***Supplementary methods***

*Cutadapt command line:*

do /home/local/.local/bin/cutadapt --discard-untrimmed -g CCTACGGGNGGCWGCAG -g GTCCTACGGGNGGCWGCAG -g AGAGCCTACGGGNGGCWGCAG -g TAGTGTCCTACGGGNGGCWGCAG -G GACTACHVGGGTATCTAATCC -G AGACTACHVGGGTATCTAATCC -G TCTGACTACHVGGGTATCTAATCC -G CTGAGTGGACTACHVGGGTATCTAATCC -o $(echo $line)\_L001\_R1\_noCOtrim.fastq -p $(echo $line)\_L001\_R2\_noCOtrim.fastq $(echo $line)\_L001\_R1\_001.fastq $(echo $line)\_L001\_R2\_001.fastq > $(echo $line)\_output.txt; done < yourfilelist.txt

*Mothur SOP pipeline modification:*

The following changes were made to the SOP parameters: maxlength = 630 in the first screen.seqs step; reference=silva.seed\_v132.align for the align.seqs step; start = 6428 and end = 23440 in the second screen.seqs step; diffs = 4 in pre.cluster.

Additionally, singleton sequences were removed with split.abunds (cutoff = 1) before running classify.seqs. The references used were the Recreated SEED database SILVA alignment release 132 for align.seqs and the RDP 16S rRNA reference (PDS) version 16 for classify.seqs.