

Table S4. *In silico* simulated read results for significant single species identified in CCREPE analysis.

For simulated 16S rRNA gene sequences (V4 region; 0.5% error rate), we recorded how frequently (1) a sequence assignment included the correct species (allowing for ambiguous assignments), (2) unambiguous assignments to the correct species, and (iii) mis-assignments that did not include the correct species. We observe low mis-assignment rates for all species, as well as high unambiguous assignment rates with the exception of *Clostridium_perfringens* which maintained higher ambiguous assignments with *Clostridium_thermophilus* (e.g. *Clostridium_perfringens*:*Clostridium_thermophilus*).

Species	Total Simulated Sequences	Assignment includes correct species (%)	Unambiguously assigned to the correct species (%)	Misassigned at the species level (%)
<i>Blautia_faecis</i>	1,000	100.0%	100.0%	0.0%
<i>Blautia_glucerasea</i>	1,000	100.0%	100.0%	0.0%
<i>Blautia_luti</i>	1,000	100.0%	100.0%	0.0%
<i>Blautia_schinkii</i>	1,000	99.8%	99.8%	0.2%
<i>Blautia_wexlerae</i>	1,000	99.7%	99.5%	0.3%
<i>Clostridium_butyricum</i>	1,000	99.9%	99.9%	0.1%
<i>Clostridium_neonatale</i>	1,000	100.0%	100.0%	0.0%
<i>Clostridium_paraputrificum</i>	1,000	99.8%	99.8%	0.2%
<i>Clostridium_perfringens</i>	1,000	100.0%	2.6%	0.0%
<i>Clostridium_scindens</i>	1,000	99.9%	99.9%	0.1%
<i>Veillonella_dispar</i>	1,000	100.0%	100.0%	0.0%