

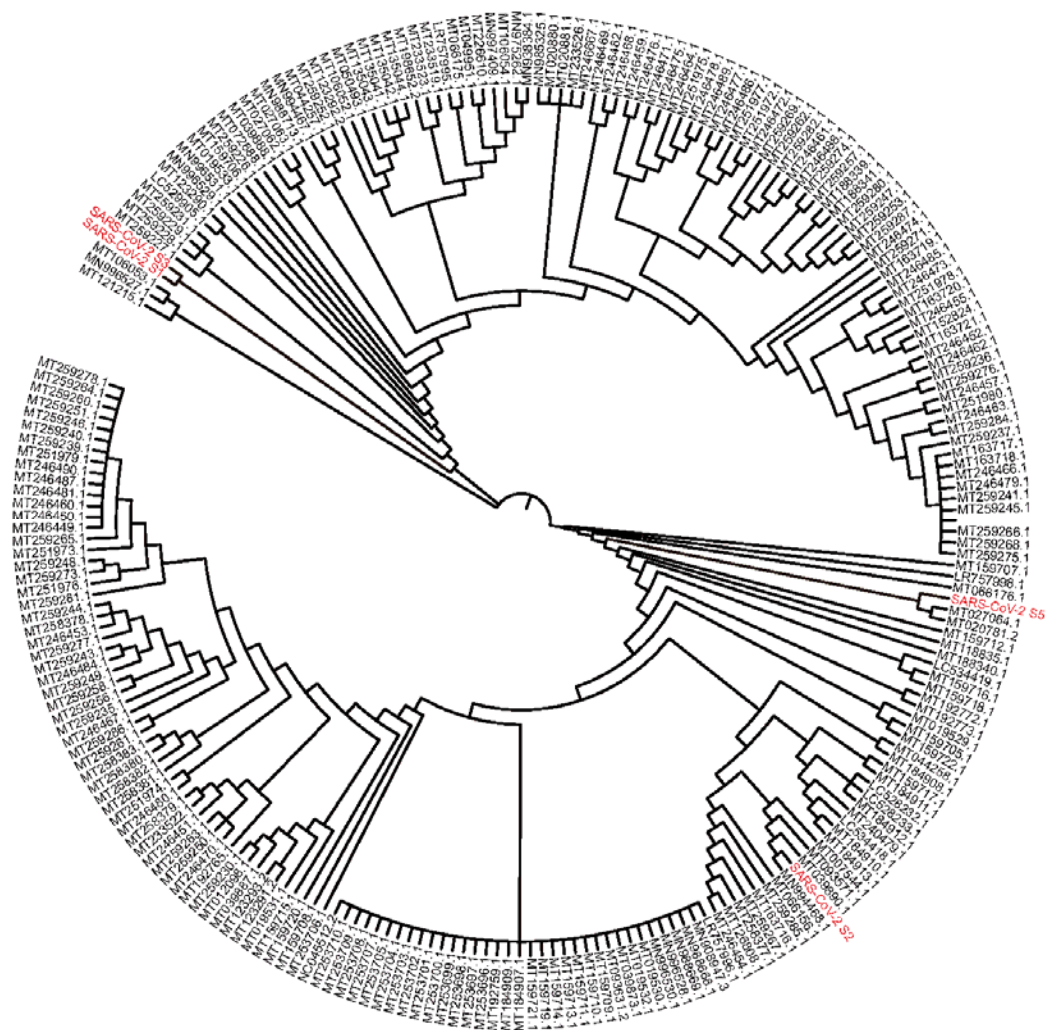
## Electronic Supplementary Material

### Enriched Opportunistic Pathogens Revealed by Metagenomic Sequencing Hint Potential Linkages between Pharyngeal Microbiota and COVID-19

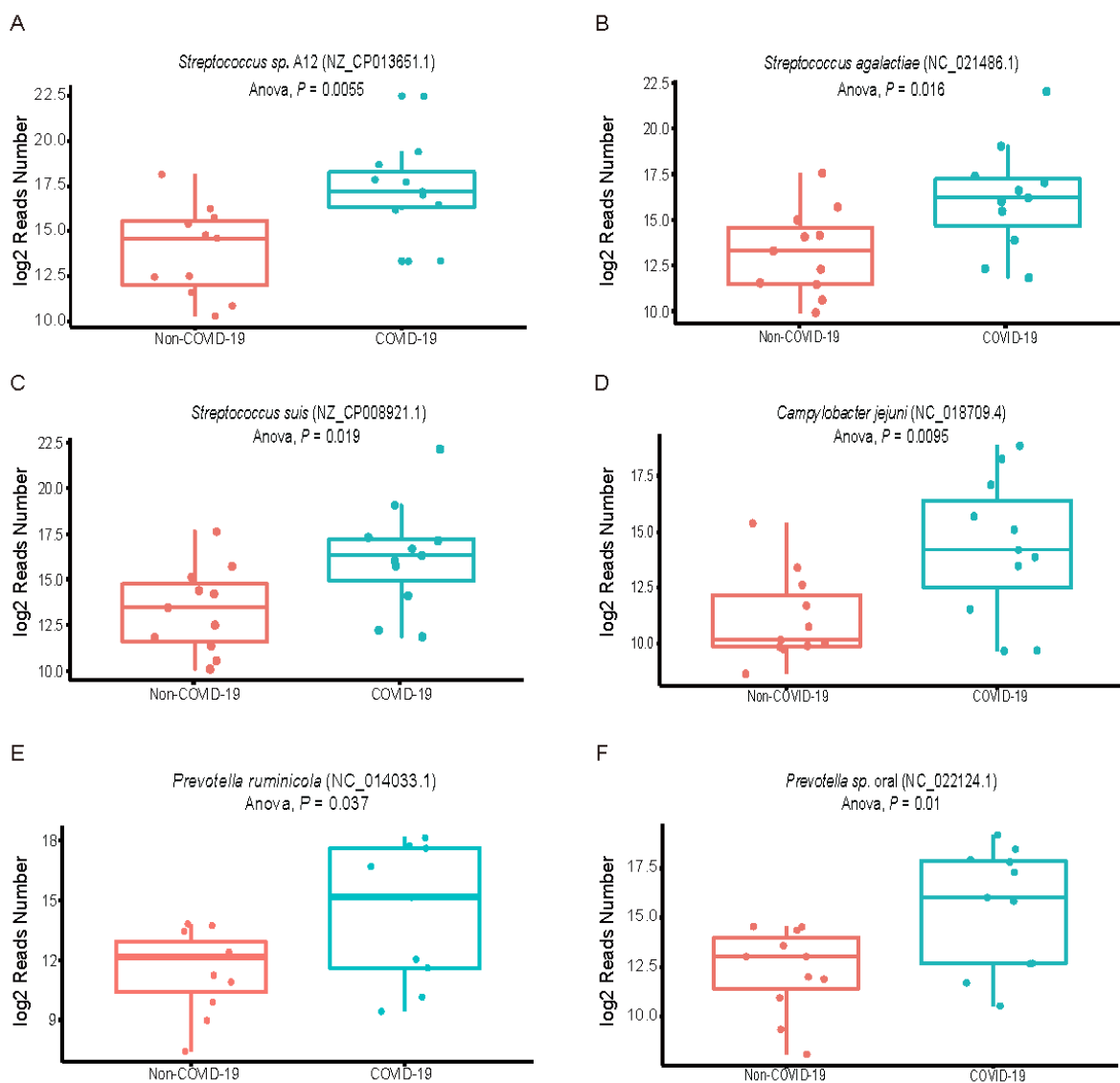
Dongyan Xiong<sup>1,2</sup> • Caroline Muema<sup>1,2</sup> • Xiaoxu Zhang<sup>1,2</sup> • Xinming Pan<sup>3</sup> • Jin Xiong<sup>1</sup> • Hang Yang<sup>1</sup> • Junping Yu<sup>1</sup>✉ • Hongping Wei<sup>1</sup>✉

1. CAS Key Laboratory of Special Pathogens and Biosafety, Centre for Biosafety Mega-Sciences, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China;
3. Jiangxia District Center for Disease Control and Prevention, Wuhan 430200, China

Supporting information to DOI: 10.1007/s12250-021-00391-x



**Fig. S1** Phylogenetic tree constructed with the four high quality SARS-CoV-2 (Red font marked in the figure) genomes assembled from the transcriptome data and 232 reference genomes (Accession ID could be seen in figure).



**Fig. S2** One-way analysis of variance for the number of pathogen reads identified in metagenomic profile of the COVID-19 group ( $n = 11$ ) and the Non-COVID-19 group ( $n = 11$ ). ANOVA for the reads number of *Streptococcus* sp. A12 (**A**,  $P = 0.0055$ ), *Streptococcus agalactiae* (**B**,  $P = 0.016$ ), *Streptococcus suis* (**C**,  $P = 0.019$ ), *Campylobacter jejuni* (**D**,  $P = 0.0095$ ), *Prevotella ruminicola* (**E**,  $P = 0.037$ ) and *Prevotella* sp.oral (**F**,  $P = 0.01$ ) under two cohorts.

**Table S1** Open access number of the healthy control group

Sample name in this study	Open access number
H1	SRS013948
H2	SRS014689
H3	SRS019027
H4	SRS019127
H5	SRS065335
H6	SRS097905
H7	SRS147151

\* Website of the Public database: [www.hmpdacc.org/hmp/resources/](http://www.hmpdacc.org/hmp/resources/).

**Table S2** Basic information about the clinical samples

Sample Name	Age	Gender	Ct Value
S1	52	F	17.5
S2	47	M	17.1
S3	24	F	23.1
N1	29	M	NA
N2	43	M	NA
S4	38	F	26.2
N3	28	F	NA
S5	56	F	23.6
N4	48	F	NA
N5	25	F	NA
N6	81	M	NA
N7	86	M	NA
S6	39	F	28.4
S7	75	M	26.3
N8	75	F	NA
N9	76	F	NA
S8	30	M	28.6
S9	43	M	31.2
N10	36	M	NA
S10	47	F	26.4
N11	44	F	NA
S11	51	F	25.7

Student's t test for age:  $t = 0.77439$ ,  $df = 16.213$ ,  $P\text{-value} = 0.4498$

Chi-square test for gender:  $X\text{-squared} = 0$ ,  $df = 1$ ,  $P\text{-value} = 1$

**Table S3** Number of metagenomic reads and contigs in the samples

Sample	Number of reads (clean data after removing host reads)	Number of contigs
S1	12175924	98995
S2	5891800	51849
S3	17318610	47011
N1	14184292	81812
N2	3063968	3420
S4	22649030	144432
N3	12677392	21451
S5	6527112	47514
N4	7711834	39090
N5	7655632	81660
N6	4626552	37694
N7	2455370	19928
S6	1984086	2405
S7	28178082	63413
N8	6704736	42279
N9	5282554	30612
S8	45350920	103557
S9	3027936	23473
N10	1164170	6470
S10	1309570	14639
N11	3495330	55679
S11	3330608	23469
H1	28929212	92749
H2	20428870	80047
H3	26393494	106997
H4	40539396	101926
H5	6384956	57305
H6	86104896	127699
H7	27452310	99841

**Table S4** Total microbiome abundance profile including SARS-CoV-2.**Table S5** Eighty-one species with differential abundance.