

## Supplementary Materials

### Unique characteristics of gut microbiota in black snub-nosed monkeys (*Rhinopithecus strykeri*) reveal an enzymatic mechanism of adaptation to dietary vegetation

Xiao-Chen Wang<sup>1,#</sup>, Jia-Li Zhang<sup>1,2,#</sup>, Hui-Juan Pan<sup>3,#</sup>, Yi-Xin Chen<sup>4,#</sup>, Shu-Xin Mao<sup>1,2</sup>, Ji-Wei Qi<sup>1</sup>, Ying Shen<sup>1</sup>, Ming-Yi Zhang<sup>3</sup>, Zuo-Fu Xiang<sup>5,\*</sup>, Ming Li<sup>1,6,\*</sup>

<sup>1</sup> CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

<sup>2</sup> College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China

<sup>3</sup> School of Ecology and Nature Conservation, Beijing Forestry University, Beijing, 100083, China

<sup>4</sup> College of Life Sciences and Technology, Central South University of Forestry and Technology, Changsha, Hunan 410004, China

<sup>5</sup> College of Forestry, Central South University of Forestry and Technology, Changsha, Hunan 410004, China

<sup>6</sup> Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan 650223, China

#Authors contributed equally to this work

\*Corresponding authors, E-mail: xiangzf@csuft.edu.cn; lim@ioz.ac.cn

## SUPPLEMENTARY MATERIALS AND METHODS

### Sample collection

We collected 42 fecal samples from wild population, including 13 *R. strykeri* (Rst) individuals from the Gaoligong Mountain National Nature Reserve in Nujiang Prefecture, Yunnan Province, 15 *R. roxellana* (Rro) individuals from the Jiuzhaigou population in Sichuan Province (three samples), Yuhuangmiao population in Shaanxi Province (five samples), and Shennongjia population in Hubei Province (seven samples), and 14 *R. bieti* (Rbi) individuals from the Xiangguging population in Yunnan Province (five samples), Lasha Mountain population in Yunnan Province (five samples), and Xiaochangdu population in Tibet (four samples) (Supplementary Table S1). As the fecal samples were collected from wild populations, we were unable to identify individual monkeys, except for the Shennongjia population. We did not obtain information on age or sex but could infer that the samples were from adult monkeys based on the size of the feces. No duplicate individuals were sampled (with simultaneous collection at overnight trees or gathering sites) and samples were collected in different seasons. The samples were used to explore potential similarities in the gut microbiota of snub-nosed monkeys from the seven different habitat sites, and the unique adaptive characteristics of black snub-nosed monkeys. Fresh fecal samples were immediately collected in 3 mL of RNeasy lysis buffer (Qiagen, Valencia, CA, USA), then transported on dry ice within one week and stored at  $-80^{\circ}\text{C}$  until DNA extraction at the Institute of Zoology, Chinese Academy of Sciences (CAS), Beijing, China.

### DNA extraction and sequencing

Microbial DNA was extracted from the fecal samples using a QIAamp DNA stool mini kit (Qiagen, Valencia, CA, USA) following standard protocols. DNA quality and quantity were determined using Nanodrop (ND-1000) spectrophotometry (Nanodrop Technologies, Wilmington, DE, USA) and agarose gel electrophoresis, respectively. DNA samples were stored at  $-20^{\circ}\text{C}$  until use. Shotgun sequencing was performed using an Illumina NovaSeq 6000, with at least 10 Gb per sample. Raw data were filtered using Trimmomatic v0.36 (Bolger et al., 2014) to trim low-quality reads: 3' tailing sequences were removed when the average quality over a 4 bp sliding window was less than 20 and reads less than 70 bp were discarded. The genomes of *R. strykeri* (assembly ASM2376470v1), *R. bieti* (assembly ASM169854v2), *R. roxellana* (assembly ASM756505v1), and *Homo sapiens* (assembly GRCh38.p13) were used to remove contamination with bowtie2 v2.3.5 (Langmead & Salzberg, 2012) to obtain clean data.

After removal of low-quality and contaminated reads, an average of 12.26 Gb of high-quality nonhost sequences were obtained per sample in the Rst, Rbi, and Rro groups (Supplementary Table S1). We assessed 4 623 bacterial terms (38 phyla, 10 classes, 14 orders, 25 families, 43 genera, and 73 species), 169 fungal terms (three phyla, 78 classes, 180 orders, 398 families, 1 197 genera, and 2 732 species), 527 viral terms (12 phyla, 16 classes, 20 orders, 39 families, 169 genera, and 271 species), and 575 CAZymes, which formed the gut microbiomes of the three snub-nosed monkey species (Supplementary Table S5).

### Determination of relative abundance of taxonomic and functional terms

Sequence taxonomy was resolved using Kraken2 and the PlusPF database (Standard plus protozoa & fungi database) with default parameters (Wood et al., 2019). The Kraken2-mapped sequence counts were subsequently re-estimated using Bracken, a rapid read-level classifier, along with information about the genomes themselves to estimate abundance at the species, genus, or higher level of bacteria, fungi, and viruses (Lu et al., 2016). We further calculated the relative abundance of each taxonomic term for bacteria, fungi, and viruses using an in-house script. Taxonomic terms that did not exceed a maximum relative abundance of  $1 \times 10^{-4}$  were excluded from further analysis, together with taxonomic terms accounting for less than 20% of the samples in each cohort (Wirbel et al., 2019).

We used single-sample metagenomic assembly and functional annotation to analyze the gut

microbiota of all three species. Briefly, assemblies were produced with MEGAHIT (v1.2.6) (Li et al., 2015), and gene identification was performed on contigs longer than 300 bp using MetaGeneMark (Zhu et al., 2010). The contigs were then annotated with carbohydrate-active enzymes (CAZymes) (Lombard et al., 2014) to obtain a complete picture of the carbohydrate digestibility capabilities of gut microbiota using DIAMOND (v0.9.24) (Buchfink et al., 2015) with parameters -d -q -e 1e-5 -k 1. We further calculated the relative abundance of each functional term, then removed terms detected (i) in less than 20% of the samples in each group or (ii) in no sample in any group. Specifically, functional terms that did not exceed a maximum relative abundance of  $1 \times 10^{-5}$  were excluded from further analysis (Wirbel et al., 2019).

### **Analysis of bacterial coexistence in three snub-nosed monkey species**

Bacterial, fungal, and viral genera present in at least 20% of individuals within at least one host group and with a relative abundance greater than 0.005 were screened for coexistence analysis (Doron et al., 2021). First, we constructed a heatmap representing relative abundance of each genus (columns) in each sample (rows). The heatmap was hierarchically clustered using average linkage based on Euclidean distance, showing coexistence of bacterial genera within host groups. Second, a correlation matrix was calculated between vectors of abundances of bacterial genera across all samples. Genera were hierarchically clustered with average linkage based on Euclidean distance. Third, clusters of coexisting genera were visually selected and their abundance in the corresponding hosts was visualized using a heatmap. To relate the clusters back to the groups that contained them, heatmaps using average relative abundance of genera for each snub-nosed monkey group were visualized for each cluster.

### **Analysis of gut microbiota diversity of three snub-nosed monkey species**

Alpha diversity (Chao1 and Shannon indices) was calculated based on genus abundance for the three groups using the R package “vegan” (Oksanen et al., 2019). The Shannon diversity index accounts for genus richness and evenness, whereas the Chao1 index extrapolates the number of rare taxa that may have been detected with deeper sampling. Each metric was further compared using the Wilcoxon rank sum test with the R package “coin” (Hothorn et al., 2006), and *P*-values were adjusted using the Benjamini-Hochberg false discovery rate (FDR) (Benjamini & Hochberg, 1995). Principal component analysis (PCoA) and non-metric multidimensional scaling (NMDS) based on PERMANOVA were performed using the R package “vegan” to analyze the gut microbiota of the three groups (Oksanen et al., 2019), with the results visualized using the “ggplot2” package. The *P*- and *R*-values were calculated based on “ANOSIM” using the R package “vegan” (Oksanen et al., 2019).

### **Analysis of functional differences among three snub-nosed monkey species**

Diversity analysis indicated that the gut microbiota structures of Rbi and Rro were highly similar but differed significantly from that of Rst. To explore the unique or significantly enriched functional terms in Rst, we compared the relative abundances of CAZymes across all individuals in the three host groups. We used the fold-change method to conduct differential analysis and selected functional terms with Log2 fold-change >2 in both Rst/Rbi and Rst/Rro, as well as functional terms with zero relative abundance in Rbi and Rro but not in Rst, as significantly more abundant terms in the Rst group (Supplementary Table S3). To explore the correlation between coexisting bacteria in Rst and significantly abundant functional terms, we constructed a correlation network diagram between the genera and functional terms. Pearson correlation analysis was performed between genera and functional terms, then a matrix of correlation coefficients was visualized using a correlation network diagram (<https://www.omicstudio.cn/tool/>).

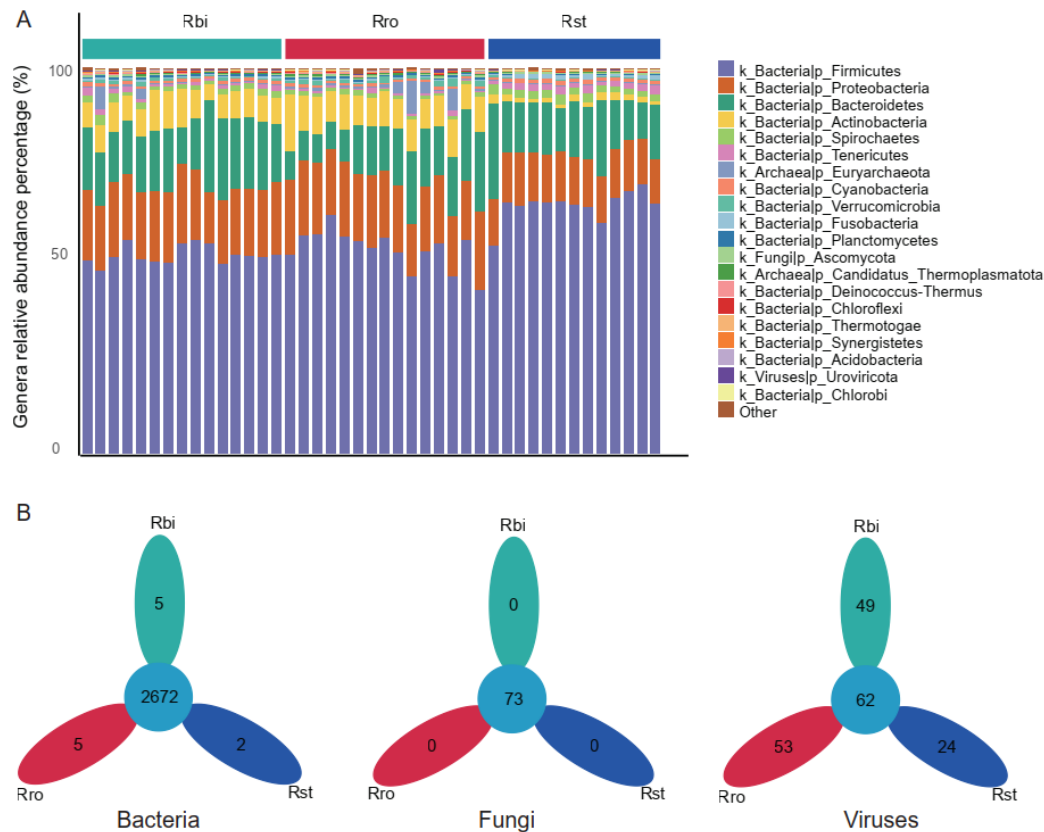
### **Analysis of core microbes and CAZymes**

We selected core microbes and CAZymes from Rst, Rro, and Rbi to explore the general characteristics of the gut microbiota in snub-nosed monkeys. Specifically, we choose microbes

present in all individuals, with average relative abundance in each population greater than 0.1%. The same approach and thresholds were used for the selection of core CAZymes. In total, 95 bacterial genera, 24 fungal genera, 12 viral genera, and 21 CAZymes were screened among all individuals of the three snub-nosed monkey species (Supplementary Table S4).

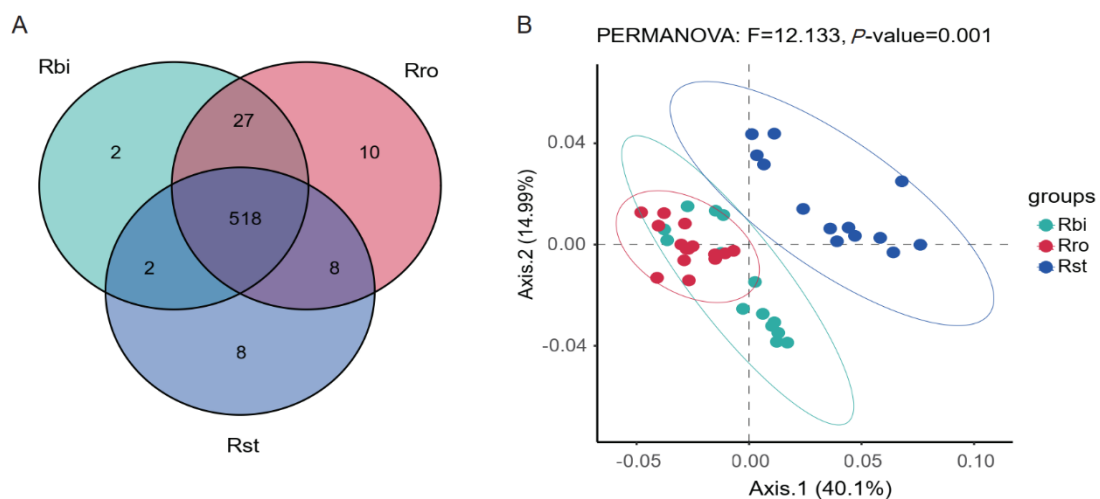
## References

- Benjamini, Hochberg Y. 1995. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple testing. *Journal of the Royal Statistical Society*, **57**(1): 289–300.
- Blekhman R, Tang K, Archie EA, Barreiro LB, Johnson ZP, Wilson ME, et al. 2016. Common methods for fecal sample storage in field studies yield consistent signatures of individual identity in microbiome sequencing data. *Scientific Reports*, **6**: 31519.
- Bolger AM, MarcLohse, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**(15): 2114–2120.
- Buchfink B, Xie C, Huson DH. 2015. Diamond : fast and sensitive protein alignment using diamond. *Nature Methods*, **12**(1): 59-60.
- Doron L, Neta R, Yishay P, Daphna R, Nadav M, Gal Z, et al. 2021. Diversity and functional landscapes in the microbiota of animals in the wild. *Science*, **372**(6539).
- Hothorn T, Hornik K, Wiel MA, Zeileis A. 2006. A Lego System for Conditional Inference. *The American Statistician*, **60**: 257–263.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. *Nature Methods*, **9**: 357–359.
- Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, **31**: 1674–1676.
- Lombard V, Golaconda RH, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, **42**: D490–495.
- Lu J, Breitwieser FP, Thielen P, Salzberg SL. 2016. Bracken: Estimating species abundance in metagenomics data. *bioRxiv*.
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. 2019. Vegan: Community Ecology Package. *R Packag*, 2.5-6.
- Vlckova K, Mrazek J, Kopečný J, Petrzeltkova KJ. 2012. Evaluation of different storage methods to characterize the fecal bacterial communities of captive western lowland gorillas (*Gorilla gorilla gorilla*). *Journal of Microbiological Methods*, **91**(1): 45–51.
- Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, et al. 2019. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nature Medthods*, **25**(4): 679–689.
- Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *bioRxiv*.
- Zhu W, Lomsadze A, Borodovsky M. 2010. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Research*, **38**(12), e132.



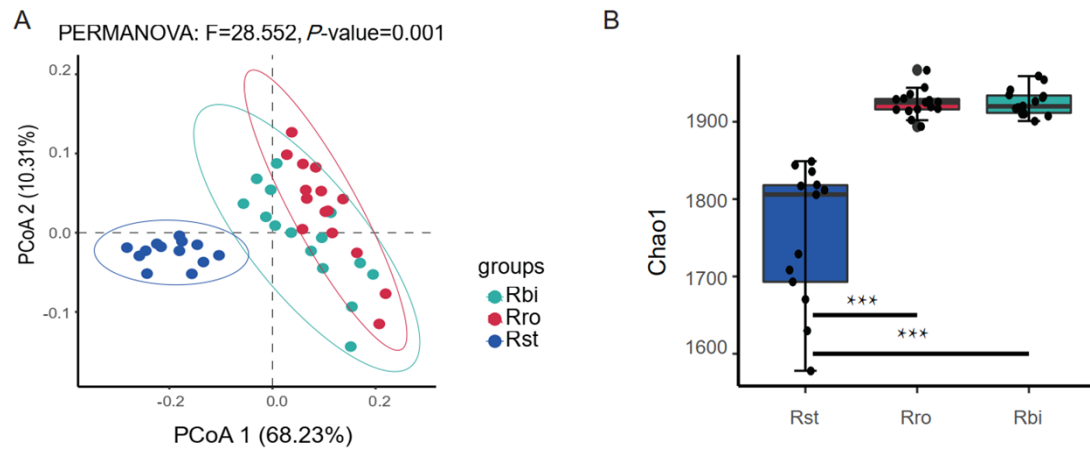
**Supplementary Figure S1 Community components of gut microbiota in three snub-nosed monkey species**

A: Relative abundances of gut microbial communities, including bacteria, fungi, and viruses, at the phylum level in three snub-nosed monkey groups. B: Venn diagram of bacteria, fungi, and viruses at the genus level in three snub-nosed monkey groups.



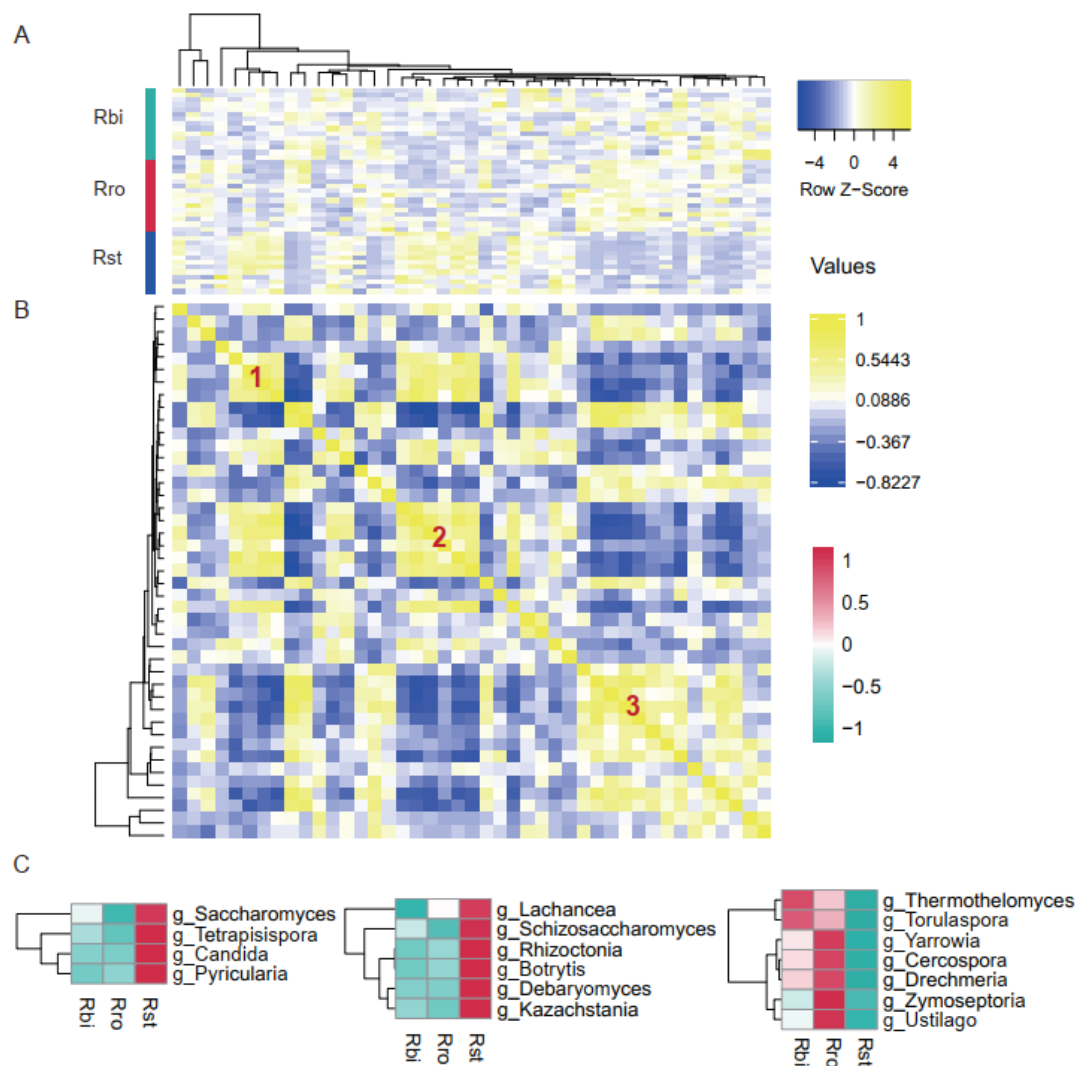
**Supplementary Figure S2 Differential analysis of taxa and CAZymes among three snub-nosed monkey species**

A: Venn diagram of microbial abundances in three snub-nosed monkey species. B: Bray-Curtis PCoA based on CAZy database functional terms.



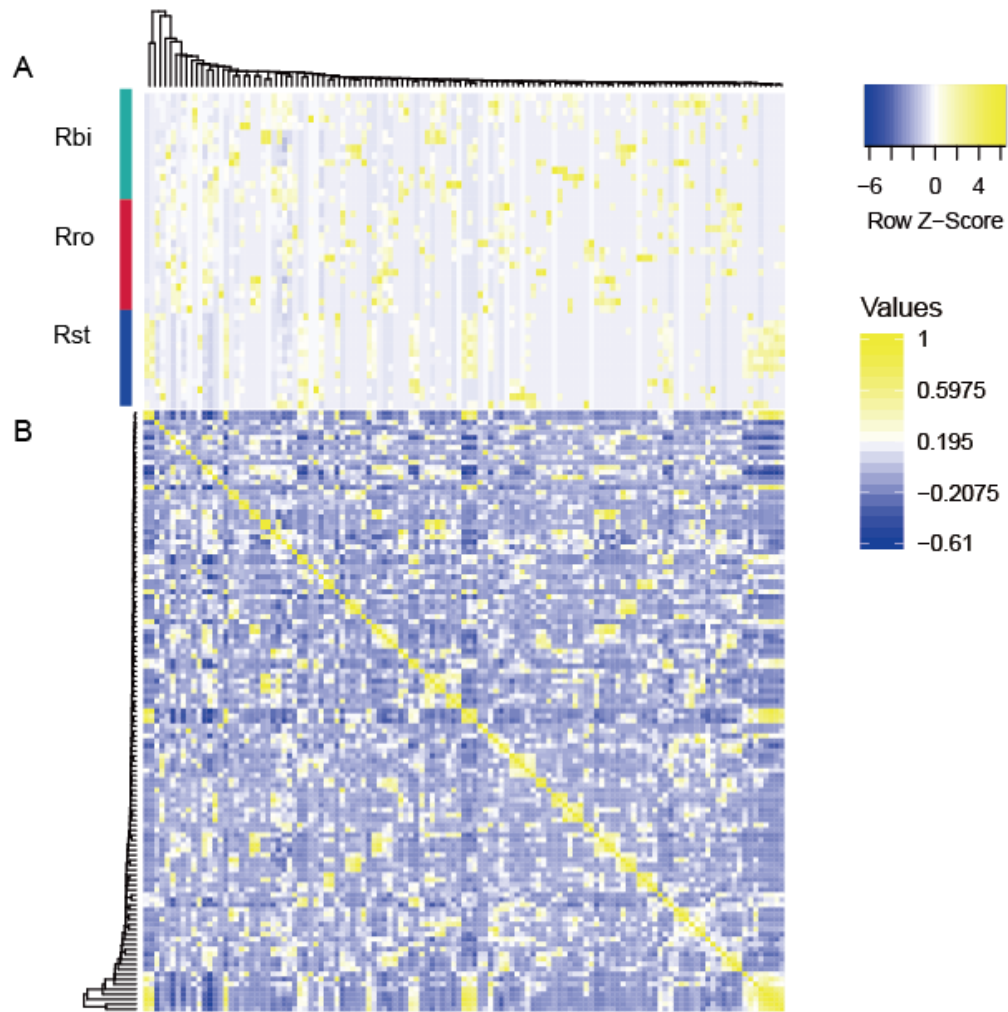
**Supplementary Figure S3 Community structure, richness, and diversity of gut microbiota of three snub-nosed monkey species**

A: Representative PCoA plots of gut microbiota community composition of three snub-nosed monkey species at the genus level. B: Representative alpha diversity of gut microbiota of three snub-nosed monkey species at the genus level.



**Supplementary Figure S4 Association between microbiota composition and host class, expressed as coexisting fungal clusters**

A: Heatmap representing relative abundance of each fungal genus in each sample. Heatmap was hierarchically clustered using average linkage based on Euclidean distance, showing coexistence of several groups of fungal genera within specific groups. B and C: Correlation matrix calculated among vectors of abundances of fungal genera across all samples. Fungal genera were hierarchically clustered using average linkage based on Euclidean distance. Three clusters of coexisting fungal genera were visually selected, with their abundances in corresponding hosts depicted in (C).



**Supplementary Figure S5 Association between microbiota composition and host class, expressed as coexisting viral clusters**

A: Heatmap representing relative abundance of each viral genus in each sample. Heatmap was hierarchically clustered using average linkage based on Euclidean distance, showing coexistence of several groups of viral genera within specific groups. B: Correlation matrix calculated among vectors of abundances of viral genera across all samples. Viral genera were hierarchically clustered using average linkage based on Euclidean distance. There were no significant clusters of coexisting viral genera.



**Supplementary Tables**

**Supplementary Table S1 Sample information of three snub-nosed monkey species**

**Supplementary Table S2 Coexisting bacterial and fungal clusters in three snub-nosed monkey species**

**Supplementary Table S3 CAZymes with high abundance in Rst**

**Supplementary Table S4 Core bacteria, fungi, viruses, and CAZymes in three snub-nosed monkey species**

**Supplementary Table S5 Relative abundance of bacteria, fungi, viruses, and CAZymes in three snub-nosed monkey species**

**Supplementary Tables S1–S5** are listed separate excel files due to their large size.