

## Supplementary information

### **Go beyond “abundance”: cross-cohort single-nucleotide-variant profiling of gut microbiota suggests a novel gut-health assessment approach**

Chenchen Ma<sup>1, 4#</sup>, Yufeng Zhang<sup>2#</sup>, Shuaiming Jiang<sup>1</sup>, Fei Teng<sup>3</sup>, Shi Huang<sup>2\*</sup>, Jiachao Zhang<sup>1,5\*</sup>

1. College of Food Science and Engineering, Key Laboratory of Food Nutrition and Functional Food of Hainan Province, Hainan University, Haikou, 570228, China

2. Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, 999077, China

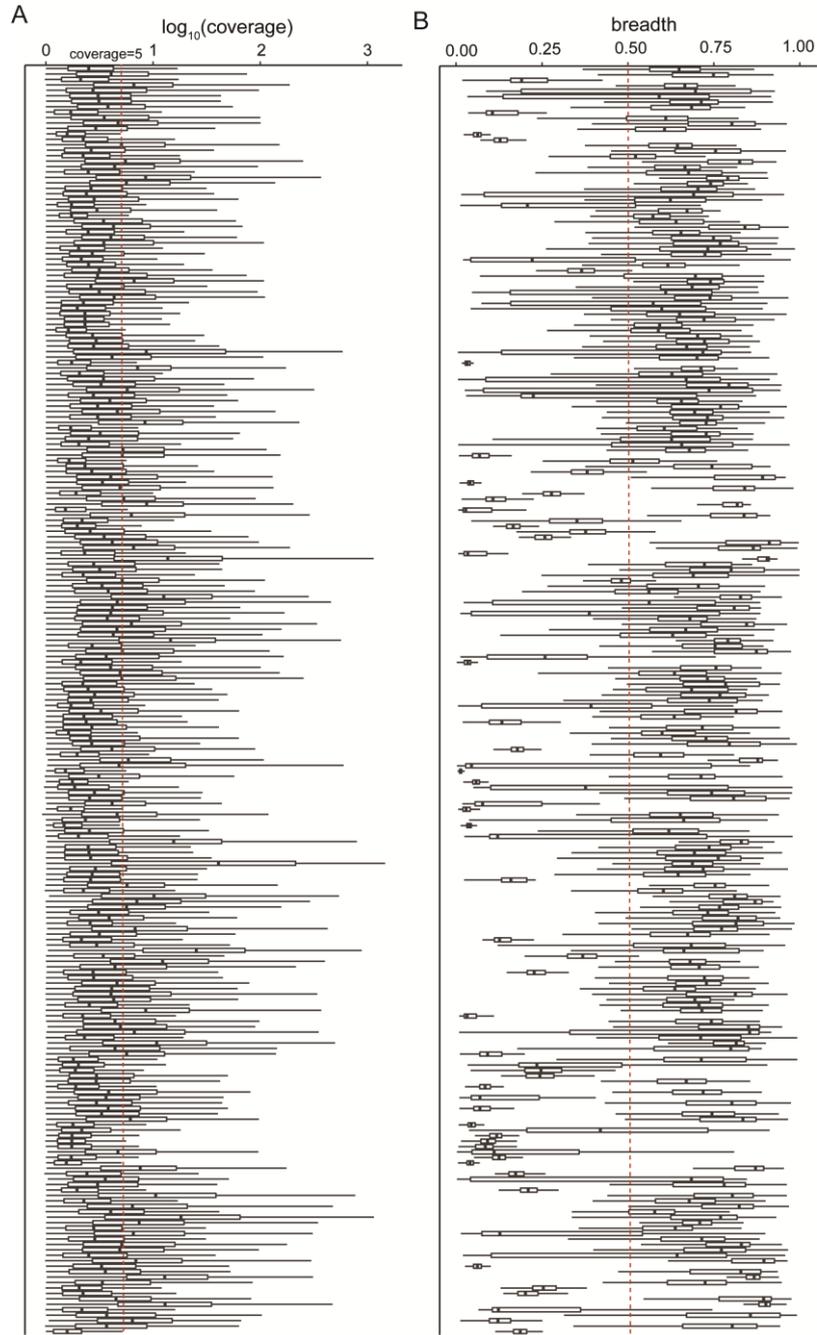
3. Qingdao Stomatological Hospital Affiliated to Qingdao University, Qingdao, 266000, China

4. School of Medicine, Southern University of Science and Technology, Shenzhen, 518055, China

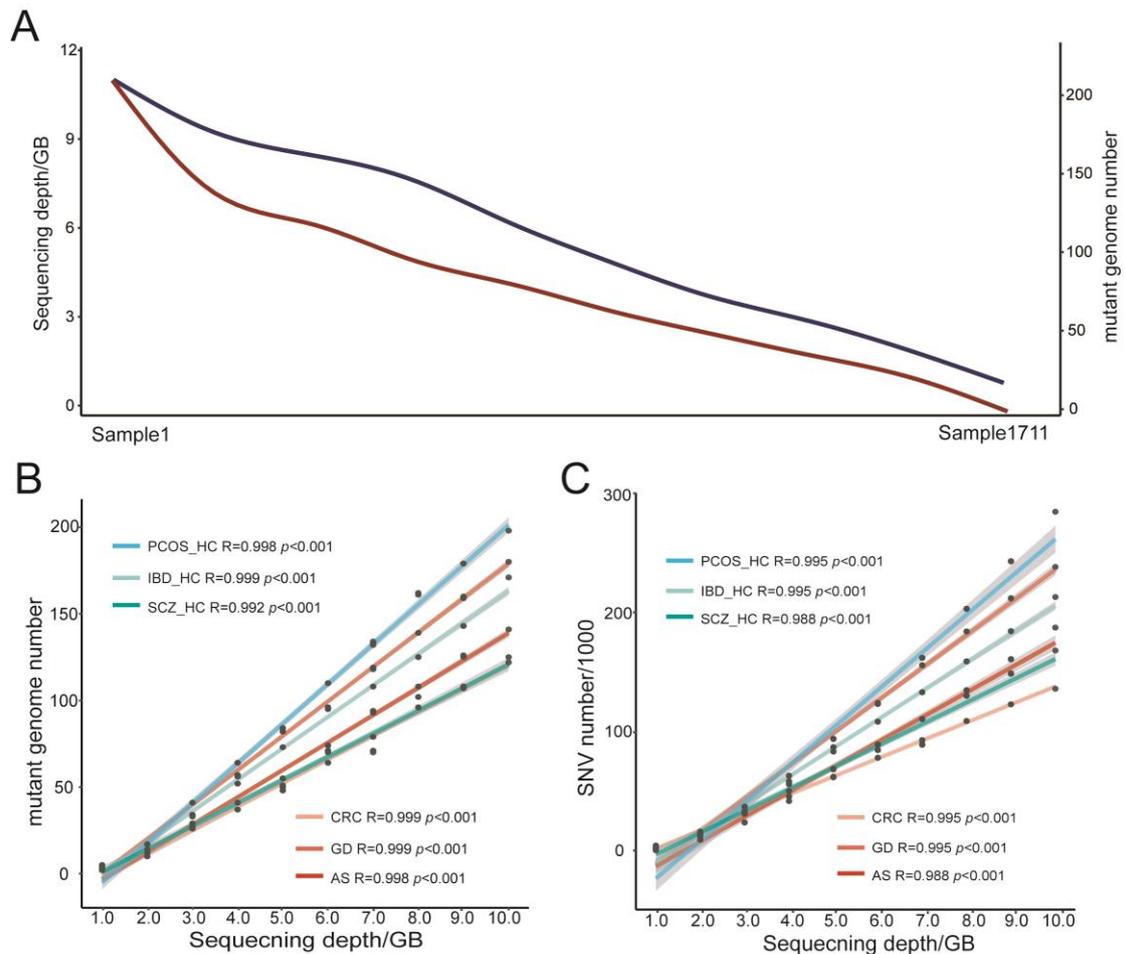
5. One Health Institute, Hainan University, Haikou, Hainan 570228, China

# These authors contributed equally to this work.

\* Correspondence: Jiachao Zhang (E-mail: [zhjch321123@163.com](mailto:zhjch321123@163.com)) and Shi Huang (E-mail: [shihuang@hku.hk](mailto:shihuang@hku.hk))

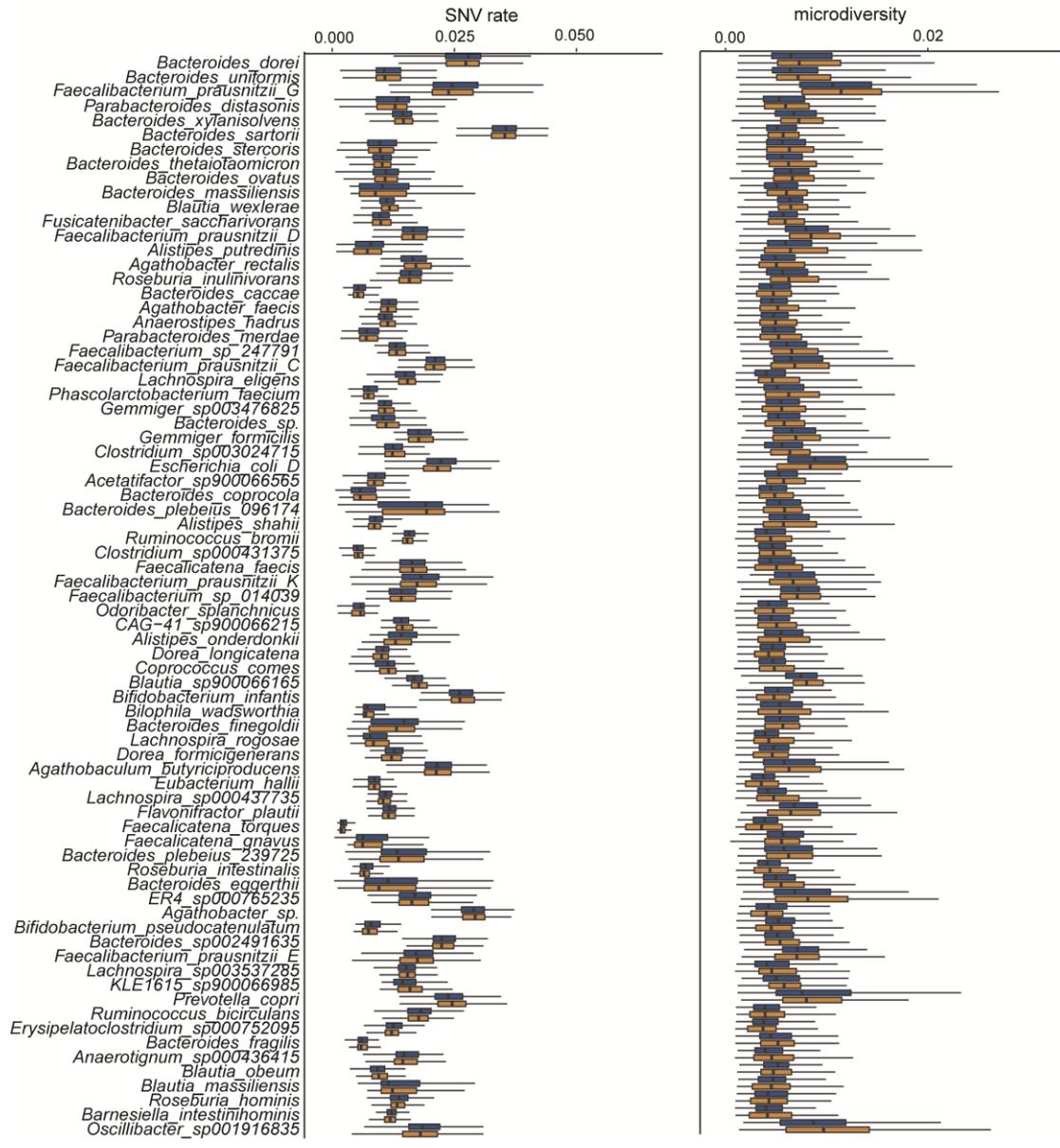


**Figure S1. Coverage and breadth profiles of the mutated genomes. A)** 233 mutated genomes in at least 10% of the samples were concerned, and their average coverage in the detected samples was evaluated. The coverage of SNV was required to be 5, and that's the minimum limit, which could cover most of the genomes. **B)** The breadth of the above samples was also evaluated and compared, and most strains have high breadth, which can meet the annotation of SNV.



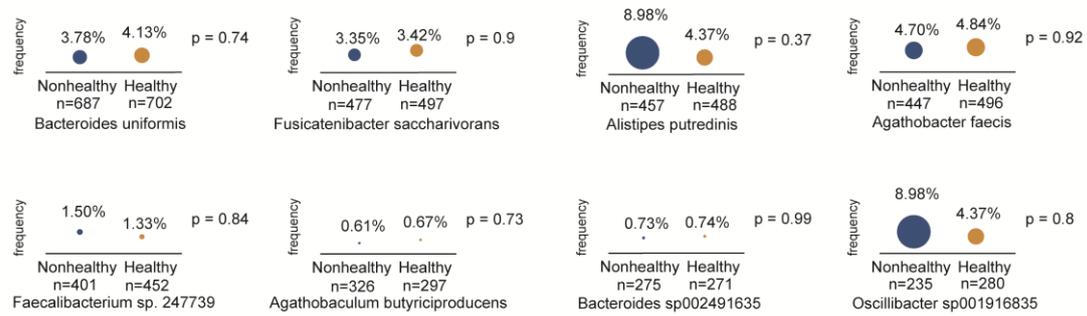
**Figure S2. Evaluation of the influence of sequencing depth by simulated data. A)**

The decreasing trend of the sequencing depth and the number of mutated genomes of 1711 samples, and the cross section represents the same sample. The result showed that the number of mutated genomes is related to the sequencing depth. **B-C)** The simulated data showed the strong positive correlation in the number of mutated genomes and SNV and the sequencing depth of samples and the pipeline of simulated data can be found in the “Methods”.

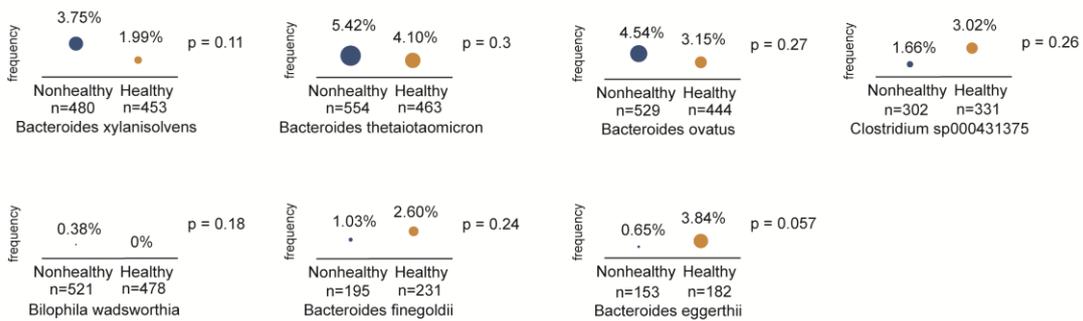


**Figure S3. SNV rate and microdiversity of 75 strains.** The boxplot showed that the values of SNV rate and microdiversity of 75 strains.

ack



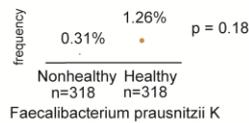
buk



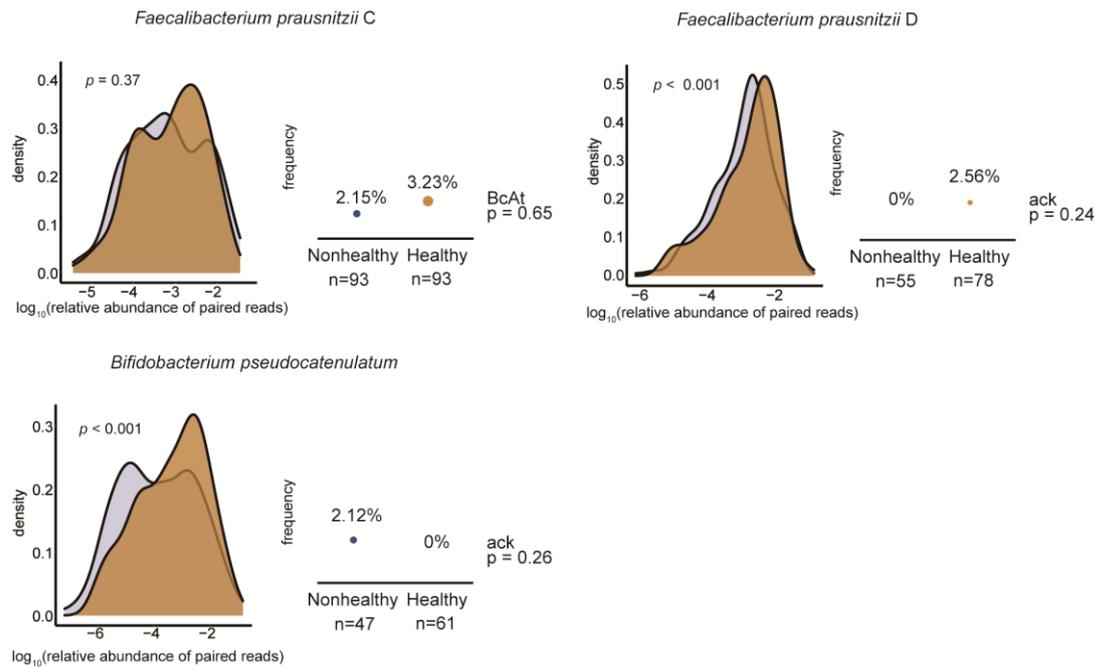
ydiF/atoD/A



BcAt



**Figure S4. Codon mutation bias in SCFAs gene on initiator and terminator.** All codon mutation bias of SCFAs gene of the strain marked with red font in Figure 3 was shown. N represented the number of individuals with SNV on these genes. When the prevalence was calculated, the denominator was N, and the numerator was the number of individuals with codon mutation bias. The p values were determined with the Wilcoxon rank-sum test.



**Figure S5. Validation cohorts of codon mutation bias in SCFAs gene.** The abundance of three strains and the prevalence of characteristic codon mutation bias were described. Validation cohorts confirmed the characteristic codon mutation bias on ack gene of *Faecalibacterium prausnitzii* C and scpC gene of *Bacteroides stercoris*. However, BcAt gene of *Faecalibacterium prausnitzii* C ( $p=0.65$ ), ack gene of *Faecalibacterium prausnitzii* D ( $p=0.24$ ) and ack gene of *Bifidobacterium pseudocatenulatum* ( $p=0.26$ ) have not been confirmed.