Raw sequence were download from …

Sequence reads of 4x (20 men and 20 women) stool samples were downloaded from the Human Microbiome Project (HMP).

Megahit (v1.2.9) was used for the assembly for the sequences from all the samples (cross-assembly) with min length of contigs > 200. CD-hit (4.8.1) was used to get the no-redundant sequences.

Prokka (1.14.6) was used for the prediction of ORFs and protein and nucleotide sequence were obtained based on the cross-assembly generated contig sequeneces.

HMMER (hmmsearch, v 3.3.2) software was used for searching the dmd\_tmd, dmm, mauA, mauB, tdm, tmm sequence in the protein sequence obtained from Prokka. Protein sequence with E-value less than xx were obtained as the sequence for each of the genes, respectively.

Taxmony annotation for prokka sequence was conducted by using Kraken2 (2.1.2).

Databases for antibiotic resistance were downloaded from the The Comprehensive Antibiotic Resistance Database (<https://card.mcmaster.ca/home>). rgi (5.1.0) software was used to annotate Prokka predicted sequences for antibiotic resistance profile.

The sequence for virulence factor was downloaded from The virulence factor database (http://www.mgc.ac.cn/VFs/main.htm). Functional annotation based on virulence factor sequence for Prokka predicted sequences was conducted using Diamond (0.8.36).

Bowtie2 (2.4.2) was used for the mapping of the raw sequence to the prokka predicted DNA sequences. The sequence read counts for each of the annotations (6 genes, Taxmony, antibiotic resistance, virulence factor) was obtained by summing the reads number that belong to the corresponding domain based on each annotation method.

Individual reads number belong to each prokka contig sequences were obtained by samtools (1.12) idxstats.