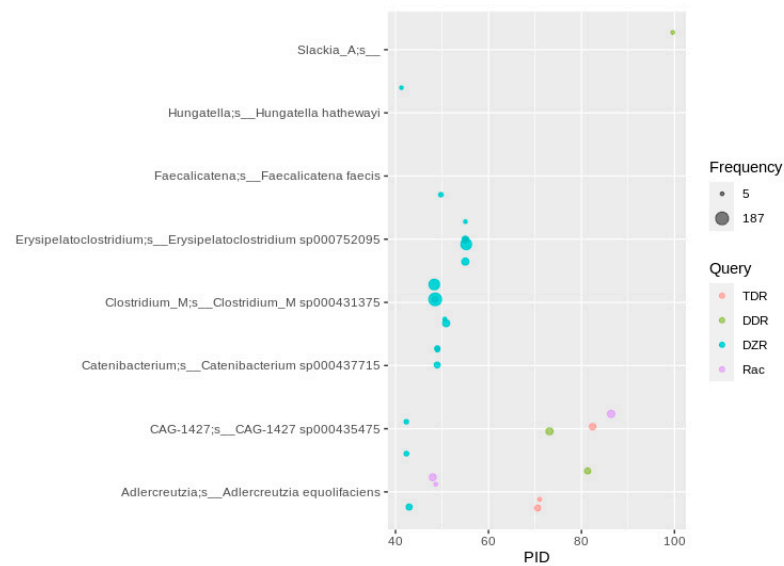
**Supplementary Figure S4:** Hits with 40 to 65 PID to flavonoid rhamnosidase sequences. Filtered for at least 50 occurrences per redundant sequence so that each bubble represents 50 to 1009 sequences.



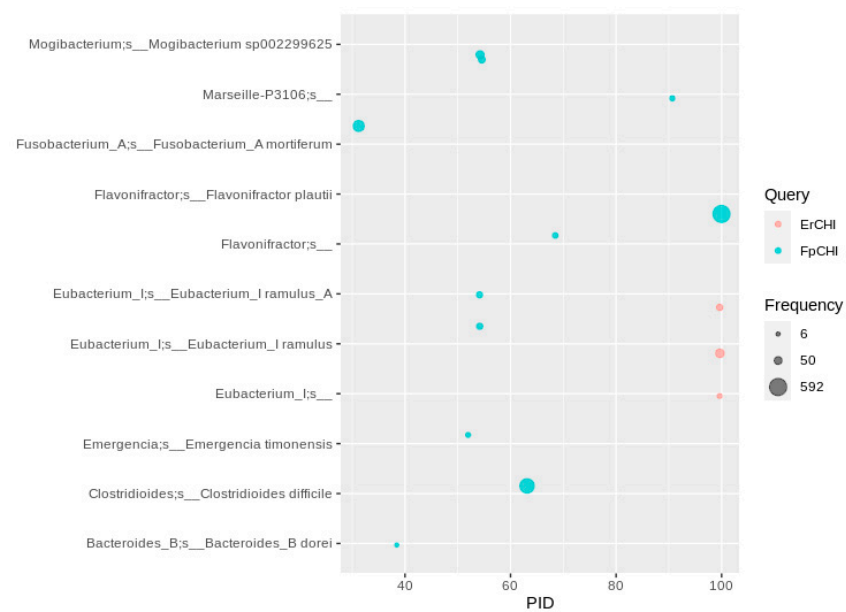
**Supplementary Figure S5:** Phylogenetic tree of rhamnosidase sequences. Most abundant species genomes are indicated besides the triangles. Colored in red are nodes containing the nine reference sequences (in brackets, for details see Table 1). Only nodes with an overall number of at least 500 genomes (including isolate genomes) were included as triangles, the lengths of the triangles represent the number of genomes as indicated by the scale.



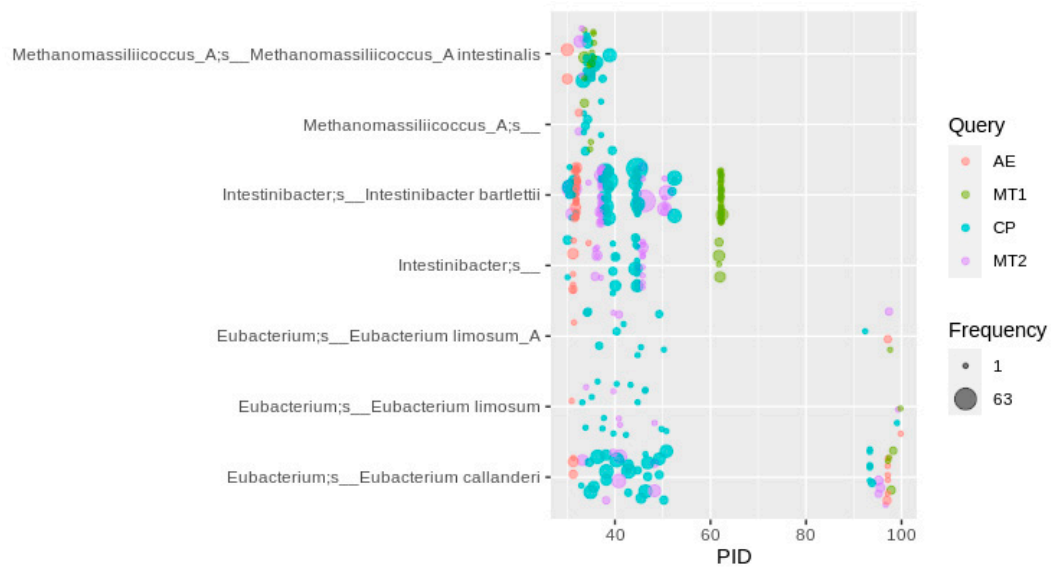
**Supplementary Figure S6:** Hits to the enzyme query sequences of the C-deglycosylation pathway of *E. cellulosolvens*. Filtered for the co-occurrence of at least three of the five enzymes and at least two identical/redundant sequences per bubble. There were no hits to DfgC.



**Supplementary Figure S7:** Hits to daidzein-to-equal enzyme query sequences. Filtered was only for a PID of at least 40 and at least five occurrences of sequences per bubble, not for co-occurrences of the genes in one genome. Note the presence of a number of DZR hits at about 50 PID.



**Supplementary Figure S8:** CHI-like enzymes in human gut bacteria. Hits were filtered for at least five occurrences, so that each bubble represents a number of redundant sequences ranging from 6 to 592.



**Supplementary Figure S9:** Putative *O*-demethylase systems in human gut bacteria. Hits were filtered for the co-occurrence of all four genes encoding the enzymes required for *O*-demethylation (MT1, MT2, AE, CP) in one MAG.



**Supplementary Figure S10:** *O*-Demethylase gene cluster of *Intestinibacter bartlettii*. The representative MAG genome MGYG-HGUT-00062 (NCBI accession number NZ\_CABIXZ010000001.1, GUT\_GENOME000245) was used. Selected locus tags are given below the genes encoding the corresponding enzymes of *O*-demethylation (for abbreviations see Table 1). Accessory proteins colored violet. CP2 with a lower PID than CP1 to the *E. limosum* CP in light blue. NCBI protein accession numbers of main proteins: MT1: WP\_147616460.1; CP1: WP\_007286937.1; AE: WP\_147616461.1; MT2 WP\_007286933.1.

**Supplementary Table S1:** Distribution of BLAST hits (PID > 40) to flavonoid-modifying enzymes across all MAGs of species discussed in the text. For the number of MAGs with hits to flavonoid-modifying enzyme sequences, we counted the MAGs (genomes assigned to isolates in the UHGG metadata were excluded) carrying at least one BLAST hit to the corresponding enzyme sequence. Each genome was counted only once to avoid double counts when different hits to different queries were present.

Species	Number of MAGs with hits	Overall number of species MAGs
<b>O-Deglycosylation</b>		
<i>Bifidobacterium adolescentis</i>	2229	2404
<i>Bifidobacterium pseudocatenulatum</i>	1016	1063
<i>Bifidobacterium infantis</i>	1708	2717
<i>Bacteroides thetaiotaomicron</i>	489	553
<i>Bacteroides ovatus</i>	474	1027
<i>Bacteroides cellulosilyticus</i>	420	452
<i>Bacteroides eggerthii</i>	198	409
<i>Faecalicatena gnavus</i>	1155	1197
Gemmiger sp003476825	3184	3216
CAG-180 sp000434995	2226	2419
CAG-217 sp000436335	1124	1404
<i>Blautia wexlerae</i>	1871	2179
<i>Eubacterium</i> sp000434995	584	683
<i>Faecalibacterium prausnitzii</i> K	1243	1380
<b>De-rhamnosylation</b>		
<i>Bacteroides dorei</i>	5246	5588
<i>Bacteroides thetaiotaomicron</i>	492	553
<i>Bifidobacterium pseudocatenulatum</i>	604	1063
Gemmiger sp003476825	2852	3216
<i>Alistipes onderdonkii</i> (unclassified Alistipes)	1800	1997
<i>Parabacteroides distasonis</i>	3782	4543
<i>Paracacteroides merdae</i>	1636	1723
<b>C-Deglycosylation</b>		
<i>Faecalicatena gnavus</i> ( <i>Ruminococcus gnavus</i> )	82	1197
<i>Agathobacter faecis</i> (unclassified <i>Roseburia</i> )	804	2819
<b>Dai-to-equol transformation</b>		
CAG1427 sp000435475	35	81
<i>Adlercreutzia equolifaciens</i>	33	271
<b>Phy</b>		
<i>Dialister succinatiphilus</i>	318	345
<i>Flavonifractor plautii</i>	563	610
<b>FLR</b>		
<i>Fusobacterium mortiferum</i>	165	270
<i>Flavonifractor plautii</i>	610	610
<i>Clostridoides difficile</i>	114	338
<b>Fcr</b>		
<i>Flavonifractor plautii</i>	509	610
<b>Chalcone isomerization/Taxifolin isomerization</b>		
<i>Flavonifractor plautii</i>	579	610
<i>Clostridoides difficile</i>	117	338
<b>O-Demethylation</b>		
<i>Intestinibacter bartlettii</i>	69	85

In brackets: Different taxonomy database entries (e.g. old/non-GTDB entries).

**Supplementary Table S2:** Potential flavonoid-modifying bacteria from a screening of the BRENDA enzyme database performed in a previous study [60].

Bacterium	Number of MAGs
<i>Pantoea agglomerans</i>	2
<i>Acinetobacter johnsonii</i>	1
<i>Pseudomonas putida</i>	1
<i>Desulfitobacterium hafniense</i>	1
<i>Leuconostoc mesenteroides</i>	5
<i>Lactobacillus pentosus</i>	14
<i>Bacillus licheniformis</i>	2
<i>Bacillus subtilis</i>	1
<i>Bacillus cereus</i>	4