

Table S2^a

Taxonomy	Illumina paired-end		Illumina single-end	
	Percentage	BLASTn result ^b	Percentage	BLASTn result
<i>Adlercreutzia</i>	NA ^c	NA	1 (0.02)	Exact
<i>Allobaculum</i>	2 (0.04) ^d	Exact	1 (0.04)	Exact
<i>Anaeroplasma</i>	1 (0.09)	1 off	1 (0.04)	1 off
<i>Bacteria</i>	1 (0.00)	NA	8 (0.55)	Exact
<i>Bacteroidales</i>	NA	NA	1 (0.02)	3 off
<i>Bacteroides</i>	1 (0.09)	Exact	1 (0.04)	Exact
<i>Candidatus Arthromitus</i>	1 (0.07)	Exact	1 (0.04)	Exact
<i>Clostridia</i>	NA	NA	1 (0.04)	3 off
<i>Clostridiales</i>	27 (3.85)	Exact	15 (1.73)	Exact
<i>Coprobacillus</i>	1 (0.04)	Exact	1 (0.02)	Exact
<i>Coprococcus</i>	3 (0.20)	1 off	1 (0.02)	Exact
<i>Coriobacteriaceae</i>	2 (0.15)	Exact	2 (0.04)	Exact
<i>Dehalobacterium</i>	1 (0.07)	Exact	1 (0.02)	Exact
<i>Desulfovibrio</i>	1 (0.16)	Exact	1 (0.09)	Exact
<i>Desulfovibrionaceae</i>	2 (0.13)	Exact	3 (0.18)	Exact
<i>Dorea</i>	1 (0.09)	Exact	1 (0.04)	Exact
<i>Erysipelotrichaceae</i>	1 (0.02)	Exact	1 (0.04)	Exact
<i>Firmicutes</i>	NA	NA	2 (0.18)	Exact
<i>Helicobacter</i>	2 (0.04)	Exact	1 (0.02)	Exact
<i>Helicobacteraceae</i>	1 (0.13)	Exact	1 (0.11)	Exact
<i>Lachnospiraceae</i>	17 (1.81)	Exact	24 (1.68)	Exact
<i>Lactobacillaceae</i>	1 (0.97)	Exact	1 (0.44)	Exact
<i>Lactobacillus</i>	4 (2.88)	Exact	4 (1.64)	Exact
<i>Mogibacteriaceae</i>	1 (0.02)	Exact	NA	NA
<i>Mucispirillum</i>	NA	NA	1 (0.02)	Exact
<i>Mycoplasmataceae</i>	2 (0.97)	6 off	1 (0.51)	3 off
<i>Oscillospira</i>	10 (2.97)	Exact	9 (1.86)	Exact
<i>Prevotella</i>	1 (0.02)	5 off	NA	NA
<i>Prevotellaceae</i>	NA	NA	1 (0.02)	Exact
<i>RF39</i>	2 (0.02)	Exact	1 (0.02)	Exact
<i>Rickettsiales</i>	1 (0.02)	2 off	NA	NA
<i>Rikenella</i>	1 (0.02)	4 off	NA	NA
<i>Rikenellaceae</i>	3 (0.20)	Exact	3 (0.11)	Exact
<i>Roseburia</i>	1 (0.02)	Exact	NA	NA
<i>Ruminococcaceae</i>	9 (0.89)	Exact	9 (0.69)	Exact
<i>Ruminococcus</i>	8 (0.62)	Exact	6 (0.22)	Exact

<i>S24-7</i>	4 (0.44)	Exact	4 (0.51)	Exact
<i>Sutterella</i>	1 (0.07)	Exact	1 (0.05)	Exact
<i>Turicibacter</i>	2 (2.44)	Exact	1 (1.26)	Exact
Sum	116 (19.53)		111 (12.27)	

^a Results were based on DADA2 analysis pipeline with Greengenes database rarefied at 5483 reads per sample.

^b Exact represents that there is 100% match between the ASV sequence and the sequences in the NCBI nr/nt database. Each number followed by “off” represents that there are the number of mismatched bases between the ASV sequences and the best matched sequences in the NCBI nr/nt database.

^c NA: Not available. The ASV does not exist for the particular Illumina MiSeq read analysis.

^d Each value represents the number of ASVs and mean percentage (in parenthesis) of that taxon from triplicate gDNA mock community samples. The sum reflects the % of all spurious reads and not just the most abundant ASVs represented in Table 5.