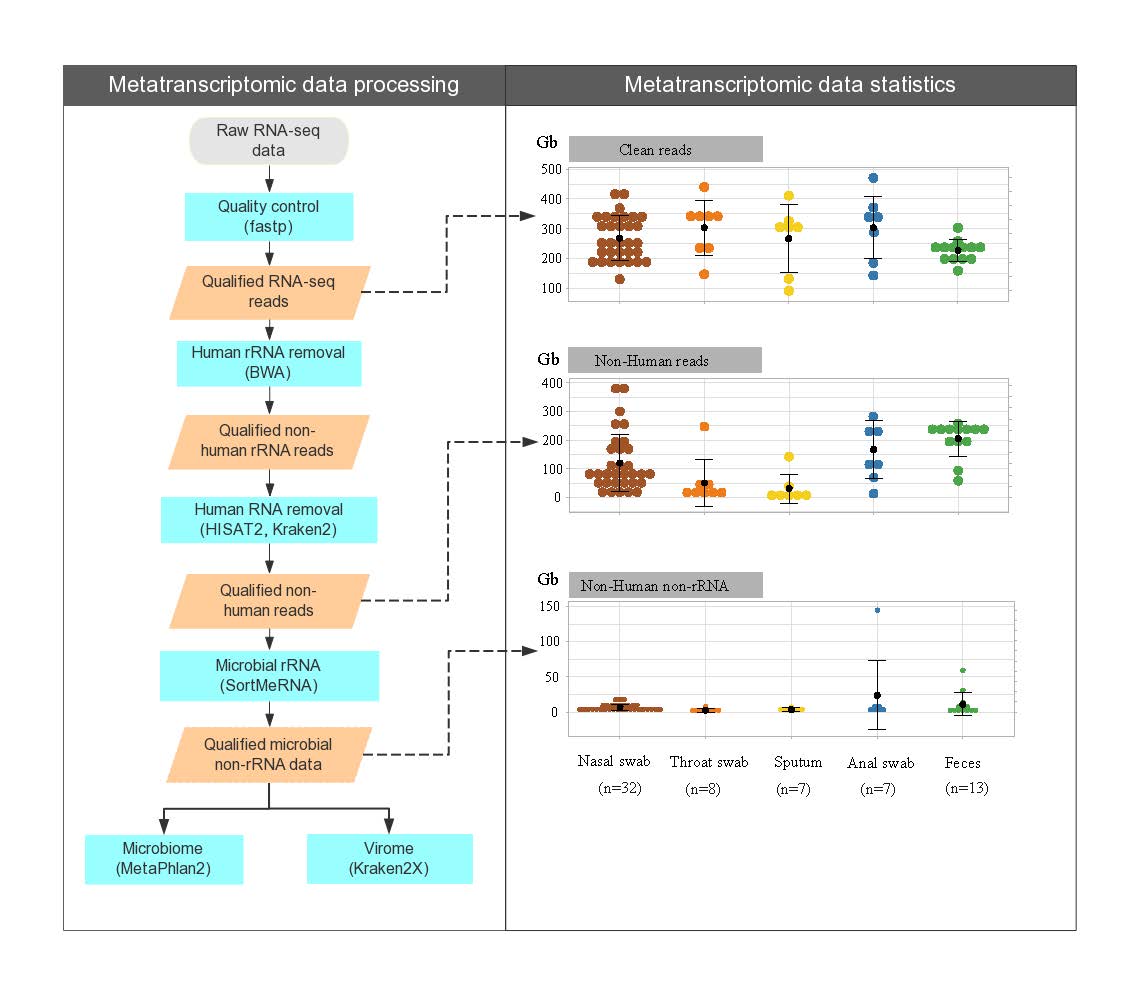
**Supplementary appendix 1 for**

**Characterization of Microbial Co-infections in the Respiratory Tract of hospitalized COVID-19 patients**

Huanzi Zhong1,3\*, Yanqun Wang2\*, Zhun Shi1\*, Lu Zhang4\*, Huahui Ren1,3\*, Weiqun He2\*, Zhaoyong Zhang2\*, Airu Zhu2\*, Jingxian Zhao2\*, Fei Xiao5\*, Fangming Yang1,6, Tianzhu Liang1,7, Feng Ye2, Bei Zhong8, Shicong Ruan9, Mian Gan2, Jiahui Zhu1,10, Fang Li2, Fuqiang Li1,11, Daxi Wang1,7, Jiandong Li1,7,12, Peidi Ren1,7, Shida Zhu1,13, Huanming Yang1,14,15, Jian Wang1,14, Karsten Kristiansen1,3, Hein Min Tun16, Weijun Chen17,12, Nanshan Zhong2, Xun Xu1,18#, Yi-min Li2#, Junhua Li1,6,19#, Jincun Zhao2,4#

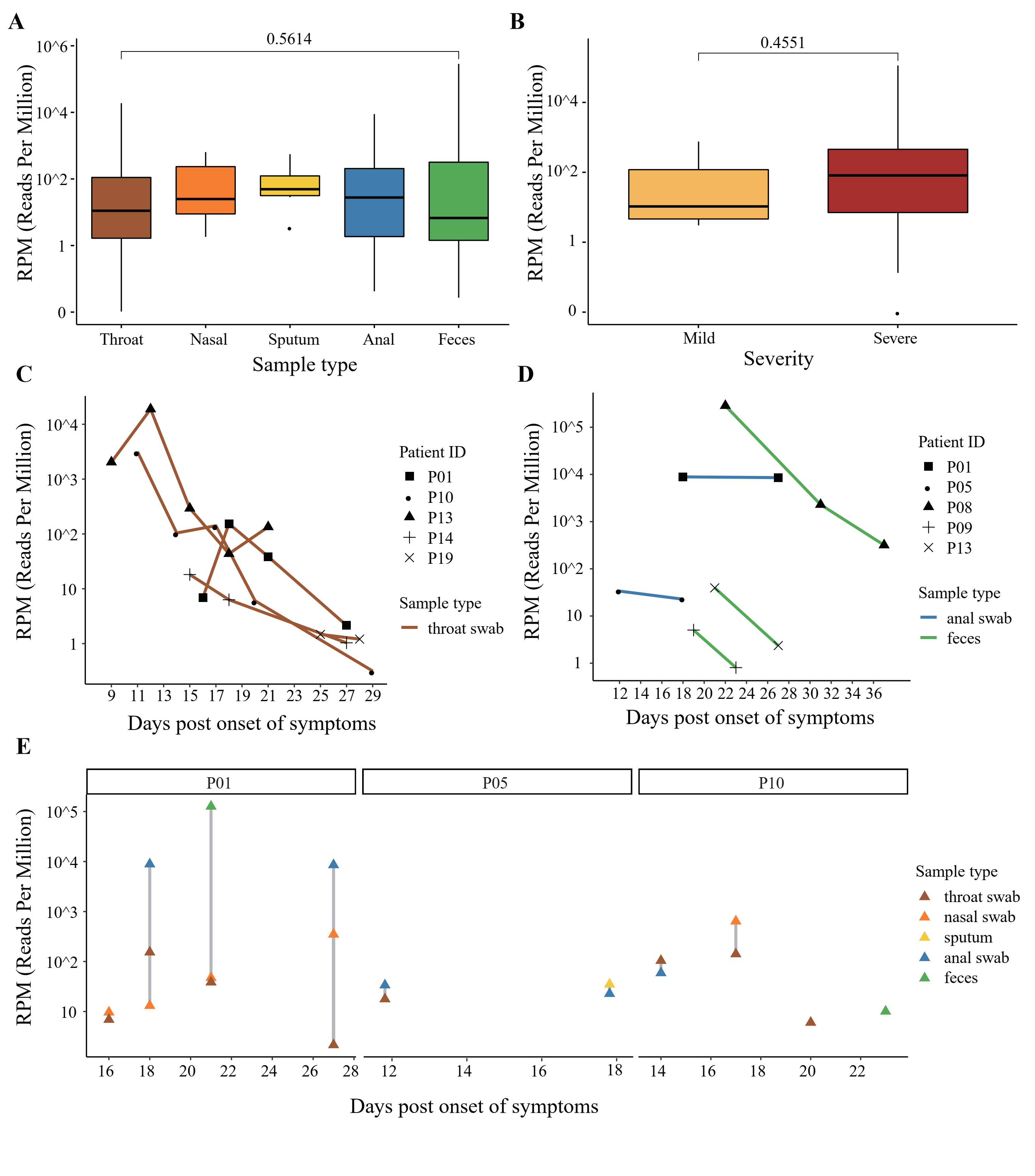
\* These authors contributed equally to this work.

#Correspondence should be addressed to **Jincun Zhao** (zhaojincun@gird.cn), **Junhua Li** (lijunhua@genomics.cn), **Yi-min Li** (dryiminli@vip.163.com), and **Xun Xu** ([xuxun@genomics.cn](mailto:xuxun@genomics.cn)).



**Supplementary figure 1: Flowchart of the metatranscriptomic analysis pipeline**

Raw metatranscriptomic sequencing data from 67 clinical samples were processed as follows: 1) quality control by fastp and generation of clean data, 2) identification and removal of human rRNA from high-quality clean data by BWA, 3) identification and removal of human non-rRNA reads by HISAT2(hg19) and Kraken2 (hg38) and generation of non-human RNA-seq reads, 4) identification and removal of microbial rRNA reads by SortMeRNA and generation of non-rRNA microbial data for subsequent analyses of virome (by Kraken2X) and non-viral microbiome (by MetaPhlan2). Total clean RNA-seq data (upper panel), total non-human RNA-seq data (middle panel) and total non-human non-rRNA microbial RNA-seq data (lower panel) of 67 clinical samples are shown. Brown, throat swab (n=32); orange, nasal swab (n=8); yellow, sputum(n=7); blue, anal swab(n=7); and green, feces (n=13). Detailed statistics for each processing step are shown in **appendix 2 p4, Table S3.**



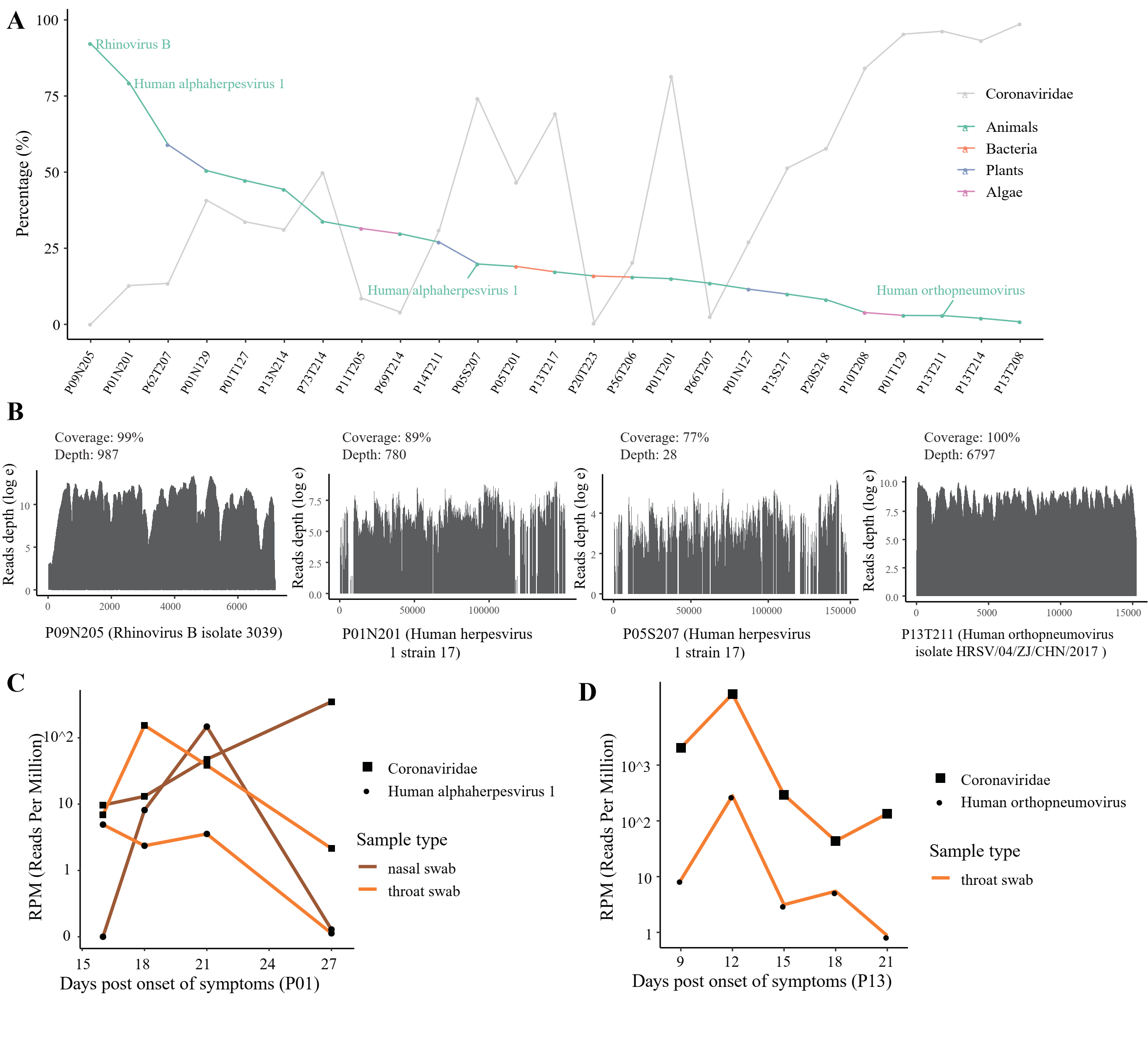
**Supplementary figure 2: Comparison of SARS-CoV-2 viral loads across specimens and individuals**

**A,** Bar plot showing thenumber of SARS-CoV-2-like reads in different types of clinical specimens.

**B**, Bar plot showing thenumber of SARS-CoV-2-like reads in respiratory tract specimens from mild (orange) and severe (brick red) cases.

**C-D,** Individual-based changes inthenumber of SARS-CoV-2-like reads in specimens from the respiratory tract (**c**) and the gastrointestinal tract (**D**). The x-axis indicates the days post-onset of symptoms (**appendix 2 p2, Table S1**).

**E,** Individual-based patterns in thenumber of SARS-CoV-2-like reads in different types of clinical specimens. Data have been normalized to total sequencing reads in reads-per-million (RPM). The number of SARS-CoV-2-like (*Coronaviridae*) reads of different sample types is colored as follows: brown, throat swab; orange, nasal swab; yellow, sputum; blue, anal swab; green, feces. Gray, number of non-*Coronaviridae* viral reads.



**Supplementary figure 3: Highly abundant non-SARS-CoV-2 viruses identified in respiratory tract specimens**

**A,** Percentage of *Coronaviridae* reads and highly abundant non-SARS-CoV-2 viral species to the total viral reads. Only non-SARS-CoV-2 viruses with >10,000 mapped reads and ranked in top 1 of all non-SARS-CoV-2 viruses are shown (See details in **appendix 2 p6, Table S5**). Viral species are colored according to their natural hosts: green, animals; pink, bacteria; light blue, plant; and purple, algae.

**B,** Depth and coverage of known respiratory viruses in four severe patients.Four representative specimens collected from four severe patients are co-detected (genome coverage>50%) with rhinovirus B (P09N205), human herpesvirus 1 (P01N201 and P05S207) and human orthopneumovirus (P13T211).

**C-D,** Individual-based changes inthenumber of SARS-CoV-2-like reads and other known respiratory viruses in respiratory tract specimens from P01 (**C**) and P13 (**D**). The x-axis indicates the days post-onset of symptoms. Data have been normalized to total sequencing reads in reads-per-million (RPM).



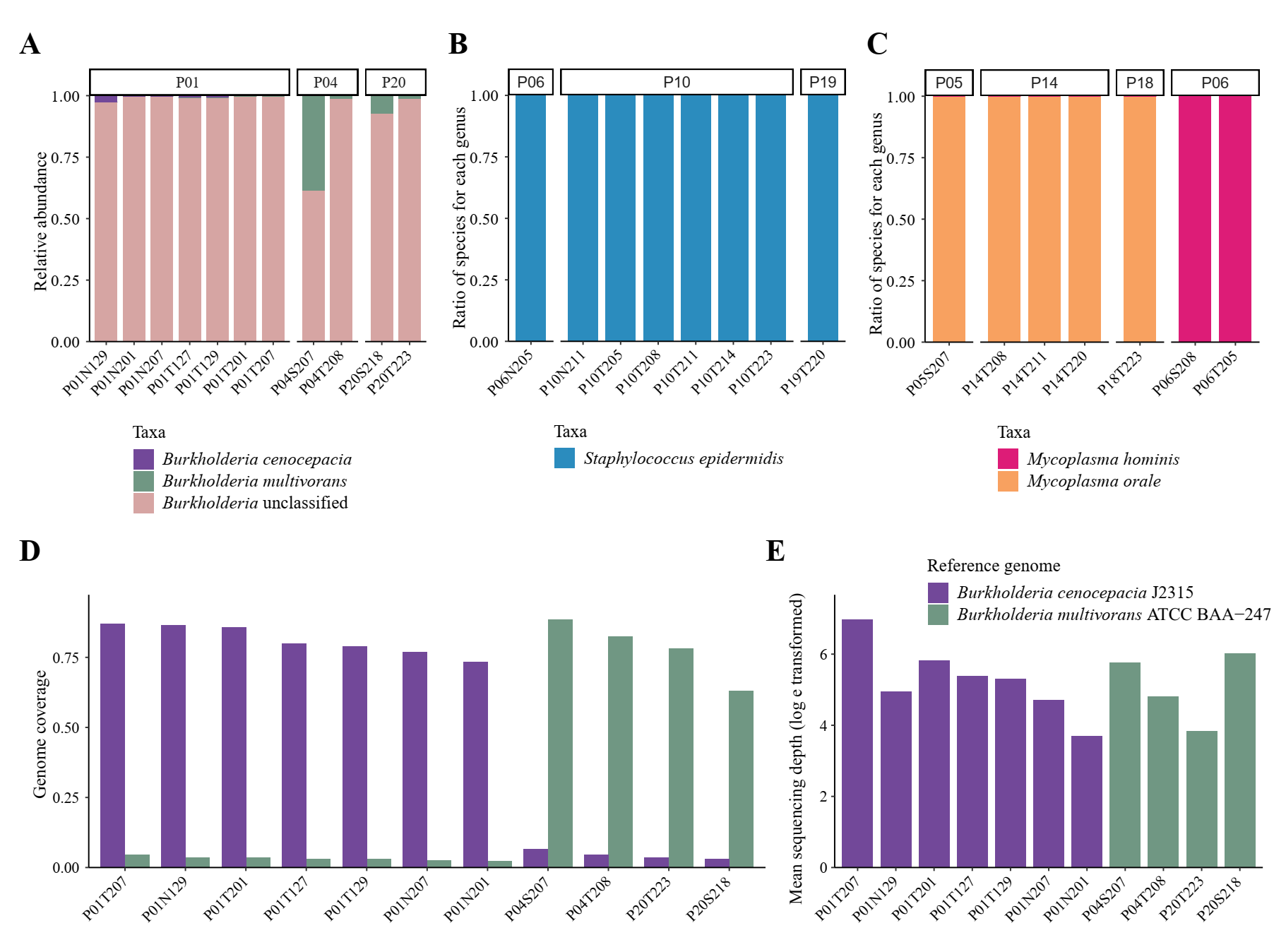
**Supplementary figure 4: Distinct respiratory microbial signatures in respiratory specimens of mild and severe cases**

**A,** Number of detected non-viral microbial genus and species in all respiratory specimens of mild (orange, n=7) and severe cases (brick red, n=40). \*\*\*, P < 0.001, Wilcoxon rank-sum test.

**B,** Non-metric multidimensional scaling (NMDS) plot displaying distinct respiratory microbial communities in specimens from mild (orange) and severe (brick red) cases. Manhattan distances based on presence/absence profile at the genus level are used for NMDS ordination.

**C**, Presence/absence profile of non-viral microbial genera in all respiratory specimens. Only genera presented in at least two samples are shown.

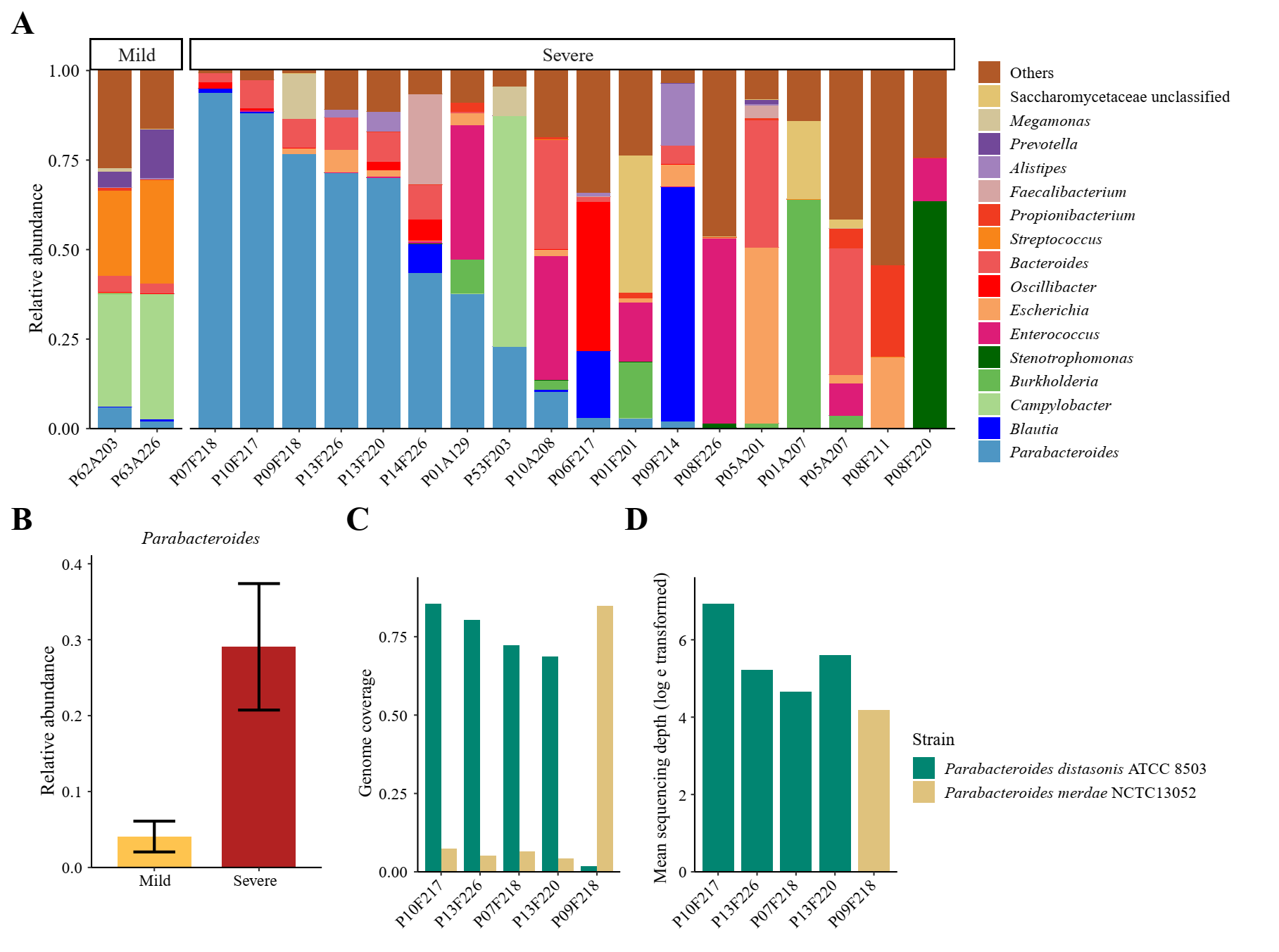
**D,** Bar plot showing the relative expression levels of non-viral microbes in all respiratory specimens.

**Supplementary figure 5: Species-level identification of genera *Burkholderia*, *Staphylococcus* and *Mycoplasma***

**A-C,** Bar plot showing the ratio of assigned species within *Burkholderia* (**A**), *Staphylococcus* (**B**) and *Mycoplasma* (**C**) by MetaPhlAn2. Only respiratory samples with mono-dominance of one single genus (relative abundance >60%) are shown.

**D,** Average coverage of two references of *Burkholderia cepacia*complex. References are shown if 10% or more of the reference genome is covered by reads in at least one sample (**appendix 2 p8, Table S7**).

**E**, Average sequencing depth (log e-transformed) of the two *Burkholderia* references. Purple, *Burkholderia cenocepacia* J2315,green, *Burkholderia multivorans* ATCC BAA-247*.*



**Supplementary figure 6: Gut microbial signatures in mildly and severely ill patients**

**A,** Bar plot showing the relative expression levels of non-viral microbes in all gastrointestinal tract specimens.

**B,** Relative expression levels of *Parabacteroides* between gut specimens from mild and severe cases. The bar chart and black error bars denote the mean and standard error values of expression levels in mild (orange) and severe (brick red) cases.

**C,** Average coverage of two references of genus *Parabacteroides*. References are shown if 10% or more of the reference genome are covered by reads in at least one sample (**appendix 2 p8, Table S7)**.

**D**, Average sequencing depth (log e-transformed) of the two *Parabacteroides* references. Green, *Parabacteroides distasonis* ATCC 8503*,* yellow, *Parabacteroides merdae* NCTC13052*.*