Supplementary Bioinformatics Methods

For the 16S ribosomal RNA gene sequence analysis, read pairs were trimmed and filtered for quality with BBDuk v38.90 [1] and then processed with DADA2 v1.20.0 [2] for denoising, merging, and taxonomic assignment of the resulting amplicon sequence variants (ASV) with rdp and the SILVA v138.1 database.[3] ASVs unable to be identified by SILVA were further queried against NCBI's 16S ribosomal RNA database using BLAST+ v2.12.0. Differential species abundance testing was performed using DESeq2 [4] contrasting the SARS-CoV-2 positive and control samples, accounting for the paired sampling design with p-values adjusted by the Benjamini-Hochberg method for multiple comparisons. ASV counts were aggregated at the species level and averaged for each pair of matched controls prior to testing.

For SARS-CoV-2 genome sequencing, RNA from the stool was extracted using the QIAamp Viral RNA Mini Kit 250 (Qiagen), and amplified following the CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel method. The amplified RNA was transformed to cDNA and sequenced on 2x75bp Nextseq platform. For genome assembly, raw reads were trimmed and quality filtered using BBDuk v38.90 [1] and assembled with a reference-mapping-based approach using the Wuhan-Hu-1 reference genome. Additional details of the bioinformatic pipeline can be found here: https://github.com/TheDBStern/viral-assembly-variant-calling.

References

1. Bushnell B. BBTools. DOE Joint Genome Institute [accessed November 16th 2021.
2. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *Isme j* 2017;11(12):2639-43. doi: 10.1038/ismej.2017.119 [published Online First: 2017/07/22]

3. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;41(Database issue):D590-6. doi: 10.1093/nar/gks1219 [published Online First: 2012/11/30]
4. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15(12):550. doi: 10.1186/s13059-014-0550-8 [published Online First: 2014/12/18]