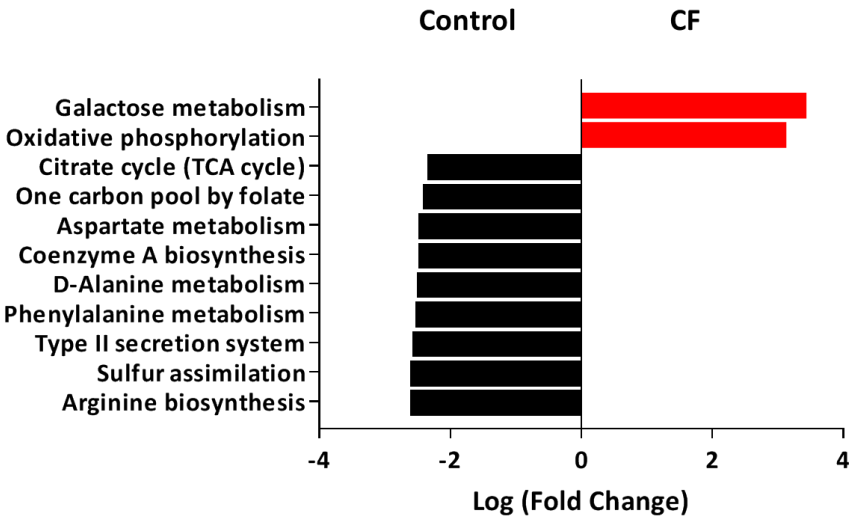
Supplementary Material

Opportunistic bacteria confer the ability to ferment prebiotic starch in the adult cystic fibrosis gut

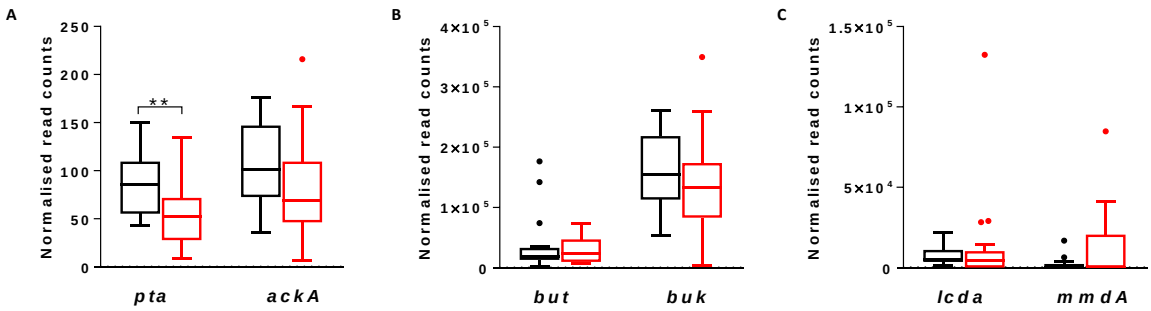
Yanan Wang, Lex E.X. Leong, Rebecca L. Keating, Tokuwa Kanno, Guy C.J. Abell, Fredrick M. Mobegi, Jocelyn M. Choo, Steve L. Wesselingh, A. James Mason, Lucy D. Burr, Geraint B. Rogers

Supplementary Figure 1-8

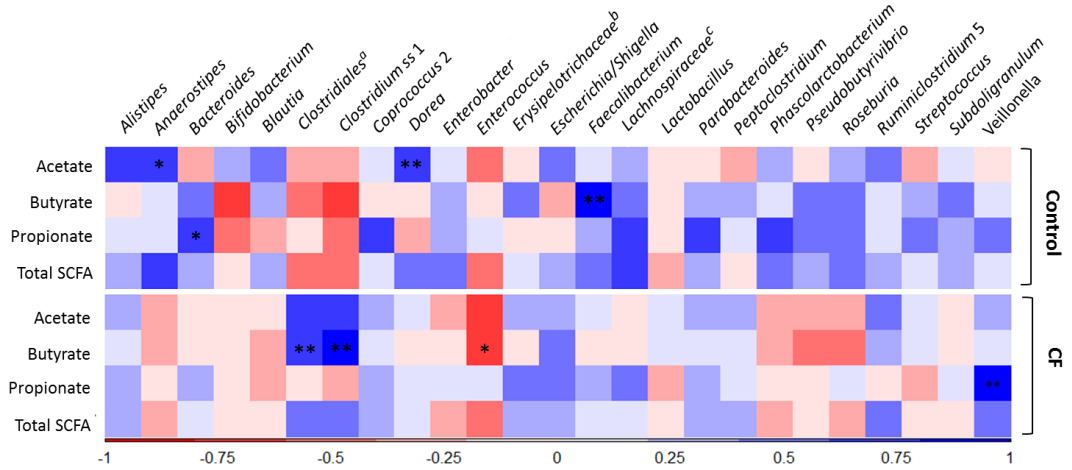
Supplementary Methods



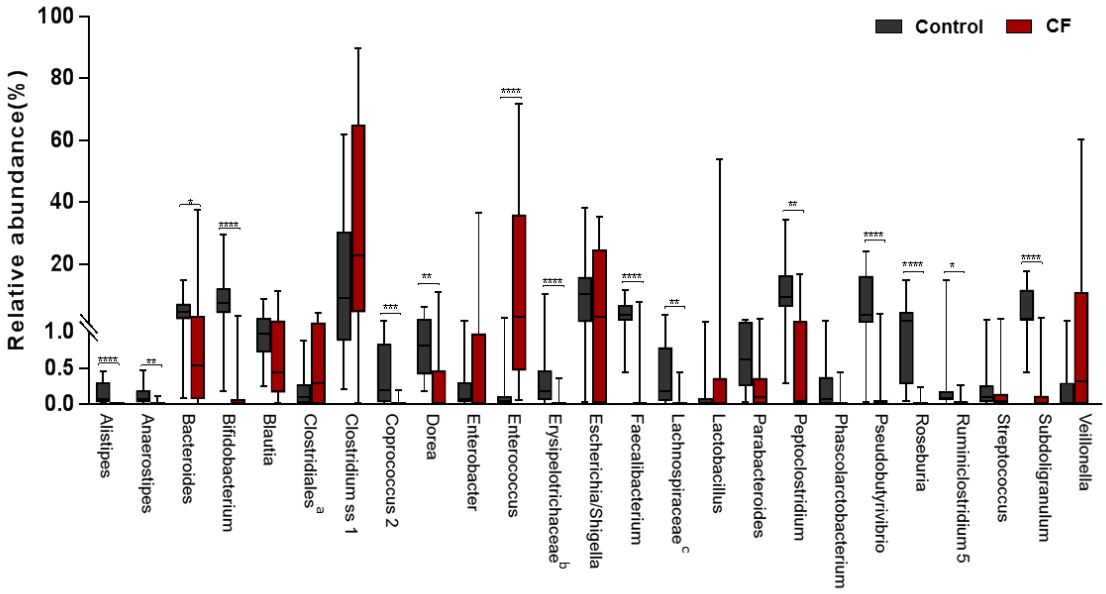
Supplementary Figure 1. Pathways (KEGG) significantly differed between non-CF (Control) and CF faecal microbiota at baseline. Differential abundance analysis of the genes were computed using Limma package on R, and pathways were enriched with MinPath (v1.4) using genes with $-2 \leq \text{Log[Fold Change]} \geq 2$.



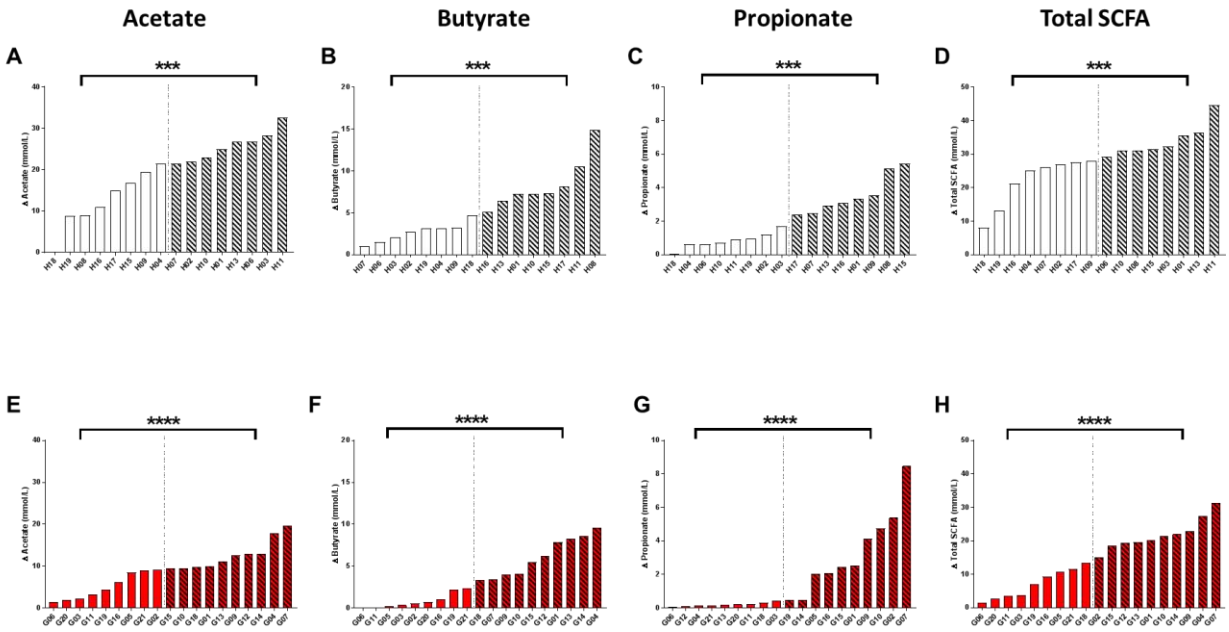
Supplementary Figure 2. Metagenomic analysis of key genes responsible for SCFA biosynthesis. The non-CF control group are shown in black, the CF group are shown in red. The following key enzymes involved in acetate (**A**), butyrate (**B**) and propionate (**C**) synthesis from hexoses were focused in this study: phosphotransacetylase (*pta)* and acetate kinase (*ackA)* in the acetate kinase –phosphate acetyltransferase pathway for acetate synthesis[1]; butyryl-CoA:acetate CoA transferase (*but*) and butyrate kinase (*buk*) in the acetyl-CoA butyrate synthesis pathway [2] for butyrate synthesis; methylmalonyl-CoA decarboxylase (*mmdA*) and lactoyl-CoA dehydratase (*lcdA*) in in succinate pathway and acrylate pathway, respectively for propionate synthesis. [3]



Supplementary Figure 3. Association between SCFA concentrations and relative abundance of taxa (spearman). Strength of the correlation is indicated by the shade of the colour, red for -1 and blue for +1. Significance of the correlation passed the Benjamini-Hochberg FDR procedure are indicated by starts * Adjust p< 0.05; ** Adjust p<0.01.[a,] Clostridiales unclassified; [b,] Erysipelotrichaceae UCG-003; [c,]Lachnospiraceae NK4A136 group; *Clostridium* ss1, *Clostridium* sensu stricto cluster 1.

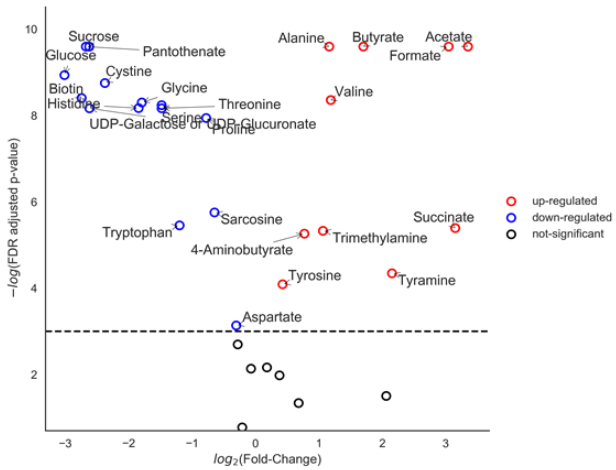


Supplementary Figure 4. Relative abundance of taxa post fermentation in non-CF control and CF samples (Mann-Whitney test adjusted with BH-FDR). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. [a,] Clostridiales unclassified; [b,] Erysipelotrichaceae UCG-003; [c,] Lachnospiraceae NK4A136 group; *Clostridium* ss1, *Clostridium* sensu stricto cluster 1.

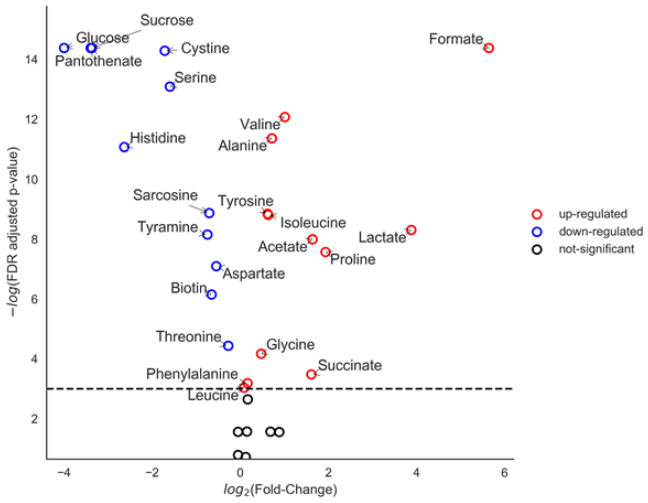


Supplementary Figure 5. SCFA production for individual samples showing varied responses to HAMS fermentation in both Control (black, A-D) and CF (red, D-H) samples. Samples produced SCFA ≥ median (showing by the vertical dashed line) in each group were defined as high responders (cross-hatched columns). *** $p < 0.001$, **** $p < 0.0001$.

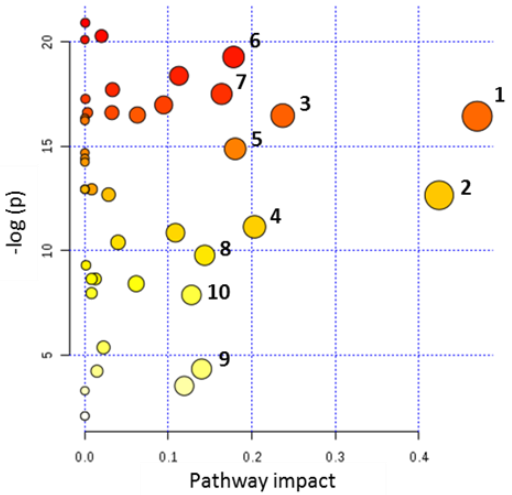
A



B



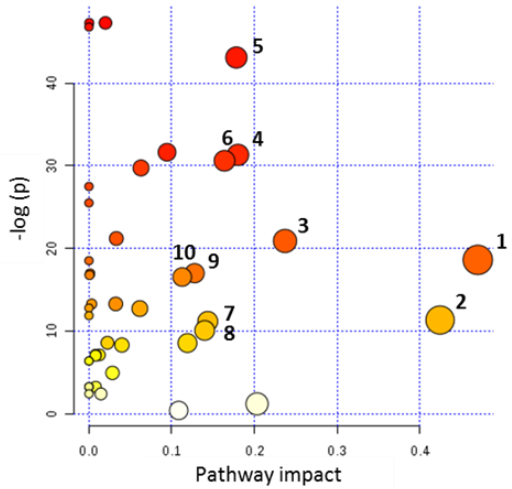
C



Most impacted pathways

1. Glycine, serine and threonine metabolism
2. Alanine, aspartate and glutamate metabolism
3. Pyruvate metabolism
4. Biotin metabolism
5. Pantothenate and CoA biosynthesis
6. Starch and sucrose metabolism
7. Methane metabolism
8. Glyoxylate and dicarboxylate metabolism
9. Histidine metabolism
10. Arginine and proline metabolism

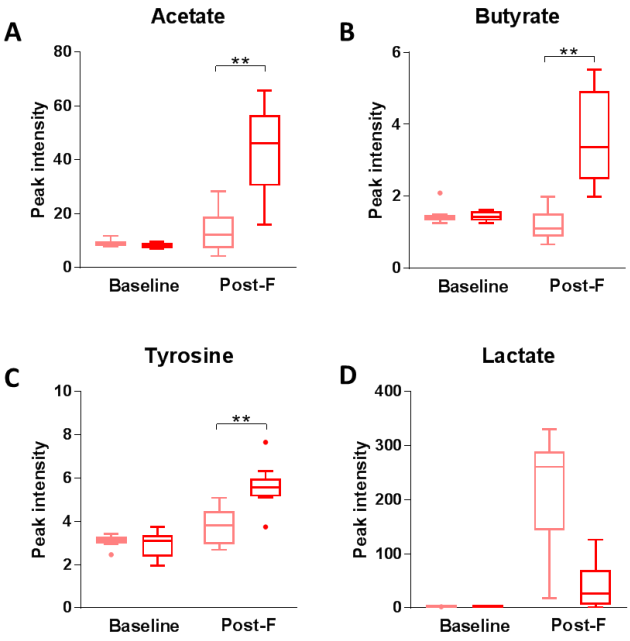
D



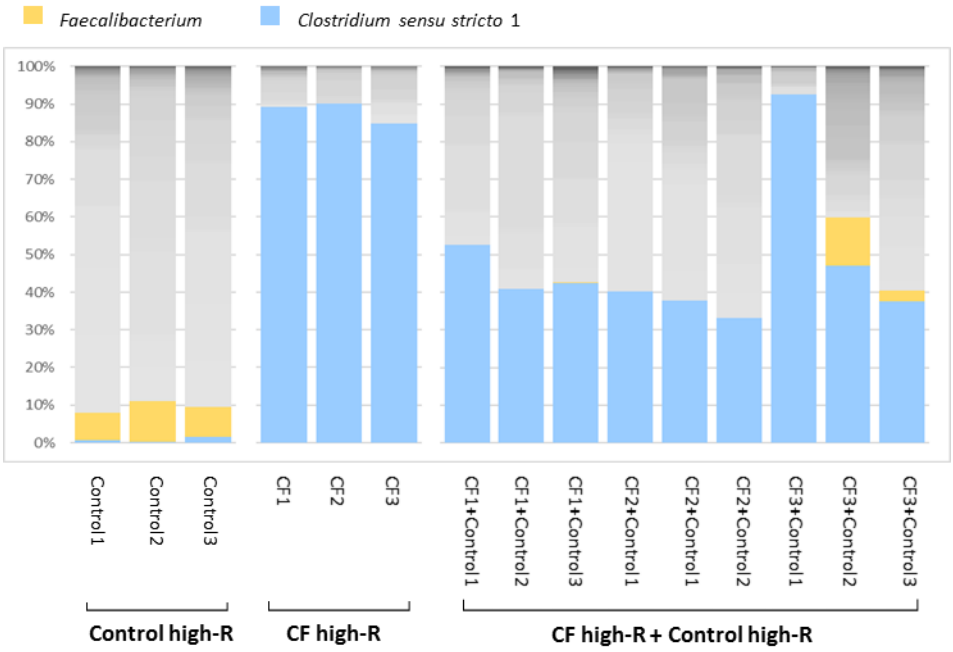
Most impacted pathways

1. Glycine, serine and threonine metabolism
2. Alanine, aspartate and glutamate metabolism
3. Pyruvate metabolism
4. Pantothenate and CoA biosynthesis
5. Starch and sucrose metabolism
6. Methane metabolism
7. Glyoxylate and dicarboxylate metabolism
8. Histidine metabolism
9. Arginine and proline metabolism
10. Phenylalanine metabolism

Supplementary Figure 6. Metabolites and predicted pathway changed during HAMS fermentation.Volcano plots showing 1H NMR metabolomic differences between baseline and post-fermentation for non-CF control **(A)** and CF **(B)** samples (red = up-regulated, post-fermentation compare against baseline; blue = down-regulated, post-fermentation compare against baseline; black, no significant difference).Weighted-scatter plots based on pathway analysis showing major differences in metabolome composition between baseline and post-fermentation for non-CF control **(C)** and CF **(D)** samples.



Supplementary Figure 7. Metabolites that differed between low responders (low-R) and high responders (high-R) of CF samples post HAMS fermentation **(A-D)**. P values were from Mann-Whitney test adjusted by FDR-BH. *, p <0.05; **, p <0.01. Pink, CF low-R; red, CF high-R.



Supplementary Figure 8. Post-fermentation taxa profile with *Clostridium* sensu stricto 1 (*Clostridium* ss1) and *Faecalibacterium* highlighted for non-CF control and CF high responding (high-R) samples as well as samples with non-CF control and CF high-R mixed in 1:1 ratio. Grey shades are non- *Clostridium* ss1 or *Faecalibacterium* taxa.

Supplementary methods

Composition of basal medium and anaerobic diluent used in the *in vitro* fermentation.

	Medium	**Anaerobic diluent**
Tryptone	2.5g/L	-
Yeast extract	0.5g/L	-
K_2HPO_4	1.20g/L	0.46g/L
NaCl	1.21g/L	0.46g/L
$(NH_4)_2SO_4$	0.60g/L	0.23g/L
$MgSO_4$.7H2O	0.28g/L	0.095g/L
$CaCl_2$.2H2O	0.16g/L	0.061g/L
$N(CH_2CO_2H)_3$	1.5mg/L	-
$MnSO_4$.7H2O	0.5mg/L	-
$FeSO_4$.7H2O	0.1mg/L	-
$ZnSO_4.7H_2O$	0.1mg/L	-
$CoCl_2.6H_2O$	0.1mg/L	-
$NiSO_4.6H_2O$	0.03mg/L	-
Na_2SeO_3	0.02mg/L	-
$AlK(SO_4)_2.12H_2O$	0.01mg/L	-
H_3BO_3	0.01mg/L	-
$CuSO_4$.5H2O	0.01mg/L	-
Haemin	0.05g/L	-
Resazurin	0.1mg/L	0.1mg/L
L-cysteine HCl	0.25g/L	0.25g/L

Pre-digestion procedure

HAMS (Hylon VII, Ingredion Incorporated, Westchestor, IL, USA) used in the *in vitro* fermentation was pre-digested with an *in vitro* method modified from that described by Woolnough *et al* [4] to simulating gastric and small intestinal starch digestion. In brief, 2.5 g of HAMS were incubated with 12.5 ml of pepsin solution (1 mg/ml in 0.02 M HCl) at 37°C for 30 min with shaking (150 rpm). After the incubation, pH was adjusted to 6.0 with the addition of 12.5 ml of 0.02M sodium hydroxide and 62.5 ml 0f 0.2 M of sodium acetate, followed by the addition of 12.5 ml of enzyme mixture containing 1.5 U/ml α-amylase (Sigma-Aldrich) and 200 U/ml of amyloglucosidase (Sigma-Aldrich). The second incubation was performed at 37°C with shaking (150 rpm) for 5 h. Subsequently, samples were precipitated for overnight with 400 ml of ethanol (100%). The precipitates were collected and washed for three times with 100 ml of 80% ethanol and one time with 25 ml of sodium acetate by centrifugation. Pellets recovered from centrifugation were collected.

Metagenomic sequencing

Illumina paired-end reads were adapter- and quality-filtered using Trimmomatic v0.32. [5] These high-quality interleaved reads were used for de novo assembly of contigs of at least 900 bp with IDBA-UD v1.1.1.[6] Gene prediction was performed using MetaGeneMark [7], with genes less than 100 bp discarded. A non-redundant gene catalogue of 1,900,153 genes was constructed using CD-HIT [8] and parameters "-c 0.95 –aS 0.9" (genes with greater than 95% identity and aligned length covering over 90% of the shorter gene were grouped together). The gene catalogue was annotated to the KEGG database (release 2017-04-17) using BLASTP with e-values $\leq 1 \times 10^{-5}$. High quality reads from each sample were aligned against the gene catalogue using SOAPAligner [9], and gene-length normalized read counts calculated using soap.coverage [9]. Relative gene abundances were estimated by dividing the number of the gene-

length normalized read counts for each gene by the total of reads from that sample that uniquely mapped to a gene in the catalogue.

References

1. Ferry JG. Acetate kinase and phosphotransacetylase. Methods Enzymol 2011; 494:219-31.
2. Vital M, Howe AC, Tiedje JM. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. MBio 2014; 5:e00889.
3. Reichardt N, Duncan SH, Young P, Belenguer A, McWilliam Leitch C, Scott KP, et al. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. Isme J 2014; 8:1323-35.
4. Woolnough JW, Bird AR, Monro JA, Brennan CS. The effect of a brief salivary alpha-amylase exposure during chewing on subsequent in vitro starch digestion curve profiles. Int J Mol Sci 2010; 11:2780-90.
5. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30:2114-20.
6. Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 2012; 28:1420-8.
7. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res 2010; 38:e132.
8. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006; 22:1658-9.
9. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, et al. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 2009; 25:1966-7.