

# Supplementary Information

## 1 Simulated Metagenome

Genome	taxonomy id	# reads
<i>Aliivibrio salmonicida</i> LFI1238	316275	1205
<i>Bdellovibrio bacteriovorus</i> HD100	264462	967
<i>Brucella melitensis</i> biovar Abortus 2308	359391	856
<i>Burkholderia mallei</i> SAVP1	320388	1297
<i>Burkholderia multivorans</i> ATCC 17616	395019	1789
<i>Chlamydia trachomatis</i> D/UW-3/CX	272561	253
<i>Clostridium phytofermentans</i> ISDg	357809	1276
<i>Colwellia psychrerythraea</i> 34H	167879	1377
<i>Cyanothece</i> sp. ATCC 51142	43989	1307
<i>Escherichia coli</i> B str. REL606	413997	1123
<i>Haemophilus parasuis</i> SH0165	557723	598
<i>Helicobacter pylori</i> B38	592205	404
<i>Mycobacterium abscessus</i> ATCC 19977	561007	1305
<i>Orientia tsutsugamushi</i> str. Ikeda	334380	499
<i>Pseudomonas aeruginosa</i> PA7	381754	1714
<i>Rhodopseudomonas palustris</i> BisB5	316057	1224
<i>Shigella boydii</i> CDC 3083-94	344609	1202
<i>Sinorhizobium meliloti</i> 1021	266834	973
<i>Staphylococcus aureus</i> subsp. aureus Mu50	158878	736
<i>Staphylococcus epidermidis</i> ATCC 12228	176280	684
<i>Streptococcus pneumoniae</i> G54	512566	534
<i>Sulfurovum</i> sp. NBC37-1	387093	681
<i>Synechococcus</i> sp. CC9605	110662	650
<i>Vibrio cholerae</i> M66-2	579112	1020
<i>Vibrio parahaemolyticus</i> RIMD 2210633	223926	1326
Total		25000

Supplementary Table 1: The 25 genomes from NCBI used to simulate the metagenomic reads. The number of reads refers to the simulated metagenome with average read length 265 bp.

## 2 Evaluation of CARMA3 on reads of different lengths

	order filtered		species filtered		all	
	TP	FP	TP	FP	TP	FP
superkingdom	14190	1175	21238	151	23010	19
phylum	9905	1153	20034	201	22919	28
class	4307	1289	16723	269	20826	30
order	–	2082	15851	246	21456	31
family	–	959	11696	213	18805	25
genus	–	151	7078	452	16357	99
species	–	2	–	49	417	17

Supplementary Table 2: Reads simulated using the MetaSim 454 error model with 400 bp read length.

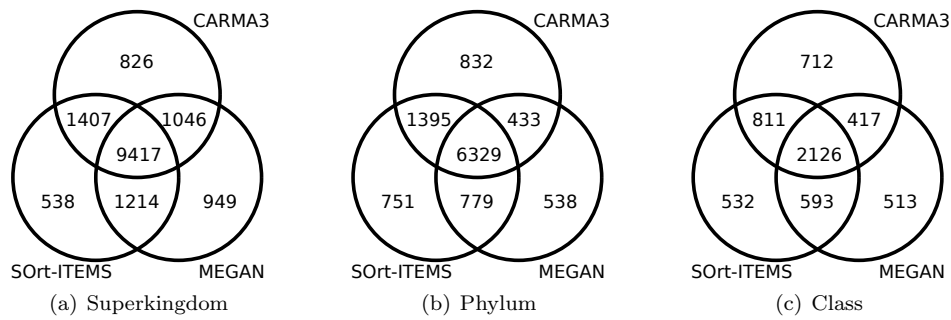
	order filtered		species filtered		all	
	TP	FP	TP	FP	TP	FP
superkingdom	3174	174	8558	39	12411	14
phylum	2208	531	8107	146	12218	41
class	884	564	6517	174	10745	47
order	–	1114	6469	251	11062	76
family	–	665	4916	299	9526	73
genus	–	190	2854	354	8597	90
species	–	42	–	485	2724	44

Supplementary Table 3: Reads simulated using the MetaSim 454 error model with 80 bp read length.

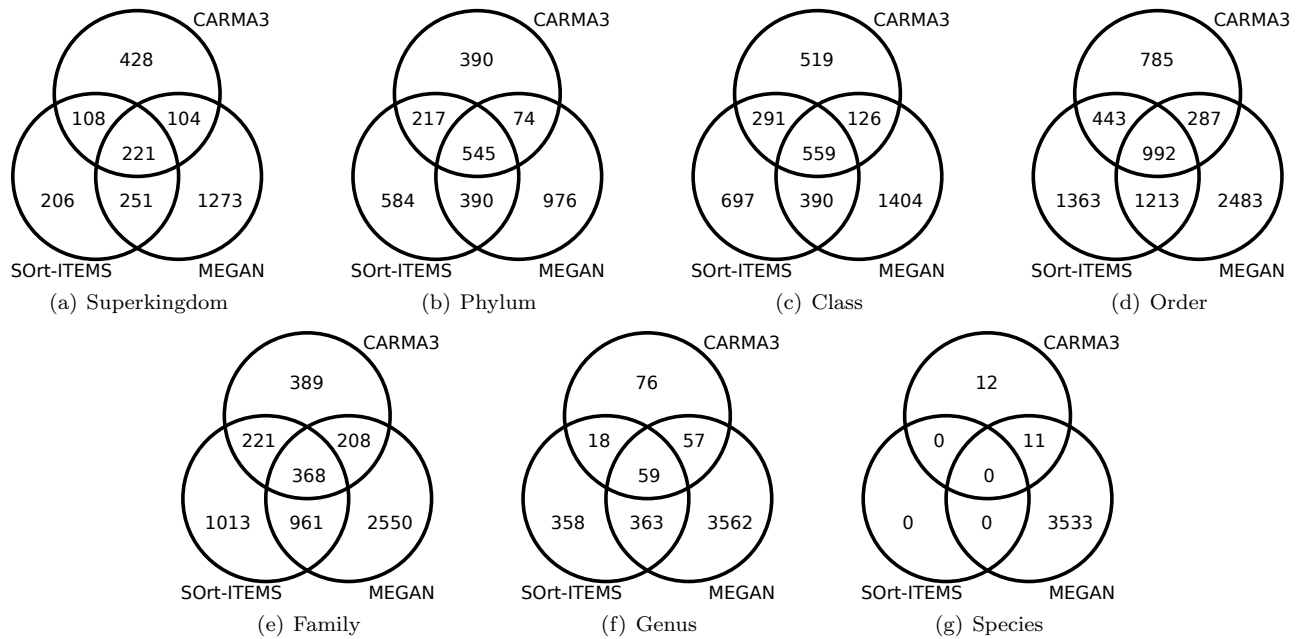
	order filtered		species filtered		all	
	TP	FP	TP	FP	TP	FP
superkingdom	7112	335	15692	34	20305	6
phylum	5139	904	15012	182	20157	22
class	2223	1021	12483	226	18331	24
order	–	2143	12291	342	18809	37
family	–	1320	9769	439	16736	38
genus	–	363	6420	672	15759	82
species	–	99	–	1190	6773	114

Supplementary Table 4: Reads simulated using the MetaSim Illumina error model with 80 bp read length.

### 3 Overlap of classifications for order-filtered dataset

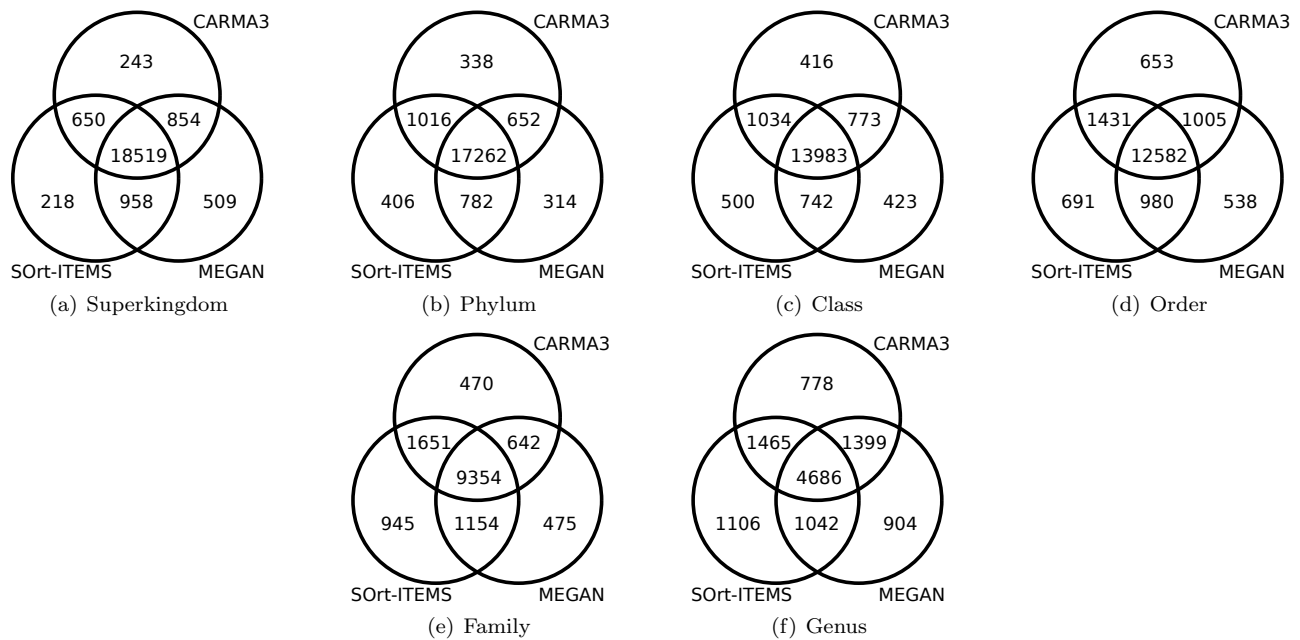


Supplementary Figure 1: True positives overlap of CARMA3, SOrt-ITEMS and MEGAN for the order-filtered dataset at taxonomic ranks superkingdom to class.

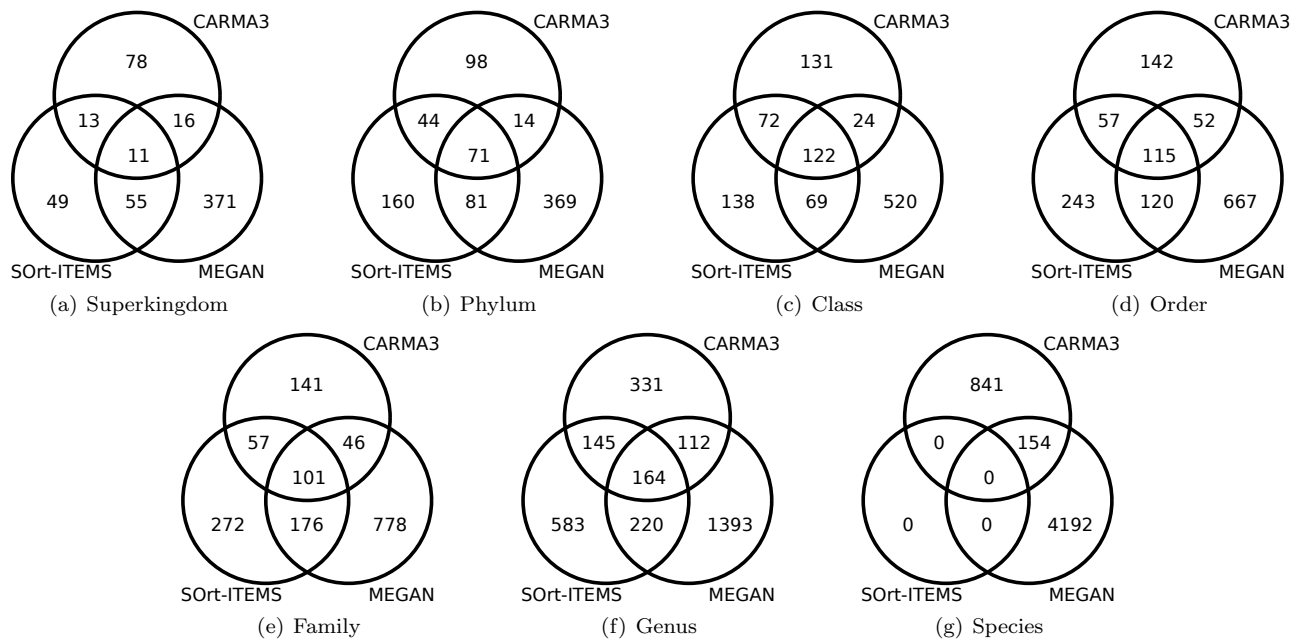


Supplementary Figure 2: False positives overlap of CARMA3, SOrt-ITEMS and MEGAN for the order-filtered dataset at taxonomic ranks superkingdom to species.

## 4 Overlap of classifications for species-filtered dataset

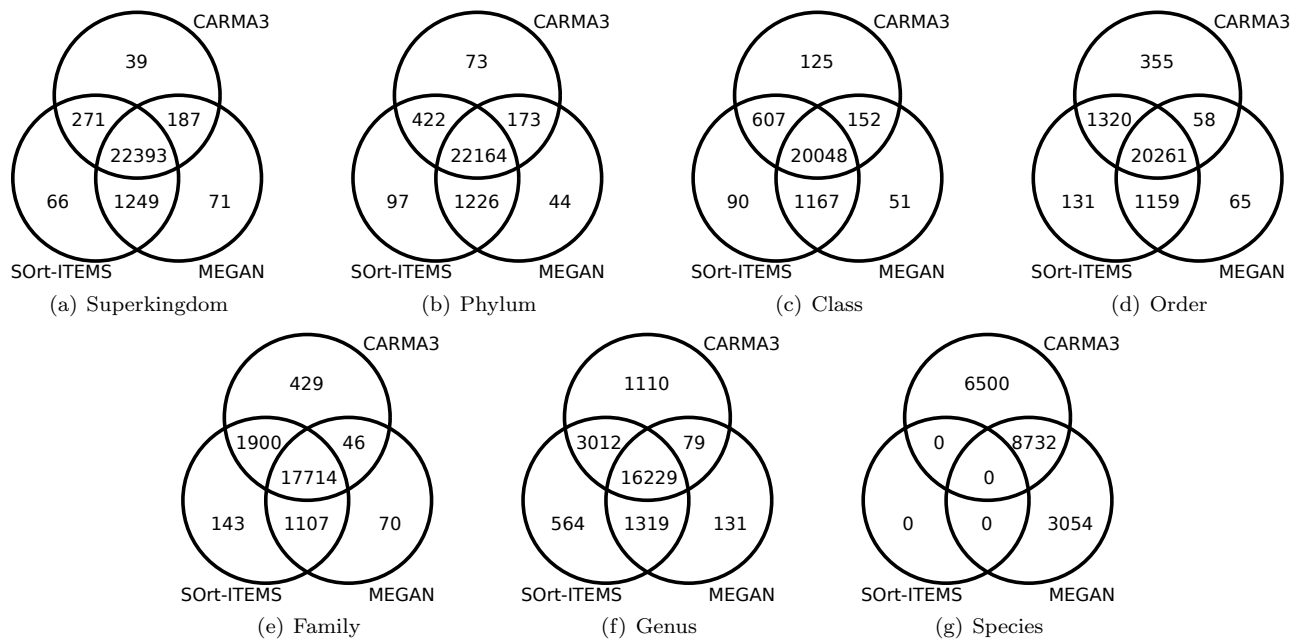


Supplementary Figure 3: True positives overlap of CARMA3, SOrt-ITEMS and MEGAN for the species-filtered dataset at taxonomic ranks superkingdom to genus.

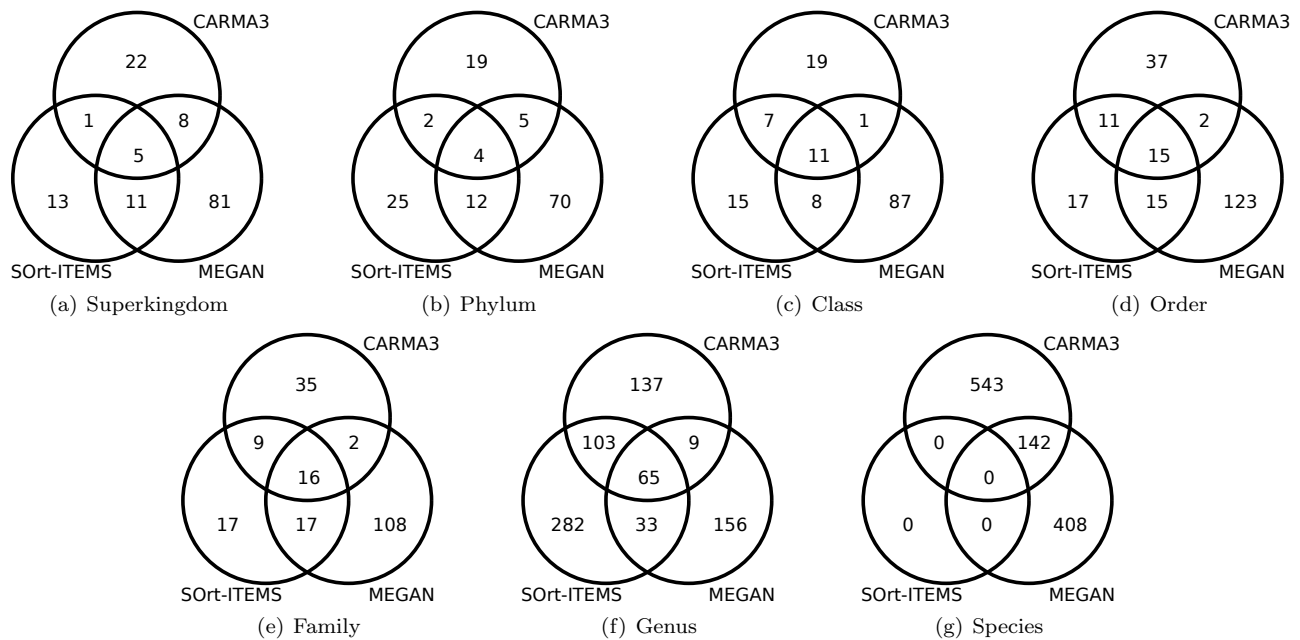


Supplementary Figure 4: False positives overlap of CARMA3, SOrt-ITEMS and MEGAN for the species-filtered dataset at taxonomic ranks superkingdom to species.

## 5 Overlap of classifications for unfiltered dataset

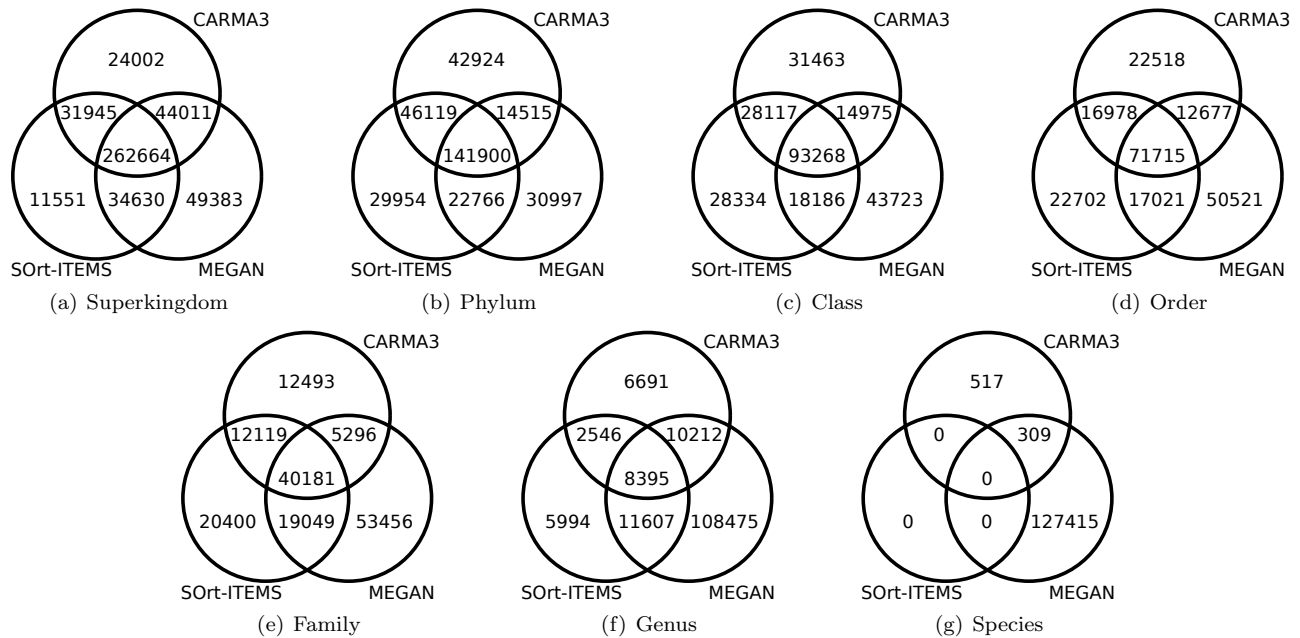


Supplementary Figure 5: True positives overlap of CARMA3, SOrt-ITEMS and MEGAN for the unfiltered dataset at taxonomic ranks superkingdom to species.



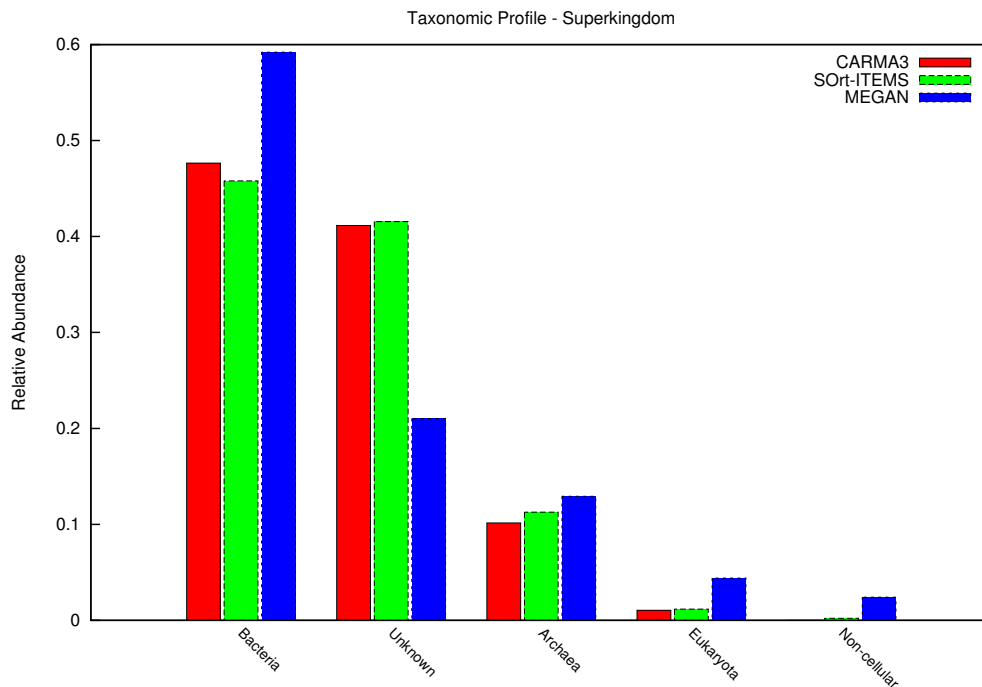
Supplementary Figure 6: False positives overlap of CARMA3, SOrt-ITEMS and MEGAN for the unfiltered dataset at taxonomic ranks superkingdom to species.

## 6 Biogas Plant Microbial Community – Overlap

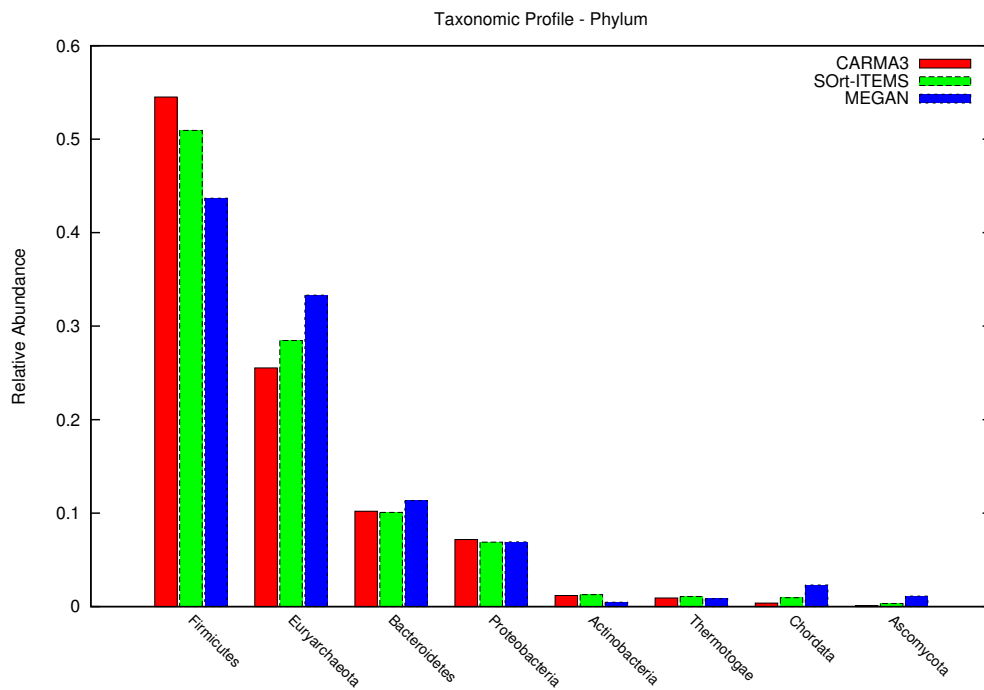


Supplementary Figure 7: Number of reads classified by each method at the corresponding taxonomic rank.

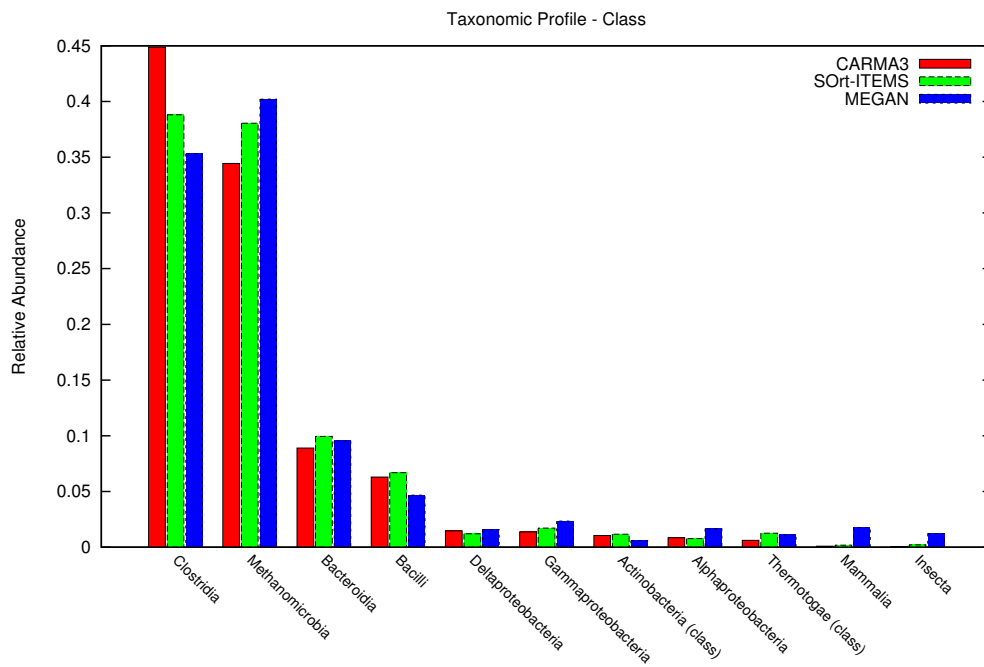
## 7 Biogas Plant Microbial Community – Comparative Taxonomic Profile



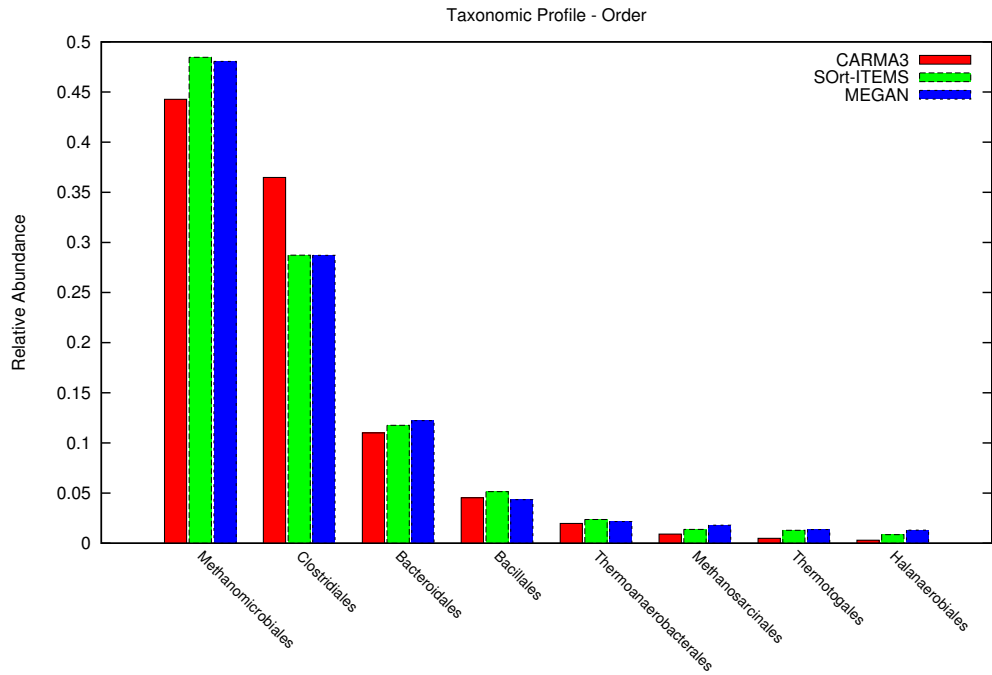
Supplementary Figure 8: Comparative taxonomic profile of a biogas plant microbial community at taxonomic rank superkingdom.



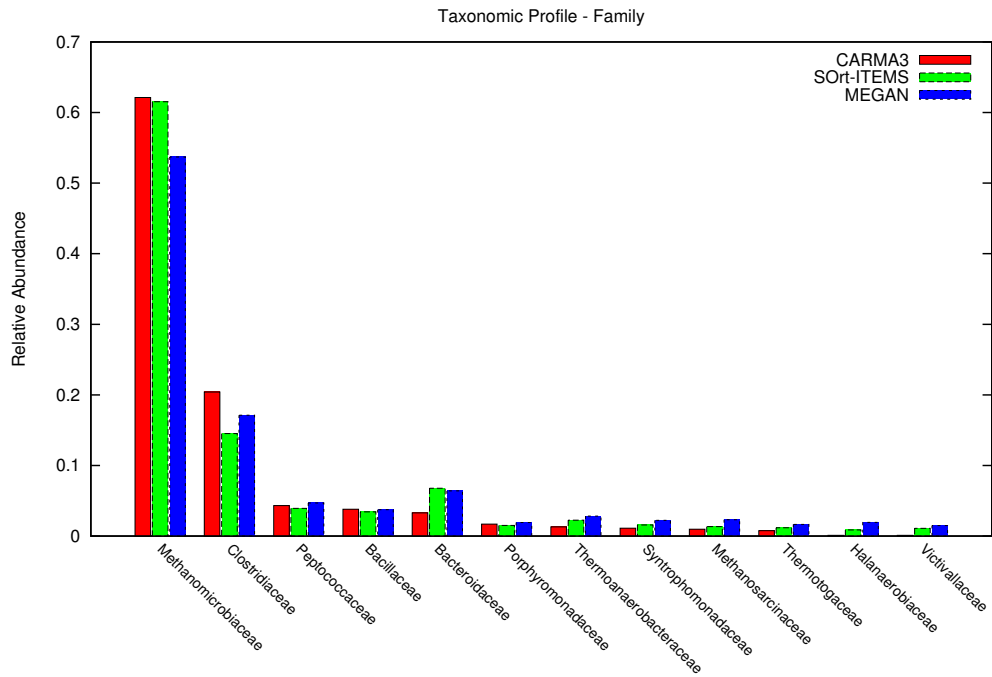
Supplementary Figure 9: Comparative taxonomic profile of a biogas plant microbial community at taxonomic rank phylum.



Supplementary Figure 10: Comparative taxonomic profile of a biogas plant microbial community at taxonomic rank class.

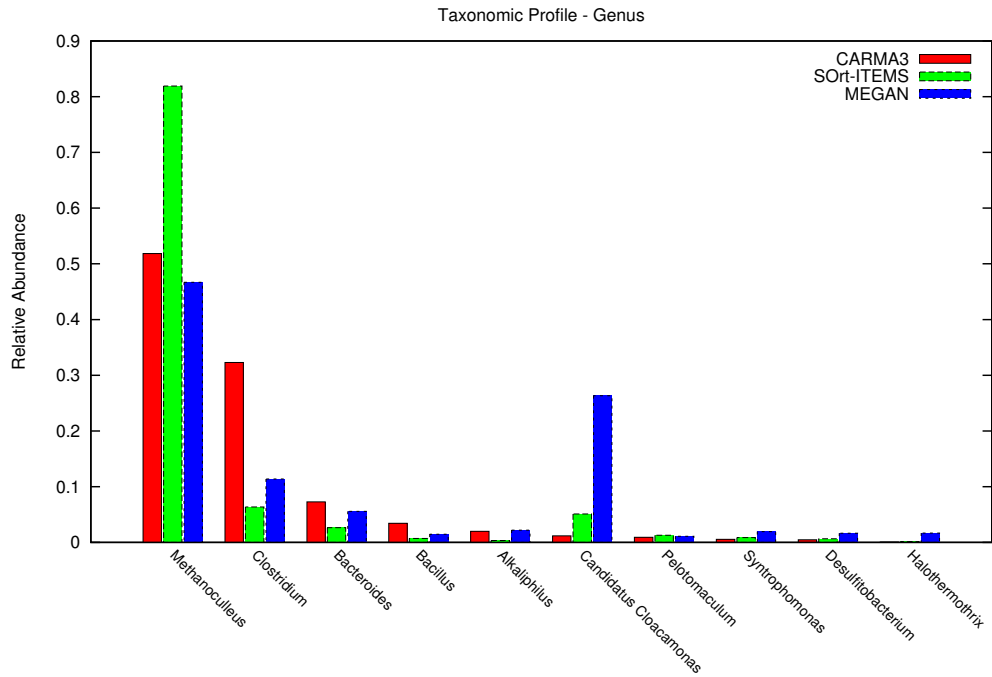


Supplementary Figure 11: Comparative taxonomic profile of a biogas plant microbial community at taxonomic rank order.

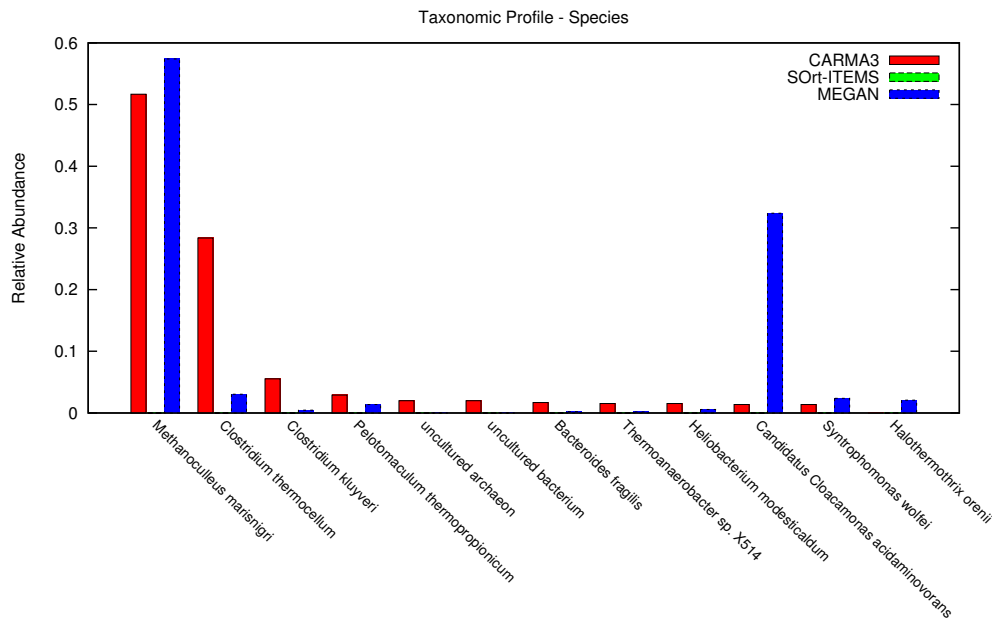


Supplementary Figure 12: Comparative taxonomic profile of a biogas plant microbial community at taxonomic rank family.





Supplementary Figure 13: Comparative taxonomic profile of a biogas plant microbial community at taxonomic rank genus.

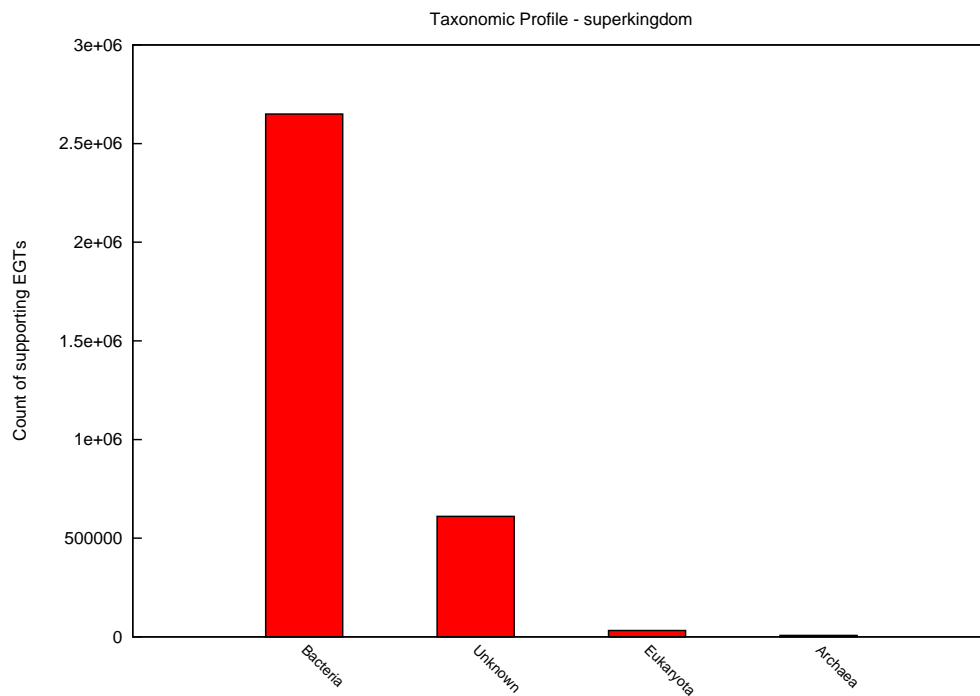


Supplementary Figure 14: Comparative taxonomic profile of a biogas plant microbial community at taxonomic rank species. Note that SOrt-ITEMS does not make predictions at taxonomic rank species.

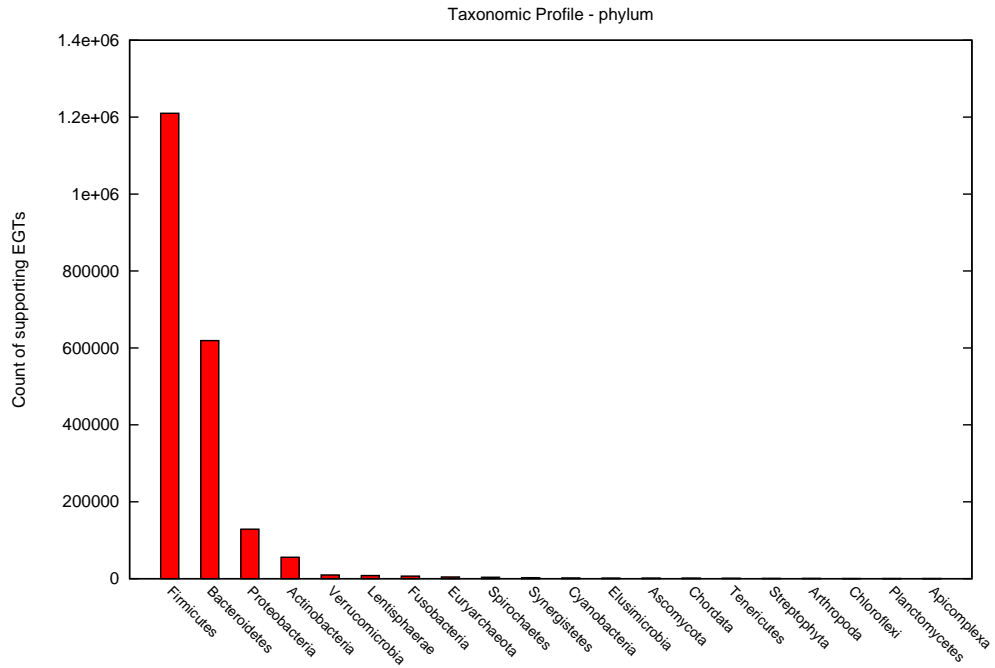
	CARMA3	SOrt-ITEMS	MEGAN
superkingdom	0	0	0
phylum	0.045	0.054	0.073
class	0.036	0.054	0.120
order	0.065	0.106	0.277
family	0.087	0.142	0.456
genus	0.091	0.090	0.495
species	0.267	–	0.689

Supplementary Table 5: Fraction of reads that have been discarded from the comparative taxonomic profile due to the cut-off threshold 0.01.

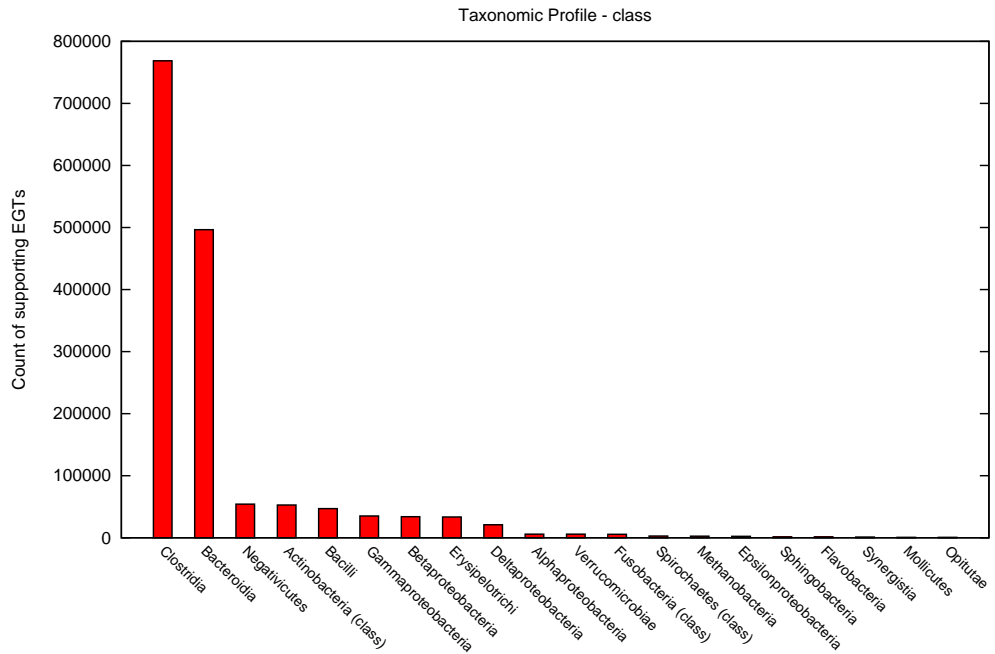
## 8 Taxonomic profile of the human gut microbial gene catalogue



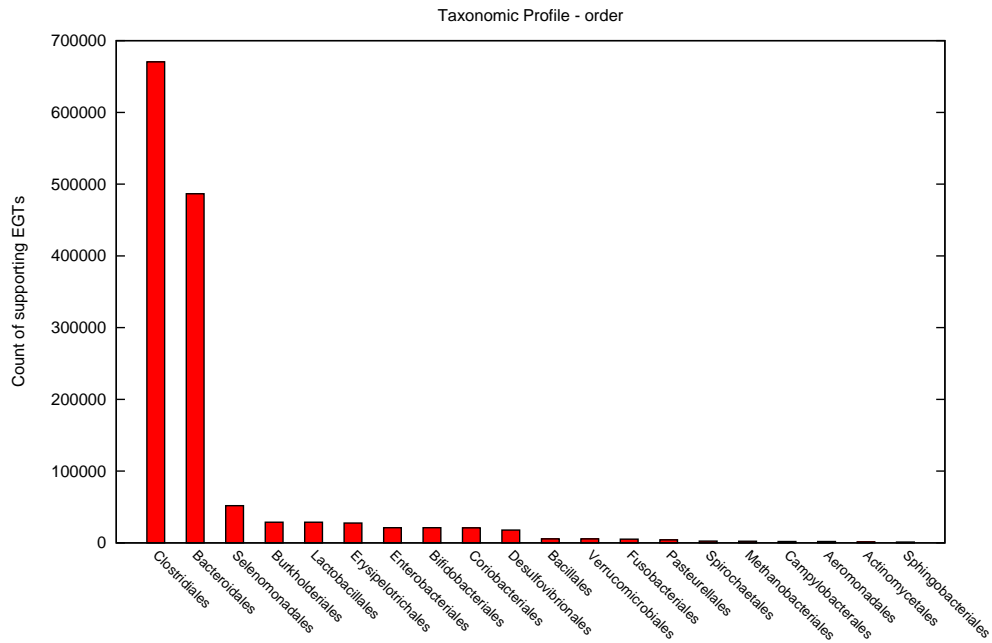
Supplementary Figure 15: All taxa of the human gut microbial gene catalogue at taxonomic rank superkingdom.



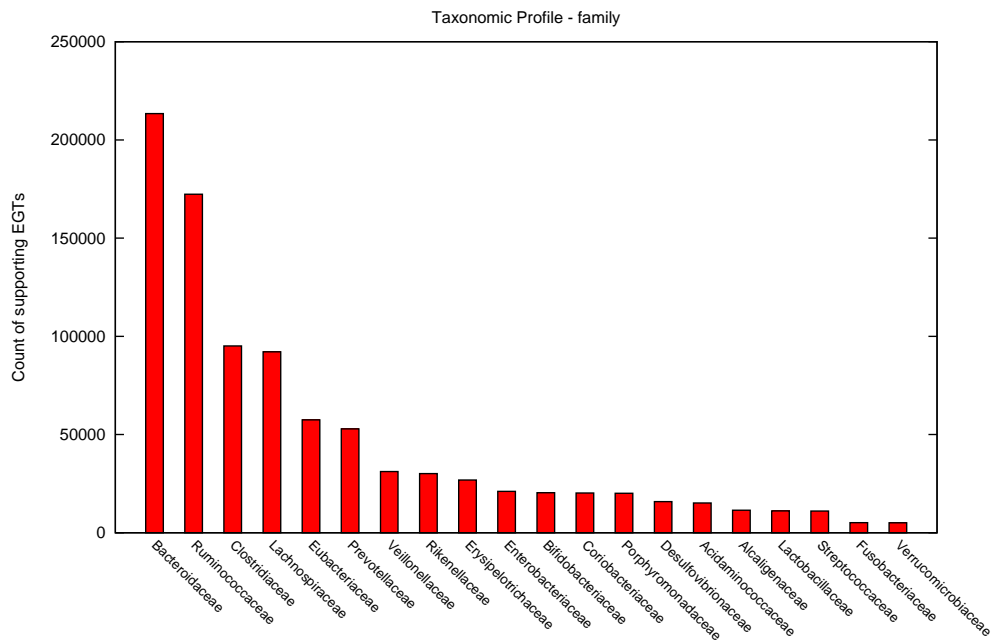
Supplementary Figure 16: The 20 most abundant taxa of the human gut microbial gene catalogue at taxonomic rank phylum.



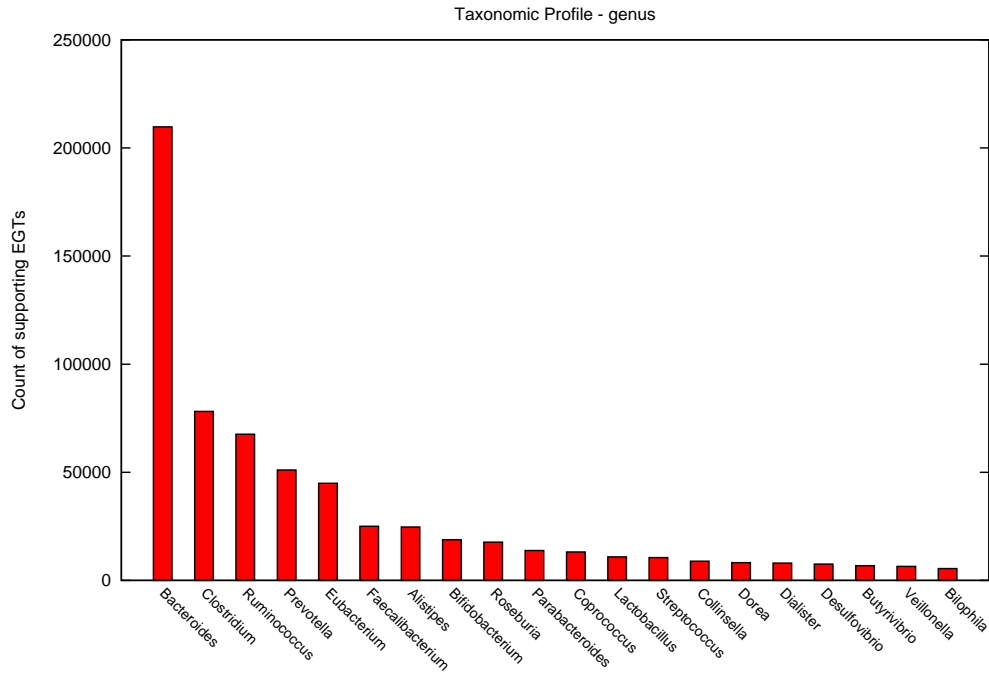
Supplementary Figure 17: The 20 most abundant taxa of the human gut microbial gene catalogue at taxonomic rank class.



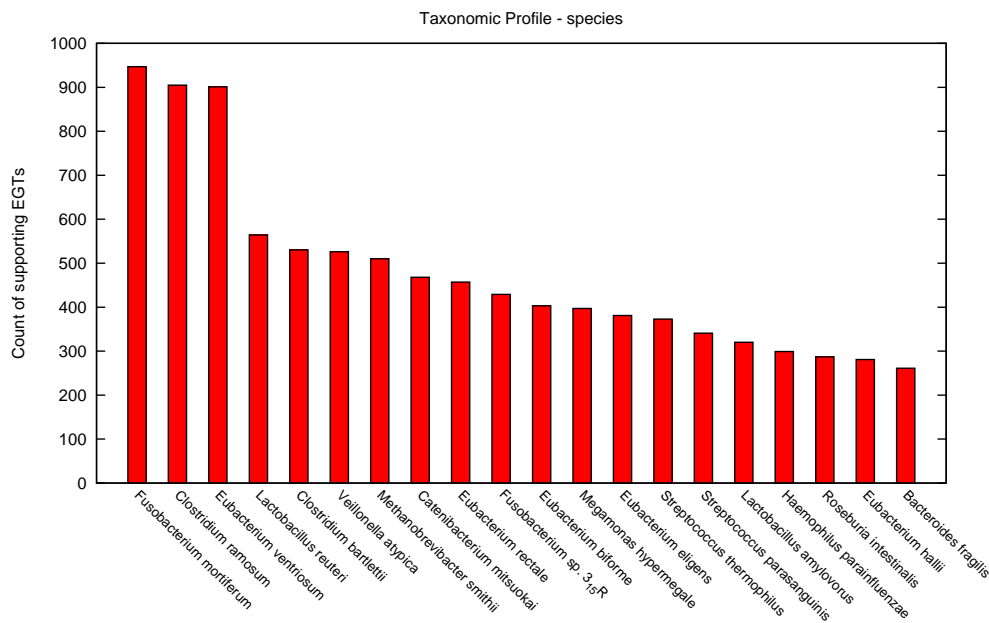
Supplementary Figure 18: The 20 most abundant taxa of the human gut microbial gene catalogue at taxonomic rank order.



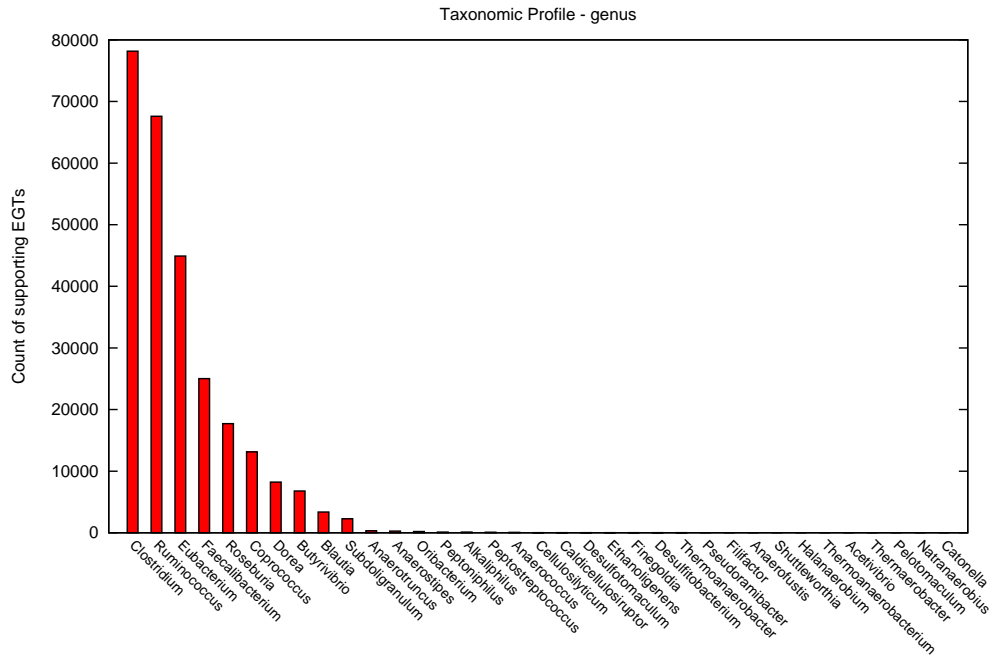
Supplementary Figure 19: The 20 most abundant taxa of the human gut microbial gene catalogue at taxonomic rank family.



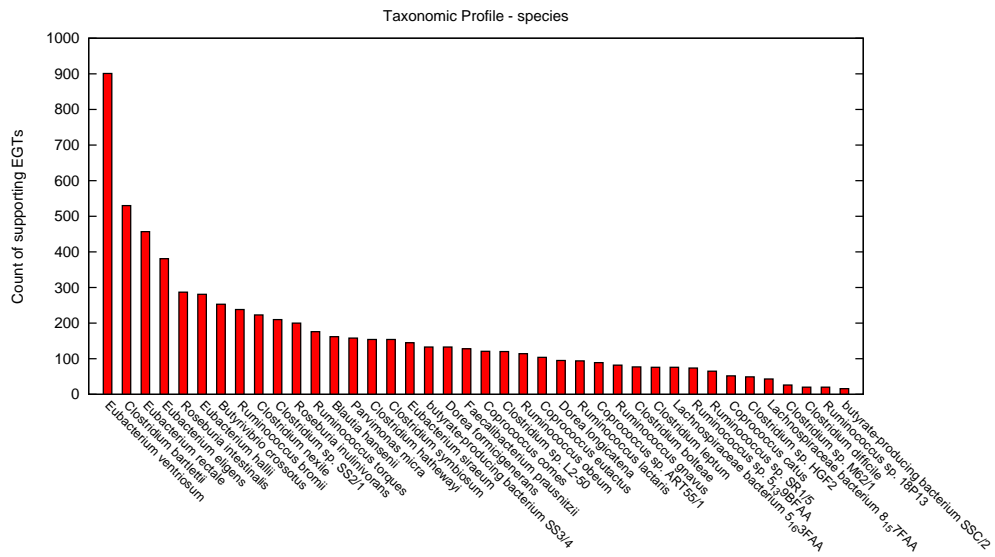
Supplementary Figure 20: The 20 most abundant taxa of the human gut microbial gene catalogue at taxonomic rank genus.



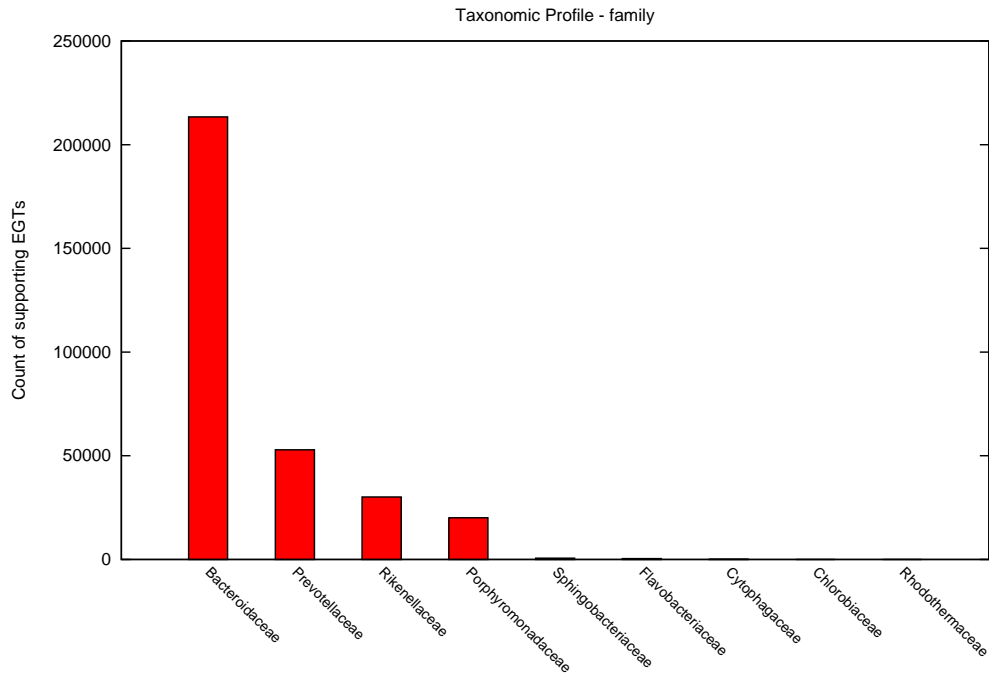
Supplementary Figure 21: The 20 most abundant taxa of the human gut microbial gene catalogue at taxonomic rank species.



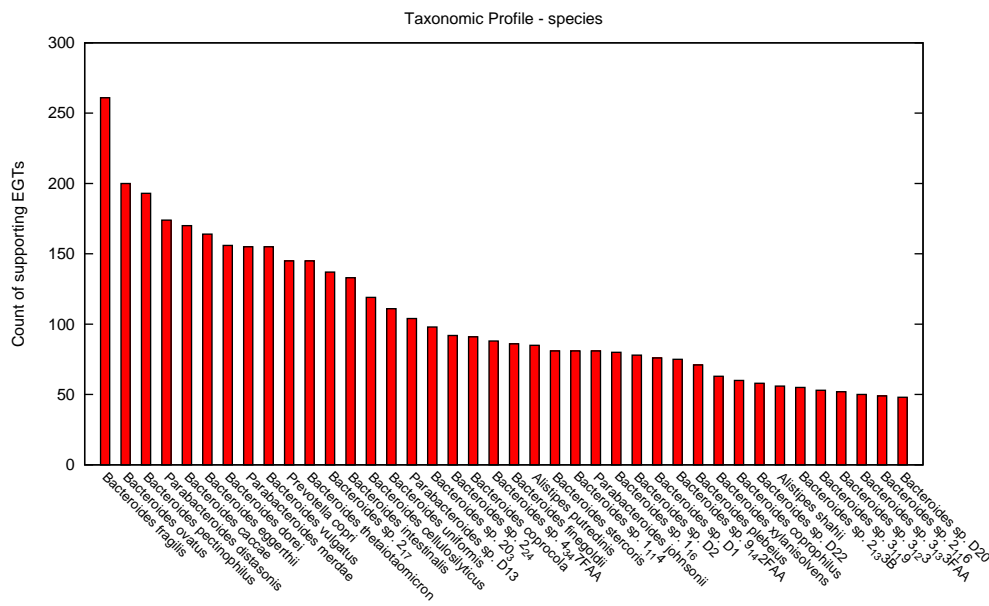
Supplementary Figure 22: All genera from the class *Clostridia* in the human gut microbial gene catalogue.



Supplementary Figure 23: The 40 most abundant species from the class *Clostridia* in the human gut microbial gene catalogue.



Supplementary Figure 24: All families from the phylum *Bacteroidetes* in the human gut microbial gene catalogue.



Supplementary Figure 25: The 40 most abundant species from the phylum *Bacteroidetes* in the human gut microbial gene catalogue.