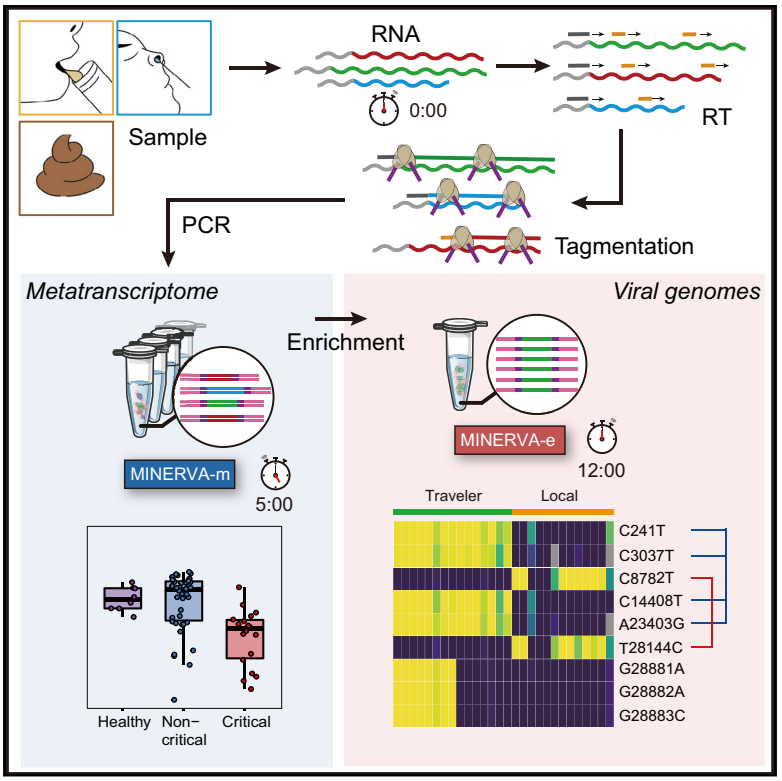
Molecular Cell

MINERVA: A Facile Strategy for SARS-CoV-2 Whole-Genome Deep Sequencing of Clinical Samples

Graphical Abstract



Highlights

- A facile and robust approach for transcriptomic sequencing of COVID-19 samples
- Metagenomic signatures of COVID-19 samples
- Better SARS-CoV-2 genome coverage compared with conventional strategies
- Facilitate multiple facets of COVID-19 research

Authors

Chen Chen, Jizhou Li, Lin Di, ..., Hui Zeng, Yanyi Huang, Jianbin Wang

Correspondence

angelawu@ust.hk (A.R.W.), zenghui@ccmu.edu.cn (H.Z.), yanyi@pku.edu.cn (Y.H.), jianbinwang@tsinghua.edu.cn (J.W.)

In Brief

The novel coronavirus disease 2019 (COVID-19) pandemic poses a serious public health risk. Chen et al. develop a facile and robust approach for transcriptomic sequencing of COVID-19 samples that will facilitate molecular epidemiology studies during current and future outbreaks.

Chen et al., 2020, Molecular Cell *80*, 1123–1134
December 17, 2020
https://doi.org/10.1016/j.molcel.2020.11.030

CellPress

Molecular Cell

CellPress

Technology

MINERVA: A Facile Strategy for SARS-CoV-2 Whole-Genome Deep Sequencing of Clinical Samples

Chen Chen,[1,12] Jizhou Li,[2,12] Lin Di,[3,4,12] Qiuyu Jing,[5,12] Pengcheng Du,[1,12] Chuan Song,[1] Jiarui Li,[1] Qiong Li,[2] Yunlong Cao,[3] X. Sunney Xie,[3] Angela R. Wu,[5,6,7,*] Hui Zeng,[1,*] Yanyi Huang,[3,8,9,10,*] and Jianbin Wang[2,10,11,13,*]

[1]Institute of Infectious Diseases, Beijing Ditan Hospital, Capital Medical University and Beijing Key Laboratory of Emerging Infectious Diseases, Beijing 100015, China

[2]School of Life Sciences, Tsinghua University, Beijing 100084, China

[3]Beijing Advanced Innovation Center for Genomics (ICG), Biomedical Pioneering Innovation Center (BIOPIC), Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

[4]School of Life Sciences, Peking University, Beijing 100871, China

[5]Division of Life Science, Hong Kong University of Science and Technology, Hong Kong SAR, China

[6]Department of Chemical and Biological Engineering, Hong Kong University of Science and Technology, Hong Kong SAR, China

[7]Hong Kong Branch of Guangdong Southern Marine Science and Engineering Laboratory (Guangzhou), The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China

[8]College of Chemistry and Molecular Engineering, Beijing 100871, China

[9]Institute for Cell Analysis, Shenzhen Bay Laboratory, Guangdong 518132, China

[10]Chinese Institute for Brain Research (CIBR), Beijing 102206, China

[11]Beijing Advanced Innovation Center for Structural Biology (ICSB), Tsinghua University, Beijing 100084, China

[12]These authors contributed equally

[13]Lead Contact

*Correspondence: angelawu@ust.hk (A.R.W.), zenghui@ccmu.edu.cn (H.Z.), yanyi@pku.edu.cn (Y.H.), jianbinwang@tsinghua.edu.cn (J.W.)

https://doi.org/10.1016/j.molcel.2020.11.030

SUMMARY

Analyzing the genome of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from clinical samples is crucial for understanding viral spread and evolution as well as for vaccine development. Existing RNA sequencing methods are demanding on user technique and time and, thus, not ideal for time-sensitive clinical samples; these methods are also not optimized for high performance on viral genomes. We developed a facile, practical, and robust approach for metagenomic and deep viral sequencing from clinical samples. We demonstrate the utility of our approach on pharyngeal, sputum, and stool samples collected from coronavirus disease 2019 (COVID-19) patients, successfully obtaining whole metatranscriptomes and complete high-depth, high-coverage SARS-CoV-2 genomes with high yield and robustness. With a shortened hands-on time from sample to virus-enriched sequencing-ready library, this rapid, versatile, and clinic-friendly approach will facilitate molecular epidemiology studies during current and future outbreaks.

INTRODUCTION

As of November 4, 2020, the ongoing coronavirus disease 2019 (COVID-19) viral pandemic has affected more than 46 million people in over 200 countries and territories around the world and has claimed more than 1,204,000 lives (WHO, 2020). Closely monitoring the genetic diversity and distribution of viral strains at the population level is essential for epidemiological tracking and for understanding viral evolution and transmission; additionally, examining the viral heterogeneity within a single individual is imperative for diagnosis and treatment (Wölfel et al., 2020). The disease-causing pathogen, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was identified from early disease cases, and its draft genome was sequenced within weeks because of rapid responses from researchers around the world (Lu et al., 2020c; Ren et al., 2020; Wu et al., 2020; Zhou et al., 2020b). The initial SARS-CoV-2 draft genome was obtained independent from the same early COVID-19 patient samples using various conventional RNA sequencing (RNA-seq) library construction methods. Although these library construction methods successfully generated a draft genome, several drawbacks hinder use of these methods for routine viral genome sequencing from the surge of clinical samples during an outbreak.

One direct library construction approach used to generate the SARS-CoV-2 draft genome (Lu et al., 2020c; Ren et al., 2020; Wu et al., 2020; Zhou et al., 2020b) essentially captures each sample's entire metatranscriptome, in which SARS-CoV-2 is just one species among many. The abundance of SARS-CoV-2 in clinical swabs, sputum, and stool samples is often low (Wang et al., 2020; Wölfel et al., 2020); therefore, this catch-all method

requires deeper sequencing of each sample to obtain sufficient coverage and depth of the whole viral genome, which increases the time and cost of sequencing. Target enrichment with spiked-in primers can improve SARS-CoV-2 genome coverage (Deng et al., 2020), but reliance on specific primers inherently limits this approach for profiling of evolving viruses. The same limitation applies to multiplex RT-PCR-based strategies (Xiao et al., 2020). Additionally, when the sample is subjected to targeted amplification during the initial RT steps, its metatranscriptomic information is lost forever.

Currently, the most comprehensive strategy is a combination of metatranscriptomics profiling with post-library SARS-CoV-2 target enrichment (Xiao et al., 2020). However, in most conventional RNA-seq methods, the double-strand DNA ligation (dsDL) portion of the protocol is usually the most demanding regarding hands-on time and user technique (Di et al., 2020). When superimposed on the target enrichment process, these labor-intensive and lengthy protocols become impractical for routine use in the clinic and for timely monitoring of viral genetics and evolution on large volumes of samples during an outbreak. Furthermore, because of the low molecular efficiency of dsDL, these protocols also require a high amount of input material, further restricting their application to clinical samples.

Although next-generation sequencing platforms are high throughput and have a short turn-around time, library construction from samples, including targeted enrichment or not, remains a major bottleneck. To broadly apply viral sequencing to clinical samples, especially during outbreaks, when biomedical resources are already limited, a rapid, simple, versatile, and scalable sample library construction method that does not compromise performance is urgently needed.

Recently, we reported a new RNA-seq library construction strategy that aims to address some of these challenges: sequencing hetero RNA-DNA-hybrid (SHERRY) avoids the problematic dsDL step in library construction by taking advantage of the newly discovered Tn5 tagmentation activity on RNA/DNA hybrids to directly tag RNA/cDNA fragments with sequencing adapters (Di et al., 2020). Therefore, SHERRY has minimal sample transfer and greatly reduced hands-on time, making it simple, robust, and suitable for input ranging from single cells to 200 ng total RNA. We now combine the advantages of a tailored SHERRY protocol, which improves coverage of whole metatranscriptomes, with a simplified post-library target enrichment protocol. MetagenomIc RNA enrichment viral sequencing (MINERVA) is an easy-to-use, versatile, scalable, and cost-effective protocol that yields a high-coverage, high-depth SARS-CoV-2 genome while preserving the sample's rich metatranscriptomic profile. The hands-on time required from clinical sample to sequencing-ready library using conventional approaches without enrichment is 190 min; MINERVA requires only 100 min hands-on time, and when deep viral coverage is desired, an additional 90 min for post-library enrichment, totaling 190 min for the entire workflow (Figure S1A). MINERVA also costs less than dsDL because of its simpler procedure and much lower sequencing depth (Figure S1A). These features make MINERVA practical for high-volume, routine clinical use. We applied MINERVA to various types of COVID-19 samples and successfully obtained up to 10,000-fold SARS-CoV-2 genome enrichment. This strategy will facilitate all studies regarding SARS-CoV-2 genetic variations in the current pandemic and can also be applied to other pathogens of interest.

RESULTS

Minerva

To analyze metagenomics and SARS-CoV-2 genetics from COVID-19 patient samples, we developed a two-stage metagenomic RNA enrichment viral sequencing strategy named MINERVA (Figure 1A and Methods S1). First, we employed a SHERRY-based RNA-seq pipeline for metagenomic analysis (MINERVA-m). Because clinical samples may contain DNA, RNA, and possibly carrier RNA, MINERVA starts with ribosomal RNA (rRNA) removal and optional simultaneous carrier RNA removal, followed by DNase I treatment. The remaining RNA is then subject to standard SHERRY. Previously, we observed 3′ bias in SHERRY libraries; to address this, we used 10 ng mouse 3T3 cell total RNA as starting material and tested whether adding a random decamer (N10) during reverse transcription could improve coverage evenness (Figures 1B, 1C, and S1B–S1E). Compared with the standard SHERRY protocol, which uses 1 μM T30VN primer during reverse transcription, supplementation with 1 μM N10 in MINERVA indeed improves gene body coverage evenness, presumably by improving the reverse transcription efficiency. When the N10 concentration was increased further to 10 μM, we observed almost no coverage bias in the gene body. The high N10 concentration can result in an increased rRNA ratio in the product, sometimes as high as 90%, but MINERVA employs rRNA removal as the first step prior to reverse transcription, negating this problem. We also performed enzyme titration with homemade and commercial Tn5 transposomes. Based on these N10 and Tn5 titration results, we used 10 μM N10 during reverse transcription and 0.5 μL V50 for each 20-μL tagmentation reaction in all following experiments. The whole procedure, from nucleic acid to metagenomic sequencing-ready library, including wait time, takes 5.5 h (Figure S1A).

For target enrichment (MINERVA-e), we first quantified SARS-CoV-2 abundance in each metagenomic sequencing library using an N gene qPCR assay and pooled eight libraries based on quantification results. Then we performed standard in-solution hybridization on the pooled library with biotinylated RNA probes covering the whole viral genome. The enrichment procedure takes ~7–13 h; the entire MINERVA pipeline can be completed within ~12–18 h.

MINERVA Is Compatible with COVID-19 Samples

To evaluate its performance on clinical samples, we applied MINERVA to 136 samples collected from 91 individuals with COVID-19, with samples types including pharyngeal swabs, sputum, and stool. These individuals were admitted to Ditan Hospital within a 3-month period from January to April 2020, presenting different symptom severity (Figure 2A; Tables S1 and S2). Some individuals were re-sampled longitudinally to investigate temporal and intra-host viral heterogeneity. In addition to the samples from individuals with COVID-19, we also included different sample types from eight healthy individuals as well as non-template control (NTC) samples. We first tested the effect

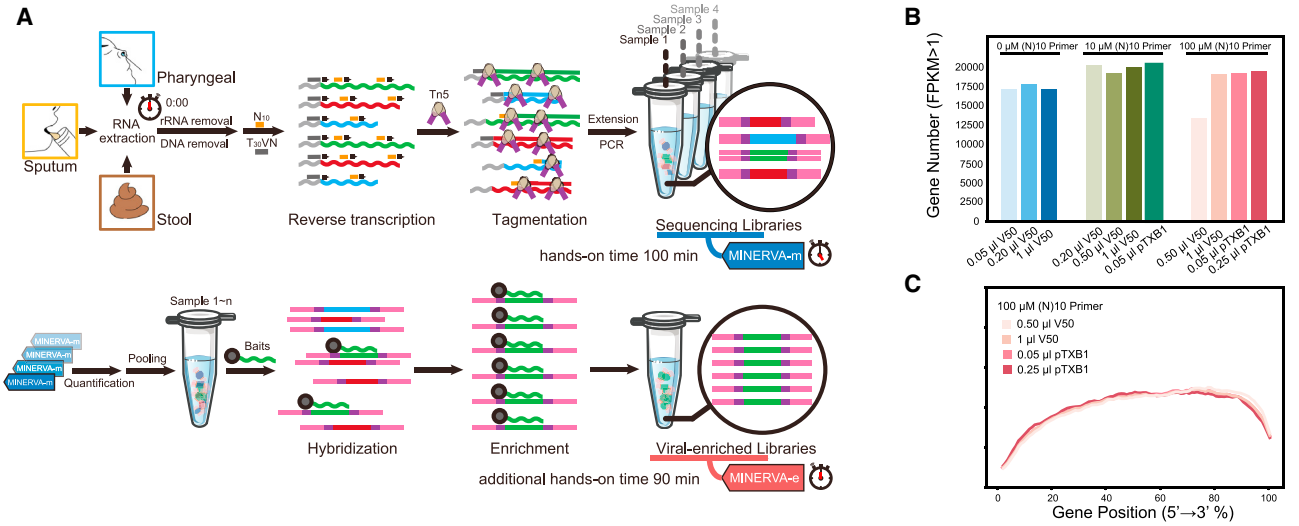


Figure 1. Scheme and Development of MINERVA

(A) RNA extracted from pharyngeal swabs and sputum and stool samples undergo rRNA and DNA removal before metagenomic sequencing library construction (MINERVA-m). Multiple libraries were then pooled for SARS-CoV-2 sequence enrichment.

(B) Effect of the N10 primer during reverse transcription and Tn5 amount on detected gene number.

(C) Effect of the N10 primer during reverse transcription and Tn5 amount on gene body coverage evenness.

of sample input volume on MINERVA results. Using just 2.7-μL of rRNA- and DNA-depleted sample led to satisfactory SARS-CoV-2 coverage, and scaling up the reaction volume and sample input further improved the MINERVA data quality (Figure 2B). Using the same samples and the same sequencing depth, more input in a higher reaction volume generated deeper SARS-CoV-2 genome coverage.

Carrier RNA, which is widely used in viral DNA/RNA extraction before qRT-PCR assays, severely affects high-throughput sequencing analysis. Therefore, most qRT-PCR-positive clinical samples are not amenable to further viral genetic studies. We explored the effect of adding polyT oligos during the rRNA removal step to simultaneously remove spike-in poly(A) RNA and carrier RNA. By incorporating this step in MINERVA, we successfully avoided overwhelming representation of unwanted RNA sequences while retaining desired metagenomic and SARS-CoV-2 information (Figures 2C and 2D).

MINERVA Captures Metagenomic Signatures of COVID-19 Samples

We benchmarked MINERVA against conventional dsDL strategies in head-to-head comparisons of the first 79 clinical samples sequenced. On average, we sequenced 2–5 gigabase pairs (Gbp) for each MINERVA-m and MINERVA-e library, and nearly 100 Gbp for each dsDL library (Figure S2A). The MINERVA-m and dsDL libraries were comparable: bacterial heterogeneity as measured by species richness and Shannon index was correlated between the two (Figures S2B and S2C). To ensure the accuracy of our metagenomic computational pipeline, we employed metagenomics and metagenome assembly strategies to profile microbial compositions of different types of samples. For assembly, we combined reads from samples of the same type for co-assembly, obtaining contigs for each sample type. After taxonomy assignment at the contig level, reads from each individual sample were mapped back to the co-assembled contigs to obtain the microbial composition of each individual sample. Overall, we found the results of microbial compositions profiled from metagenomic and assembly strategies to be comparable (Figures S2D–S2F). There were slightly more taxa identified with the metagenomic strategy, especially those with relatively lower abundance. Detailed metagenome assemblies and mapping statistics of each sample are outlined in the STAR Methods and listed in Tables S3 and S4.

To study the metagenomic signature of individuals with COVID-19, we first analyzed the data from NTC samples, which showed distinct microbial compositions and significantly lower species richness, suggesting that NTC samples contain very low-level biomass and have little influence on real samples (Figures S3A and S3B). We then assessed potential factors that may be associated with microbial compositions in all types of samples. Permutational multivariate analysis of variance (PERMANOVA) revealed significant association with sample types as well as disease severity (Figure S3C). To exclude the influence from sample type, we performed further analyses on separated sample types. Disease severity was found to have significant associations in pharyngeal and sputum samples, and no factors were found with significant effect on stool samples (Figure 3A).

We performed various analyses to explore the metagenomic signatures that correlate with disease severity. Principal coordinates analysis (PCoA) analysis based on Bray-Curtis distance among pharyngeal samples indicates differences between samples from individuals with COVID-19 and healthy control subjects as well as a difference between critical samples and samples from individuals with less severe disease status. NTC samples were also separated from other samples on the PCoA plot ($p < 0.001$ by PERMANOVA test) (Figure 3B). Based on this result,

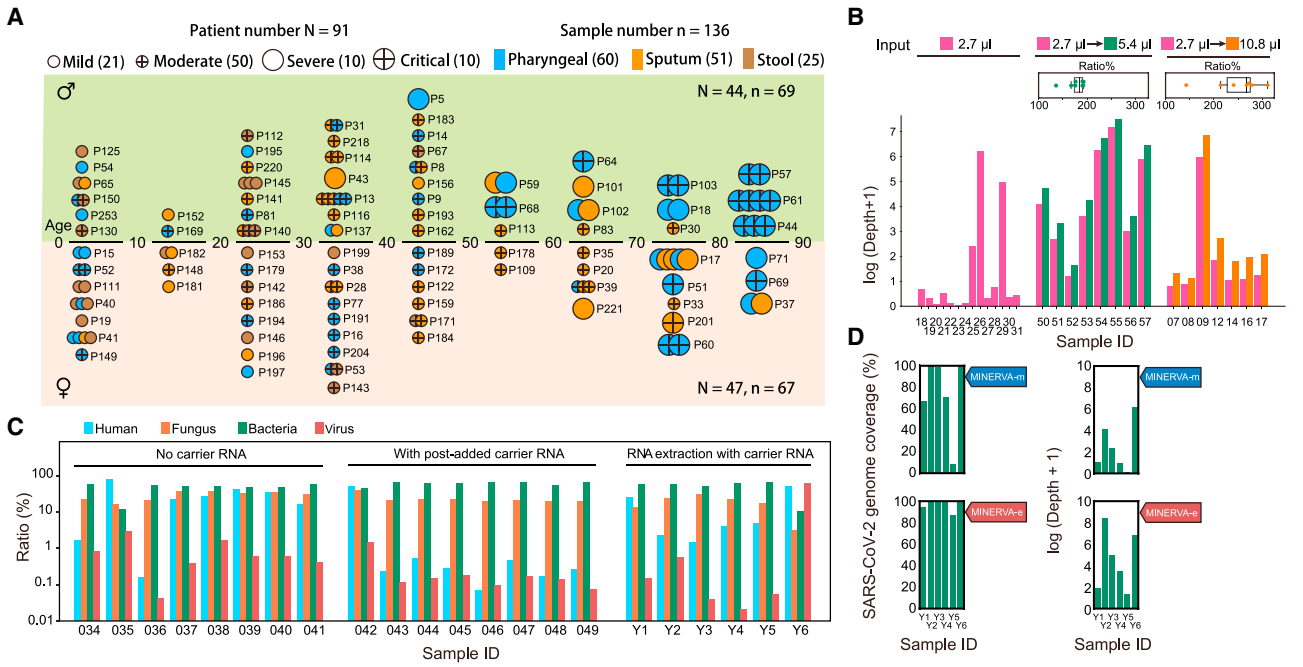


Figure 2. Optimization of MINERVA for Clinical Samples

(A) COVID-19 sample profiles, showing the age group, sex, severity, and re-sampling status of each individual.

(B) Effect of sample input and reaction volume on sequencing depth of the SARS-CoV-2 genome.

(C) Mapping ratios of human, fungus, bacterium, and virus reads showed good performance of carrier RNA removal. "No carrier RNA" refers to samples with no carrier RNA during extraction; "with post-added carrier RNA" refers to samples with post-added carrier RNA after RNA extraction. "RNA extraction with carrier RNA" refers to samples with carrier RNA during extraction. All samples with carrier RNA went through a carrier RNA removal step.

(D) SARS-CoV-2 genome coverage and depth of MINERVA-m and -e for "RNA extraction with carrier RNA" samples.

we segmented pharyngeal samples into healthy, non-critical (including "mild," "moderate" and "severe" samples) and critical groups. Decreased alpha diversities were observed in samples from individuals with COVID-19. A lower Shannon index was observed in the non-critical group compared with healthy control subjects. The most strongly and significantly decreased alpha diversity was observed in critical samples in terms of species richness and Shannon index (Figure 3C).

Furthermore, we applied multivariate analysis by using a generalized estimating equation (GEE) model to explore microbes associated with the difference among these three groups. To avoid introducing noise, especially for relatively low-abundance microbes, we set stringent filtering based on effect size and significance for the results, and only microbes with an absolute coefficient of more than 0.1 and Benjamini-Hochberg (BH) adjusted p value of less than 0.05 were kept. *Streptococcus*, *Rothia*, *Acinetobacter*, and *Acidovorax* were found to be associated with non-critical disease, whereas *Halomonas*, *Campylobacter*, *Prevotella*, and *Veillonella* were associated with critical disease (Figure 3D, left panel). Among these, *Streptococcus* and *Rothia* were highly enriched in the non-critical disease group, and *Prevotella* and *Veillonella* were abundant and prevalent in the healthy and non-critical groups. *Acinetobacter* and *Acidovorax* were enriched in the healthy and critical groups, whereas *Halomonas* was only highly enriched in critical samples (Figure 3D, right panel). Although some of these microbes were also detected in one of the NTC samples, their abundance was quite low compared with samples from individuals with COVID-19 (Figure S3I). They are also known to be abundant commensal microbes in human oral-pharyngeal samples (de Lastours et al., 2015; Human Microbiome Project Consortium, 2012; Zaura et al., 2009) and, therefore, were not removed as contaminants.

For sputum samples, PCoA analysis also showed separation between samples from individuals with COVID-19 and healthy control samples ($p < 0.001$ by PERMANOVA test), whereas samples from individuals with COVID-19 were not separated by disease severity (Figure S3D). The Shannon index was lower in samples from individuals with COVID-19 compared with healthy control samples, whereas no significant difference of species richness was observed (Figure S3E). GEE analysis found that *Streptococcus*, *Rothia*, and *Veillonella* were associated with disease (Figure S3F): *Streptococcus* and *Rothia* were enriched in in the disease group, whereas *Veillonella* was enriched and more prevalent in healthy samples. For stool samples, no difference was observed between samples from individuals with COVID-19 and healthy control subjects in terms of PCoA on Bray-Curtis distance (Figure S3G) and alpha diversities, including species richness and Shannon index (Figure S3H). These results were also consistent with PERMANOVA analysis (Figure 3A).

By surveying the metagenomic landscape of these samples, we also observed several patient samples with exceptionally

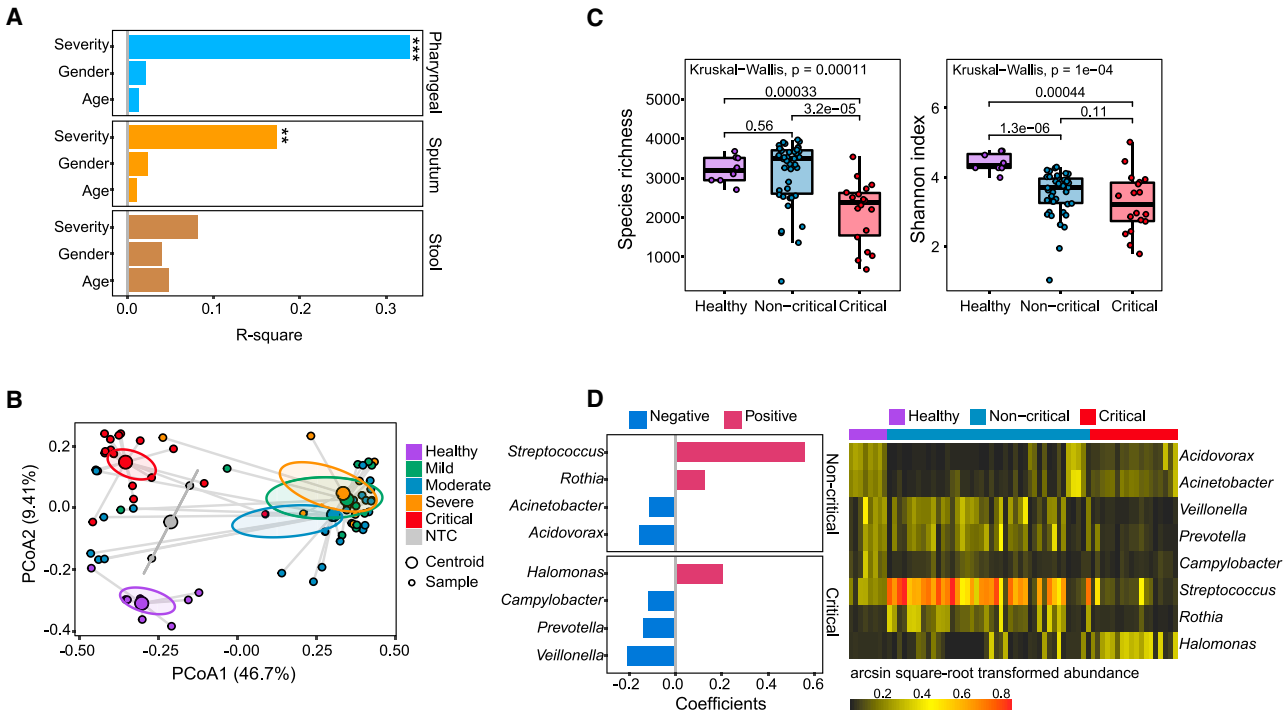


Figure 3. Metagenomics Analysis of COVID-19 Samples Using MINERVA

(A) PERMANOVA analysis highlights factors associated with microbial compositions in different sample types, including pharyngeal (n = 68), sputum (n = 59), and stool (n = 33). (PERMANOVA test, *p < 0.05, **p < 0.01, ***p < 0.001). Permutation was constrained within the same time point to account for repeated measures.

(B) PCoA analysis of pharyngeal samples based on Bray-Curtis distance, calculated at the bacterial genus, fungal genus, and viral family levels. Samples are colored according to different disease groups, including healthy controls (n = 8), mild (n = 10), moderate (n = 23), severe (n = 9), critical (n = 18), and non-template controls (NTCs; n = 2).

(C) Comparison of alpha diversity, including species richness (left panel) and Shannon index (right panel), among different groups of pharyngeal samples. Groups were segmented as healthy controls (n = 8), non-critical (including mild, moderate, and severe; n = 42) and critical (n = 18). Kruskall-Wallis test and Wilcoxon rank-sum test were used for multi-group and two-group comparisons, respectively.

(D) Analysis of association of microbial taxa with disease severity using pharyngeal samples. The generalized estimating equation (GEE) model was applied. Results were filtered based on significance (BH-adjusted p < 0.05) and effect size (absolute coefficient > 0.1). Taxa found to be significantly associated with disease are shown in the left panel, and their abundance in different groups of samples is shown in the right panel.

high abundance of additional known pathogens, including *Candida albicans*, *Staphylococcus aureus*, *Corynebacterium jeikeium*, *Corynebacterium striatum*, and *Klebsiella aerogenes* (Figure 4A, left panels). To validate these results, we directly mapped the reads to the individual genome of each species and assessed the genome coverage of each sample, normalized by sequencing depth (Figure 4A, right panel). Generally, the metagenomics results were consistent with the direct mapping results and also consistent with the assembly results (Figure S4). The occurrence rate of these species with high abundance identified by metagenomics and direct mapping is also associated with disease severity (Figure 4B). For severe viral pneumonia, co-infections can greatly affect patient outcome; one recent study showed that 50% of patients with COVID-19 who died in this pandemic had secondary bacterial infections (Cox et al., 2020; Crotty et al., 2015; Shah et al., 2016). We were unable to retrospectively confirm clinical co-infection in the cases identified by MINERVA; however, the sparse occurrence of these pathogenic microbes, combined with their high abundance, raises concerns and can be investigated in future studies.

MINERVA Achieves Better SARS-CoV-2 Genome Coverage Compared with Conventional dsDL Strategies

In MINERVA-m and dsDL data, we detected low but significant levels of SARS-CoV-2 sequences. The viral ratio is between 10^{-7} and 10^{-1}. It is worth noting that the SARS-CoV-2 sequence ratio is higher in MINERVA-m data than in dsDL data (Figures 5A and 5B), suggesting that MINERVA-m libraries capture more SARS-CoV-2 sequences. Although SARS-CoV-2 genome coverage and depth were not high in MINERVA-m results because of low viral ratios and low sequencing depth, performing MINERVA-e subsequently can enrich the SARS-CoV-2 sequence ratio up to 10,000-fold (Figures 5C and S5A–S5C). To evaluate potential false positive SARS-CoV-2 signals from targeted enrichment, we performed MINERVA-e on several 3T3 total RNA samples and three types of samples from eight healthy donors. The results showed extremely low coverage of the SARS-CoV-2 genome from a few duplicated amplicons in 3T3 data (Figure S5D). We observed low-level false-positive SARS-CoV-2 signals in a few healthy donor samples, likely because of sequencing index hopping, but the coverage depth is generally lower than in patient samples (Figure S5E). In

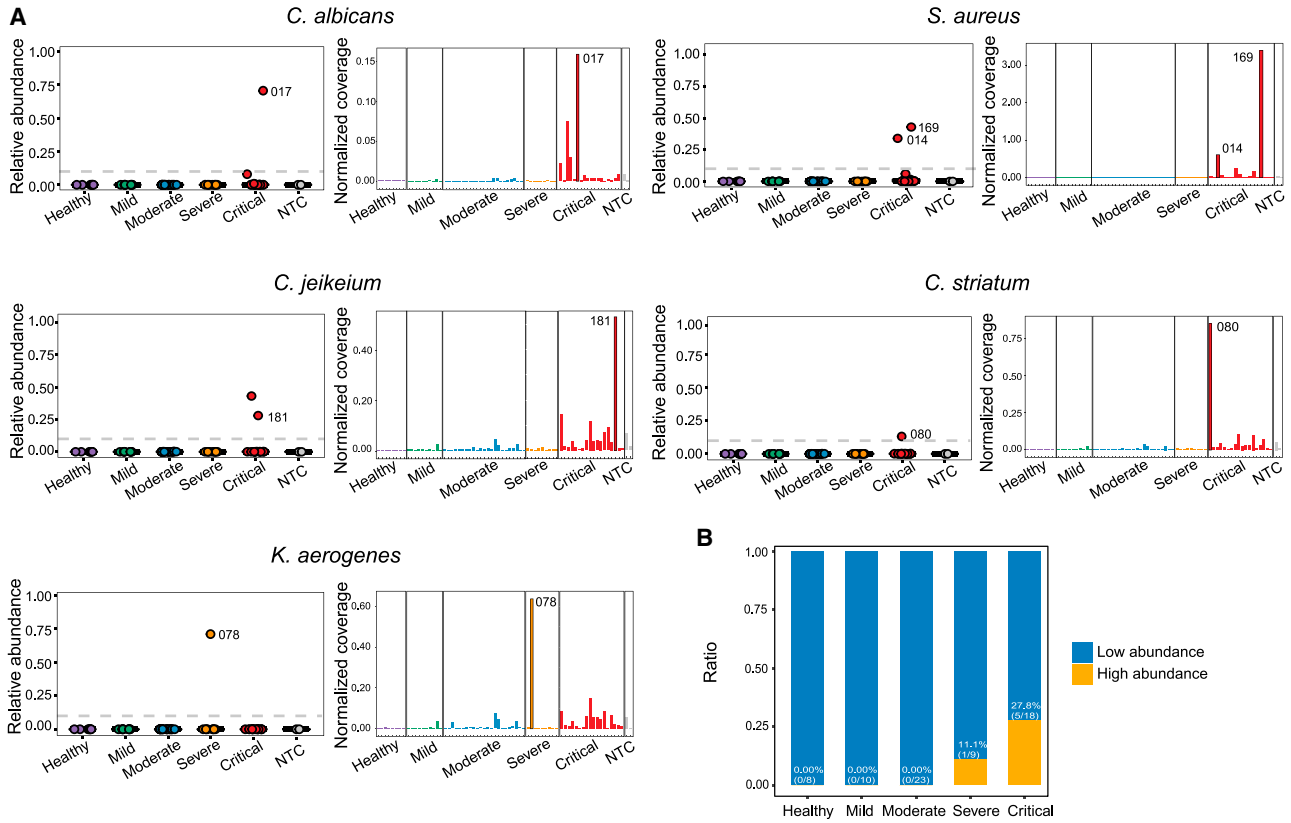


Figure 4. Co-existence of Additional Pathogens in Samples from Individuals with COVID-19

(A) Abundance of other potential pathogens. Sparse occurrence of high abundance of several pathogens with the potential to cause secondary infections was identified by metagenomics analysis (presented as relative abundance, left panels) and validated by direct mapping to their genomes (presented as coverage normalized by sequencing depth, right panels). Identified pathogens include *Candida albicans*, *Staphylococcus aureus*, *Corynebacterium jeikeium*, *Corynebacterium striatum*, and *Klebsiella aerogenes*.

(B) Occurrence rate of high-risk pathogens in different severity groups (samples with one or several high-risk pathogens identified were all considered). Only samples with a high abundance of these pathogens identified by metagenomics analysis and direct mapping were considered here and are also labeled in (A). The occurrence rate was associated with disease severity.

summary, MINERVA gives more complete and deeper coverage of SARS-CoV-2 genomes (Figures 5D and 5E), despite sequencing dsDL libraries to two orders of magnitude more depth (Figure S2A).

The superior quality of MINERVA data became clearer when we included clinical qRT-PCR results. The dsDL and MINERVA libraries detect SARS-CoV-2 sequences for samples with various cycle threshold (Ct) values, but MINERVA produced more complete and deeper genome coverage than dsDL methods (Figures 5F and 5G), and this advantage is more pronounced for low-viral-load samples, including two samples with negative qPCR results, and stool samples. By studying the relationship between SARS-CoV-2 qPCR results and read ratio, we identified two groups of samples that resulted in low SARS-CoV-2 genome coverage when processed using dsDL (Figure 5H). The first group had a low SARS-CoV-2 read ratio, which prohibited acquisition of enough SARS-CoV-2 sequencing reads. The second group, which included most stool samples, had relatively high SARS-CoV-2 Ct values and read ratios, suggesting that these samples had low total nucleic acid amounts. Because dsDL approaches are less sensitive and require more input, this may explain why MINERVA outperforms dsDL most evidently in stool samples.

MINERVA Can Facilitate Multiple Facets of COVID-19 Research

As a novel virus, little is known about the evolutionary features of SARS-CoV-2. Using 136 samples, we constructed a SARS-CoV-2 mutational profile (Figure 6A) that was distinct from the Guangdong profile (Lu et al., 2020b). A few mutation sites, including the two linked to the S and L strains (Lu et al., 2020a), were found in multiple samples. Aided by the deep genome coverage in MINERVA data, we not only detected strong linkage between position 8,782 and 28,144 but also observed high concordance of allele frequencies between these two positions. Furthermore, we detected strong linkage and high allele frequency concordance among four other positions: 241, 3,037. 14,408, and 23,403. Such allele frequency information offers additional layers of evidence supporting co-evolution of positions within the SARS-CoV-2 genome in two distinct groups of samples. It is worth noting that, in some

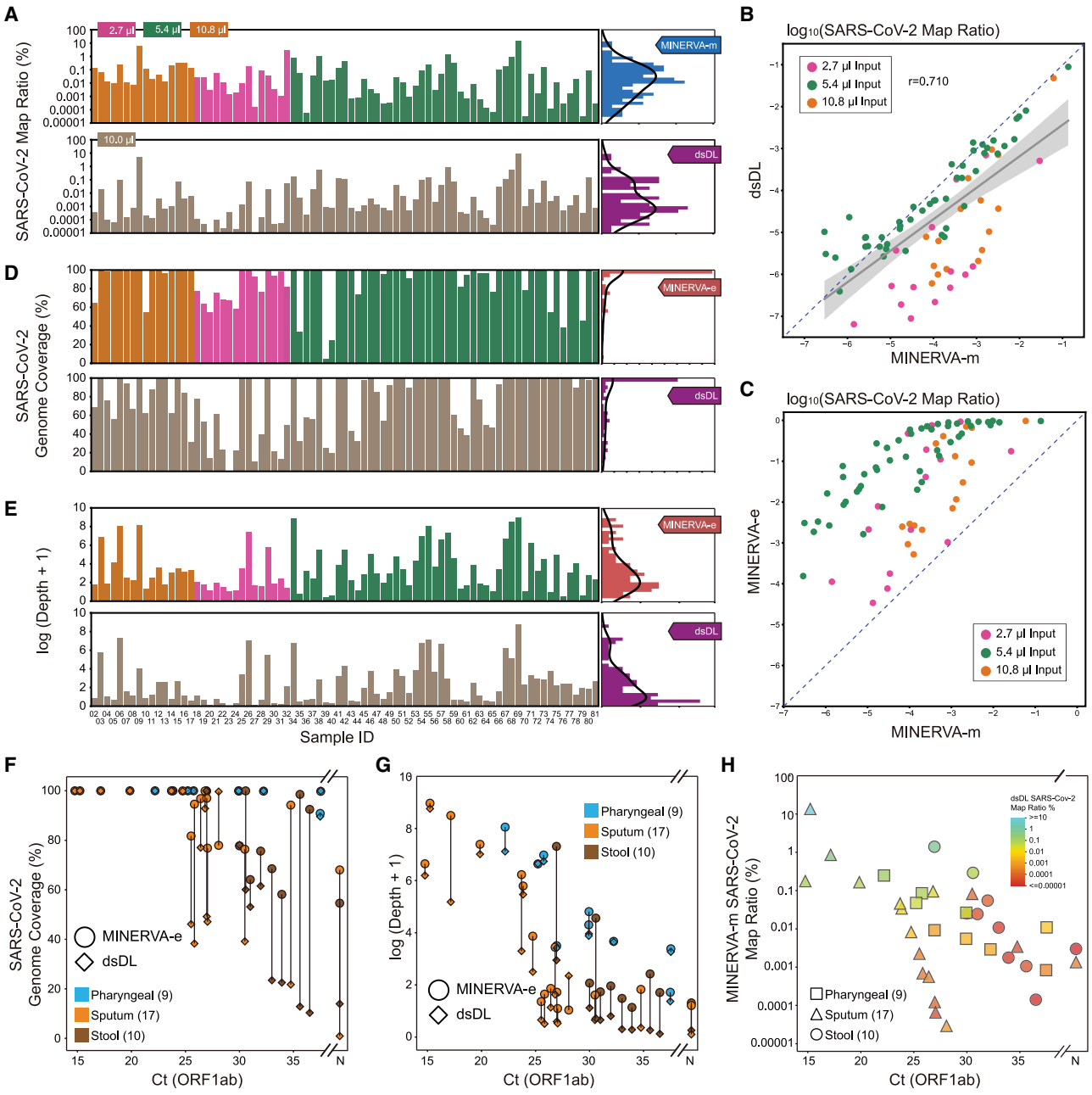


Figure 5. Direct Comparison of Sequencing Libraries Constructed from MINERVA and Conventional dsDL Strategies

(A) SARS-CoV-2 mapping ratio statistics of the MINERVA-m and dsDL libraries.

(B) Comparison of SARS-CoV-2 mapping ratios between the MINERVA-m and dsDL libraries.

(C) Comparison of SARS-CoV-2 mapping ratios between the MINERVA-m and MINERVA libraries.

(D and E) SARS-CoV-2 genome coverage and depth statistics of the MINERVA-e and dsDL libraries.

(F and G) Comparison of SARS-CoV-2 sequencing results between the MINERVA-e and dsDL libraries.

(H) Metagenomic sequencing and qPCR result features of samples with poor SARS-CoV-2 genome coverage.

samples, not all linked alleles are detected simultaneously because of low coverage at some positions; these alleles can indeed be observed at low coverage in the raw data for these samples but are missing from the post-processing data because they do not pass the stringent quality filtering steps. Nonetheless, the linkage was established by observing such linkage over many samples.

Apart from its high infectiousness, containment of SARS-CoV-2 transmission is challenging because of the existence of asymptomatic infected individuals (Bai et al., 2020). Although

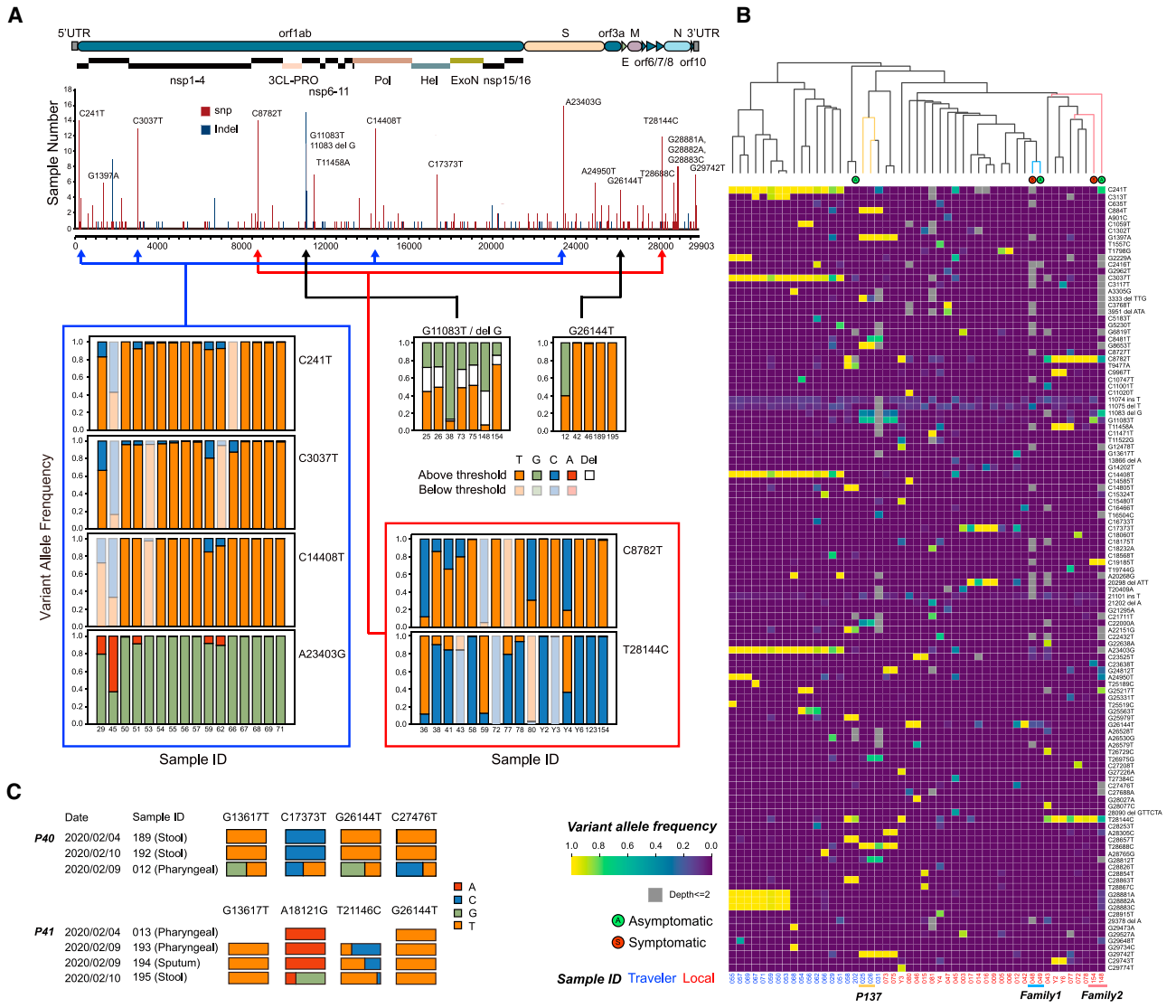


Figure 6. MINERVA Could Facilitate COVID-19 and SARS-CoV-2 Research through Accurate and Sensitive Identification of Viral Mutations

(A) SARS-CoV-2 mutation profile obtained from 136 samples.

(B) SARS-CoV-2 mutation profiles of asymptomatic individuals with COVID-19 and their infected family members. Individual origin is labeled in red (local) or blue (international traveler).

(C) Longitudinal SARS-CoV-2 mutation analysis of individuals with COVID-19.

qRT-PCR can be used to identify these individuals, elucidation of the chain of transmission requires complete SARS-CoV-2 sequences. To evaluate the performance of MINERVA when tracking SARS-CoV-2 transmission, we sequenced samples from several asymptomatic individuals and infected family members. SARS-CoV-2 SNV analysis could separate local patients from international travelers. Asymptomatic individuals each harbor viral sequences with unique signatures, and these individuals are clustered by viral SNV signature with their respective family members rather than other individuals with COVID-19, which indicates that viral SNVs within infected families are similar to each other and unique from other families (Figure 6B). Summarily, despite the asymptomatic phenotype of some infected individuals, the viral SNV signature generated by MINERVA can be used to accurately place these individuals in the chain of transmission, enabling better epidemiological tracking.

Recent studies have identified genetic variations of SARS-CoV-2 and raised the possibility that multiple variants could co-exist in the same host individual. The intra-host SNVs (iSNVs) detected in many samples (Figure 6A) suggest that SARS-CoV-2 is evolving within hosts (Wölfel et al., 2020). Through longitudinal sampling, we confirmed that iSNVs were generally relatively stable across time and body sites (Figure S6) but found that some patients harbored greater variations in iSNVs (Figure 6C). In P40 and P41, iSNVs were stable within the same sample type

across time but varied across different sample types. These results support co-existence of multiple SARS-CoV-2 variants in the same individual, and further investigation is warranted to understand this phenomenon.

In summary, MINERVA effectively converts metagenomes and SARS-CoV-2 sequences into sequencing libraries with a simple and quick experimental pipeline, and subsequent target enrichment can further improve SARS-CoV-2 genome coverage and genetic variation detection. MINERVA can facilitate study of SARS-CoV-2 genetics and can be implemented easily to fight future RNA pathogen outbreaks.

DISCUSSION

As of today, our knowledge of SARS-CoV-2 is still preliminary, and much of it is extrapolated from past studies of other betacoronaviruses such as SARS-CoV and MERS-CoV. However, the epidemiology, physiology, and biology of COVID-19 are evidently unique (Fauci et al., 2020). To speed up our investigation of this virus and the disease it causes, a practical protocol for viral genome research of clinical samples is urgently needed. Currently, methods for transforming clinical samples into sequencing libraries are laborious and painstaking, and clinical personnel at the frontlines are already strained for time and energy. MINERVA minimizes the need for expert technique and hands-on operation; we believe it will be pivotal in accelerating clinical research of SARS-CoV-2.

Recent evolutionary tracing studies suggest emergence of multiple novel, evolved subtypes of SARS-CoV-2 (Gudbjartsson et al., 2020), such as the S/L subtypes (Lu et al., 2020a). New variants will likely continue to emerge as the virus mutates, and to uncover them requires deep, complete coverage of viral genomes from a large number of patients. With the existence of asymptomatic carriers (Bai et al., 2020) and possible recurrent infections in the same individual (An et al., 2020), longitudinal re-sampling of individuals with COVID-19 is also important to uncover intra-host viral heterogeneity, but as viral load decreases with time (He et al., 2020), the sensitivity of the sample processing method becomes critical. These studies require processing large volumes of clinical samples with a highly robust and scalable method that does not compromise sensitivity. We demonstrated that MINERVA libraries from clinical samples can generate deep and complete coverage of SARS-CoV-2 genomes that can be used for evolutionary tracing and variant characterization research. Furthermore, the high sensitivity, high coverage, and high depth of the SARS-CoV-2 viral genomes obtained by MINERVA can reveal unique viral SNV signatures in individuals with COVID-19, even when they are asymptomatic. We showed that these viral SNVs allow families of infected individuals to be co-clustered but are unique between families, which enables each individual to be accurately placed in the chain of transmission. Because MINERVA is easily scalable and implementable in a clinical lab setting, it can serve as a robust strategy for timely and critical epidemiological tracking and monitoring during a pandemic.

It is well established that SARS-CoV-2 can infect multiple organ systems, tissue compartments, and cell types (Chen et al., 2020; Wang et al., 2020; Wölfel et al., 2020; Young et al., 2020). In our profiling of COVID-19 clinical samples from multiple body sites of the same individual, we found that the viral load and viral subtypes vary across different body sites, possibly affected by interactions between microbial and other viral species as well as overall metagenomic diversity in different microenvironments of each body site. The effects of metagenomic diversity and inter-compartment heterogeneity on SARS-CoV-2 biology and COVID-19 symptom severity are also not understood. The biomass of different sample types varies, which requires library construction methods compatible with various sample content and different RNA quality. It is difficult to obtain high-quality, unbiased metagenomic data using conventional library construction methods from low-quantity samples as well as samples such as stool, in which bacteria dominate the metagenomes, because conventional methods are not sufficiently sensitive. Our previous study (Di et al., 2020) demonstrated a wide tolerance of sample input amount by our DNA/RNA hybrid tagmentation strategy. Our method showed the strongest data improvement in stool samples, likely because of better performance with low-input samples.

The versatility of MINERVA as a two-part protocol integrating MINERVA-m and MINERVA-e makes it possible to use one standard sample pipeline for highly sensitive metagenomic analysis and targeted deep sequencing of specific transcripts. To further improve the integration level of MINERVA, we evaluated the possibility of analyzing the merged datasets of MINERVA-m and MINERVA-e. It was revealed that the merged datasets (MINERVA-m+e) had highly concordant metagenomic compositions as the MINERVA-m datasets alone (Figures 7A and 7B), demonstrating the feasibility to perform metagenomic analysis and deep SARS-CoV-2 sequencing in a single sequencing run by MINERVA. The MINERVA-m+e and dsDL libraries were also compatible in terms of alpha diversities, species richness, and Shannon index (Figures S7A and S7B) and microbial compositions of different sample types (Figures 7C and S7C).

Using MINERVA, we demonstrated the first large-scale profiling of metagenomic composition of different body sites in the context of COVID-19. Several studies have investigated the relationship between gut microbes with SARS-CoV-2 infection and COVID-19 severity (Gou et al., 2020; Gu et al., 2020; Zuo et al., 2020); however, there is no discussion of metagenomic composition of other body sites. As we show here with MINERVA data from a wide range of sample types, there are large body site-specific differences, and our data suggest that microbial composition in pharyngeal swab samples also correlates significantly with disease severity. The metagenomic profile of these other body sites, which are arguably more directly involved in viral infection, have not been reported or investigated elsewhere with such a large sample size (Shen et al., 2020). Using MINERVA, we highlight several new directions of clinical and basic research, and with further investigation, these could shed light on the complex interactions between SARS-CoV-2 pathology, host microbial communities, host immunity, and disease progression. We also showed that MINERVA metagenomic profiles can identify highly abundant pathogenic species in samples from individuals with COVID-19 in a non-targeted fashion, which is challenging with conventional approaches. In our

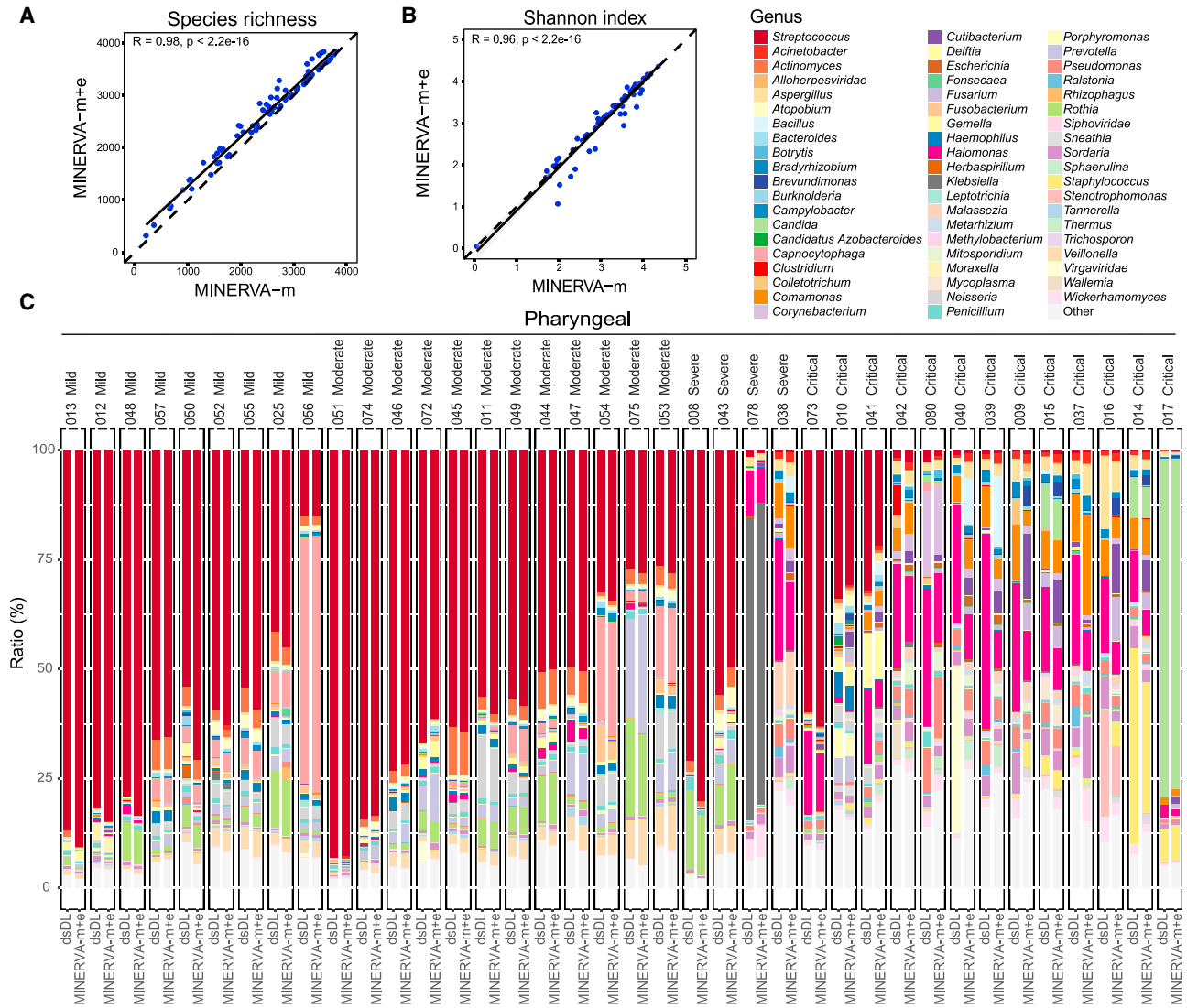


Figure 7. Evaluation of Microbial Profiles by Merged MINERVA Datasets

(A and B) Correlation of alpha diversity, including species richness (A) and Shannon index (B), between MINERVA-m and merged (m+e) datasets.

(C) Comparison of bacterial composition of pharyngeal samples between dsDL and merged MINERVA (m+e) datasets. Genera with a relative abundance over 1% are shown here.

samples, we found ~13.3% (6 of 45) of individuals with COVID-19 whose samples contained a high relative abundance of a pathogenic microbe other than SARS-CoV-2. Although this could indicate colonization or co-infection, it raises concerns regarding potentially missed co-infections. One secondary study found 8% of patients experiencing bacterial/fungal co-infection, but the rate of broad-spectrum antibiotics use for COVID-19 patients is much higher (72%) (Rawson et al., 2020). It is well known that co-infections in severe pneumonia can greatly affect patient outcome (Crotty et al., 2015; Shah et al., 2016), and it is estimated that 50% of patients with COVID-19 who died in this pandemic had secondary bacterial infections (Zhou et al., 2020a). Further primary studies using MINERVA could help to elucidate true co-infection rates to better guide strategies for antibiotics use.

Limitations

MINERVA was not created to be a rapid diagnostic assay; rather, we hope that its ease of use, versatility, scalability, sensitivity, and cost effectiveness will drive adoption of routine sequencing of COVID-19 clinical samples and facilitate multiple areas of much-needed SARS-CoV-2 and COVID-19 research for clinicians and researchers.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2020.11.030.

ACKNOWLEDGMENTS

We thank Chenyang Geng and the BIOPIC sequencing platform at Peking University for assistance with high-throughput sequencing experiments and Amelia Huang for assistance with figure preparation. This work was supported by the Ministry of Science and Technology of China (2018YFA0108100, 2018YFA0800200, and 2018YFC1002300), the National Natural Science Foundation of China (21675098, 21927802, and 21525521), the 2018 Beijing Brain Initiative (Z181100001518004), the Beijing Advanced Innovation Center for Structural Biology, the Beijing Advanced Innovation Center for Genomics, the Hong Kong Branch of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (SMSEGL20SC01), the Hong Kong Research Grants Council Theme-based Research Scheme (RGC TBRS T12-704/16R-2) and Collaborative Research Fund (RGC CRF C6002-17G), the Hong Kong RGC Early Career Support Scheme (RGC ECS 26101016), the Hong Kong Epigenomics Project (LKCCFL18SC01-E), and HKUST BDBI Labs.

AUTHOR CONTRIBUTIONS

C.C., Y.C., X.S.X., H.Z., Y.H., and J.W. conceived the project. J.L., P.D., Q.L., and C.S. conducted experiments. C.C., L.D., Q.J., J.L., Y.H., and J.W. analyzed the data. C.C., J.L., L.D., Q.J., A.R.W., Y.H., and J.W. wrote the manuscript with help from all other authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 26, 2020
Revised: October 11, 2020
Accepted: November 13, 2020
Published: November 20, 2020

REFERENCES

An, J., Liao, X., Xiao, T., Qian, S., Yuan, J., Ye, H., Qi, F., Shen, C., Liu, Y., Wang, L., et al. (2020). Clinical characteristics of the recovered COVID-19 patients with re-detectable positive RNA test. Ann. Transl. Med. *8*.

Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D.Y., Chen, L., and Wang, M. (2020). Presumed Asymptomatic Carrier Transmission of COVID-19. JAMA *323*, 1406–1407.

Bushnell, B. (2014). BBMap: A Fast (Accurate, Splice-Aware Aligner).

Carey, V.J. (2006). Ported to R by Thomas Lumley (versions 3.13, 4.4, version 4.13)., B. R. gee: Generalized Estimation Equation solver. https://CRAN.R-project.org/package=gee. R package version *4*, 13, 11.

Carl, G., and Kühn, I. (2007). Analyzing spatial autocorrelation in species distributions using Gaussian and logit models. Ecol. Modell. *207*, 159–170.

Chen, C., Gao, G., Xu, Y., Pu, L., Wang, Q., Wang, L., Wang, W., Song, Y., Chen, M., Wang, L., et al. (2020). SARS-CoV-2-Positive Sputum and Feces After Conversion of Pharyngeal Samples in Patients With COVID-19. Ann. Intern. Med. *172*, 832–834.

Cox, M.J., Loman, N., Bogaert, D., and O'Grady, J. (2020). Co-infections: potentially lethal and unexplored in COVID-19. Lancet Microbe *1*, e11.

Crotty, M.P., Meyers, S., Hampton, N., Bledsoe, S., Ritchie, D.J., Buller, R.S., Storch, G.A., Micek, S.T., and Kollef, M.H. (2015). Epidemiology, Co-Infections, and Outcomes of Viral Pneumonia in Adults: An Observational Cohort Study. Medicine (Baltimore) *94*, e2332.

de Lastours, V., Malosh, R.E., Aiello, A.E., and Foxman, B. (2015). Prevalence of Escherichia coli carriage in the oropharynx of ambulatory children and adults with and without upper respiratory symptoms. Ann. Am. Thorac. Soc. *12*, 461–463.

Deng, X., Gu, W., Federman, S., du Plessis, L., Pybus, O.G., Faria, N., Wang, C., Yu, G., Pan, C.Y., Guevara, H., et al. (2020). A Genomic Survey of SARS-CoV-2 Reveals Multiple Introductions into Northern California without a Predominant Lineage. medRxiv. https://doi.org/10.1101/2020.03.27.20044925.

Di, L., Fu, Y., Sun, Y., Li, J., Liu, L., Yao, J., Wang, G., Wu, Y., Lao, K., Lee, R.W., et al. (2020). RNA sequencing by direct tagmentation of RNA/DNA hybrids. Proc. Natl. Acad. Sci. USA *117*, 2886–2893.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Fauci, A.S., Lane, H.C., and Redfield, R.R. (2020). Covid-19 - Navigating the Uncharted. N. Engl. J. Med. *382*, 1268–1269.

Gou, W., Fu, Y., Yue, L., Chen, G.-d., Cai, X., Shuai, M., Xu, F., Yi, X., Chen, H., Zhu, Y.J., et al. (2020). Gut microbiota may underlie the predisposition of healthy individuals to COVID-19. medRxiv. https://doi.org/10.1101/2020.04.22.20076091.

Gu, S., Chen, Y., Wu, Z., Chen, Y., Gao, H., Lv, L., Guo, F., Zhang, X., Luo, R., Huang, C., et al. (2020). Alterations of the Gut Microbiota in Patients with COVID-19 or H1N1 Influenza. Clin. Infect. Dis. ciaa709.

Gudbjartsson, D.F., Helgason, A., Jonsson, H., Magnusson, O.T., Melsted, P., Norddahl, G.L., Saemundsdottir, J., Sigurdsson, A., Sulem, P., Agustsdottir, A.B., et al. (2020). Spread of SARS-CoV-2 in the Icelandic Population. N. Engl. J. Med. *382*, 2302–2315.

He, X., Lau, E.H.Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y.C., Wong, J.Y., Guan, Y., Tan, X., et al. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat. Med. *26*, 672–675.

Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

Morgan, X.C., Tickle, T.L., Sokol, H., et al. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol *13*, https://doi.org/10.1186/gb-2012-13-9-r79.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., et al. (2019). vegan: Community Ecology Package. R package version *2*, 5–6. https://cran.r-project.org/web/packages/vegan/index.html.

Kim, D., Song, L., Breitwieser, F.P., and Salzberg, S.L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. *26*, 1721–1729.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, D., Liu, C.M., Luo, R., Sadakane, K., and Lam, T.W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics *31*, 1674–1676.

Lu, J., Cui, J., Qian, Z., Wang, Y., Zhang, H., Duan, Y., Wu, X., Yao, X., Song, Y., Li, X., et al. (2020a). On the origin and continuing evolution of SARS-CoV-2. Natl. Sci. Rev. *7*, 1012–1023.

Lu, J., du Plessis, L., Liu, Z., Hill, V., Kang, M., Lin, H., Sun, J., Francois, S., Kraemer, M.U.G., Faria, N.R., et al. (2020b). Genomic Epidemiology of SARS-CoV-2 in Guangdong Province, China. Cell *181*, 997–1003.e9.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., et al. (2020c). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet *395*, 565–574.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rawson, T.M., Moore, L.S.P., Zhu, N., Ranganathan, N., Skolimowska, K., Gilchrist, M., Satta, G., Cooke, G., and Holmes, A. (2020). Bacterial and fungal co-infection in individuals with coronavirus: A rapid review to support COVID-19 antimicrobial prescribing. Clin. Infect. Dis. ciaa530.

Ren, L.L., Wang, Y.M., Wu, Z.Q., Xiang, Z.C., Guo, L., Xu, T., Jiang, Y.Z., Xiong, Y., Li, Y.J., Li, X.W., et al. (2020). Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. Chin. Med. J. (Engl.) *133*, 1015–1024.

Shah, N.S., Greenberg, J.A., McNulty, M.C., Gregg, K.S., Riddell, J., 4th, Mangino, J.E., Weber, D.M., Hebert, C.L., Marzec, N.S., Barron, M.A., et al. (2016). Bacterial and viral co-infections complicating severe influenza: Incidence and impact among 507 U.S. patients, 2013-14. J. Clin. Virol. *80*, 12–19.

Shen, Z., Xiao, Y., Kang, L., Ma, W., Shi, L., Zhang, L., Zhou, Z., Yang, J., Zhong, J., Yang, D., et al. (2020). Genomic Diversity of Severe Acute Respiratory Syndrome-Coronavirus 2 in Patients With Coronavirus Disease 2019. Clin. Infect. Dis. *71*, 713–720.

Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., and Tan, W. (2020). Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA *323*, 1843–1844.

WHO (2020). WHO Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int/.

Wölfel, R., Corman, V.M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M.A., Niemeyer, D., Jones, T.C., Vollmar, P., Rothe, C., et al. (2020). Virological assessment of hospitalized patients with COVID-2019. Nature *581*, 465–469.

Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., et al. (2020). A new coronavirus associated with human respiratory disease in China. Nature *579*, 265–269.

Xiao, M., Liu, X., Ji, J., Li, M., Li, J., Yang, L., Sun, W., Ren, P., Yang, G., Zhao, J., et al. (2020). Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. Genome Med. *12*, 57.

Young, B.E., Ong, S.W.X., Kalimuddin, S., Low, J.G., Tan, S.Y., Loh, J., Ng, O.T., Marimuthu, K., Ang, L.W., Mak, T.M., et al.; Singapore 2019 Novel Coronavirus Outbreak Research Team (2020). Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore. JAMA *323*, 1488–1494.

Zaura, E., Keijser, B.J., Huse, S.M., and Crielaard, W. (2009). Defining the healthy "core microbiome" of oral microbial communities. BMC Microbiol. *9*, 259.

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., et al. (2020a). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet *395*, 1054–1062.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., et al. (2020b). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature *579*, 270–273.

Zuo, T., Zhang, F., Lui, G.C.Y., Yeoh, Y.K., Li, A.Y.L., Zhan, H., Wan, Y., Chung, A.C.K., Cheung, C.P., Chen, N., et al. (2020). Alterations in Gut Microbiota of Patients With COVID-19 During Time of Hospitalization. Gastroenterology *159*, 944–955.e8.

"Picard Toolkit." 2019. Broad Institute, GitHub Repository. http://broadinstitute.github.io/picard/; Broad Institute.

2020. 2.5-7. https://cran.r-project.org/web/packages/vegan/index.html.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
clinical samples	Beijing Ditan Hospital	N/A
Chemicals, Peptides, and Recombinant Proteins		
DNase I (RNase-free)	NEB	Cat#M0303
Superscript II reverse transcriptase	Invitrogen	Cat#18064014
N,N-Dimethylformamide	Sigma	Cat#D4551
Recombinant RNase Inhibitor	Takara	Cat#2313
Deoxynucleotide (dNTP) Solution Set	NEB	Cat#N0446S
Betaine solution	Sigma	Cat#B0300
PEG8000	VWR Life Science	Cat#97061
ATP	NEB	Cat#P0756
Critical Commercial Assays		
RNeasy Mini Kit	QIAGEN	Cat#74104
RNA Clean & Concentrator-5 kit	Zymo Research	Cat#R1015
QIAamp Viral RNA Mini Kit	QIAGEN	Cat#52906
MGIEasy rRNA removal kit	BGI	Cat#1000005953
TruePrep DNA Library Prep Kit V2 for Illumina	Vazyme	Cat#TD501
TargetSeq One Cov Kit	iGeneTech	Cat#502002-V1
Q5 High-Fidelity 2x Master Mix	NEB	Cat#M0492
VAHTS DNA Clean Beads	Vazyme	Cat#N411
ChamQ SYBR qPCR master mix	Vazyme	Cat#Q311-02
Deposited Data		
The sequencing data generated during this study	This paper	Genome Sequencing Archive: PRJCA002533
Experimental Models: Cell Lines		
NIH/3T3	ATCC	CRL-1658
Oligonucleotides		
Decamer(N10): NNNNNNNNNN	Sangon	N/A
T30VN: TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN	Sangon	N/A
SARS-CoV-2 qPCR N gene forward primer: GGGGAACTTCTCCTGCTAGAAT	sangon	N/A
SARS-CoV-2 qPCR N gene reverse primer: CAGACATTTTGCTCTCAAGCTG	sangon	N/A
xGen Universal Blockers	IDT	Cat#1079586
Software and Algorithms		
BBmap	Bushnell, 2014	https://www.osti.gov/biblio/1241166
STAR	Dobin et al., 2013	https://github.com/alexdobin/STAR
samtools	Li et al., 2009	http://samtools.sourceforge.net/
Centrifuge	Kim et al., 2016	https://codeload.github.com/infphilo/centrifuge/zip/centrifuge-genome-research
MEGAHIT	Li et al., 2015	https://github.com/voutcn/megahit
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml
Bedtools	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/#

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
vegan	Oksanen et al., 2019	https://cran.r-project.org/web/packages/vegan/index.html
MaAsLin	(Morgan et al., 2012)	https://github.com/pooranis/maaslin/
Picard Tools	(Broad Institute, 2019)	http://broadinstitute.github.io/picard/
Other		
bench protocol	This paper	Methods S1

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jianbin Wang (jianbinwang@tsinghua.edu.cn)

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The sequencing data generated during this study have been uploaded to Genome Sequencing Archive (PRJCA002533). Detailed bench protocol is available from Mendeley Data at https://doi.org/10.17632/45w7hv53yr.1.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

The NIH/3T3 cell line was purchased from ATCC. The complete growth medium was made using DMEM (cat. No.11965-092; Life Technologies), 10% fetal bovine serum (cat. No. 16000-044; Life Technologies), and 1% penicillin and streptomycin. The cell line was incubated with 5% carbon dioxide at 37°C in a culture flask.

Ethics Approval

This study was approved by the Ethics Committee of Beijing Ditan Hospital, Capital Medical University (No. KT2020-006-01).

Patients and Clinical Samples

From January 23, 2020 to April 20, 2020, 91 patients were enrolled in this study according to the 7th guideline for the diagnosis and treatment of COVID-19 from the National Health Commission of the People's Republic of China. All patients, diagnosed with COVID-19, were hospitalized in Beijing Ditan Hospital and classified into four severity degrees, mild, moderate, severe, and critical illness, according to the 7th guideline for the diagnosis and treatment of COVID-19 from the National Health Commission of the People's Republic of China (https://www.chinadaily.com.cn/pdf/2020/1.Clinical.Protocols.for.the.Diagnosis.and.Treatment.of.COVID-19.V7.pdf). Briefly, mild cases are those with mild clinical symptoms, and there was no sign of pneumonia on imaging. Moderate cases are those showing fever and respiratory symptoms with radiological findings of pneumonia. Severe cases include the adult cases meeting any of the following criteria: (1) respiratory distress ($\geq$30 breaths/min); (2) oxygen saturation $\leq$ 93% at rest; (3) arterial partial pressure of oxygen (PaO2)/fraction of inspired oxygen (FiO2) $\leq$ 300 mmHg. We collected 136 samples (60 pharyngeal swab samples, 51 sputum samples, and 25 stool samples) from these patients. Pharyngeal swab samples were collected into the viral sampling medium (Yocon, MT0301-1), sputum samples were collected directly into sterile containers, and stool samples were collected in the storage buffer using the stool sampling kit (Longsee, LS-R-P-003). All samples were stored at −80 °C.

METHOD DETAILS

Optimization of MINERVA Protocol

We used the total RNA extracted from 3T3 cells to optimize experimental protocols. RNA extraction was performed using RNeasy Mini Kit (QIAGEN, Cat.No.74104). DNA was then removed through DNase I (NEB, Cat.No.M0303) digestion. The resulting total RNA was concentrated by RNA Clean & Concentrator-5 kit (Zymo Research, Cat R1015), and its quality was assessed by the Fragment Analyzer Automated CE System (AATI). Its quantification was done by Qubit 2.0 (Invitrogen). To optimize the MINERVA protocol, different amount of random decamer (N10) (0, 1, or 10 μM) was used to set up reverse transcription reactions. Titration of Tn5 transposome (0.2, 0.5, or 1.0 μl Vazyme V50; 0.05 or 0.25 μl home-made pTXB1) was performed in tagmentation procedure. In all tests, 10 ng 3T3 total RNA was used, and all reagents except for N10 or Tn5 transposome remain unchanged. All libraries were sequenced

on Illumina NextSeq 500 with 2x75 paired-end mode. Clean data was aligned to GRCm38 genome and known transcript annotation using Tophat2 v2.1.1. Ribosome-removed aligned reads were proceeded to calculate FPKM by Cufflinks v2.2.1 and gene body coverage by RSeQC v.2.6.4.

RNA Extraction and rRNA Removal

For all the clinical samples, nucleic acids extraction was performed in a BSL-3 laboratory. Samples were deactivated by heating at 56°C for 30 min before extraction. Total RNA was extracted using QIAamp Viral RNA Mini Kit (QIAGEN) following the manufacturer's instructions. In most (130 out of 136) samples we specifically omitted the use of carrier RNA due to its interference on the most prevalent sample preparation protocols for high-throughput sequencing. After nucleic acids extraction, rRNA was removed by rDNA probe hybridization and RNase H digestion, followed by DNA removal through DNase I digestion, using MGIEasy rRNA removal kit (BGI, Shenzhen, China). The final elution volume was 12-20 μl for each sample. For carrier RNA removal tests, 1.7 μg polyA carrier RNA was spiked into 18 μl of elute from QIAamp Viral RNA Mini Kit. To remove the carrier RNA from these spike-in samples and other samples extracted with carrier RNA, 2 μg poly(T) 59-mer (T59) oligo was added during the rDNA hybridization step.

dsDL Metagenomic RNA Library Construction and Sequencing

The libraries were constructed using MGIEasy reagents (BGI, China) following manufacture's instruction. The purified RNA, after rRNA depletion and DNA digestion, underwent reverse transcription, second strand synthesis, and sequencing adaptor ligation. After PCR amplification, DNA was denatured and circularized before being sequenced on DNBSEQ-T7 sequencers (BGI, China).

MINERVA Library Preparation

A step-by-step protocol is available (Methods S1). Briefly, 2.7 μl RNA from rRNA and DNA removal reaction was used for standard SHERRY reverse transcription, with the following modifications: 1) 10 pmol random decamer (N10) was added to improve coverage; 2) initial concentrations of dNTPs and oligo-dT (T30VN) were increased to 25 mM and 100 μM, respectively. For 5.4 μl and 10.8 μl input, the entire reaction was simply scaled up 2 and 4 folds, respectively. The RNA/DNA hybrid was tagmented in TD reaction buffer (10 mM Tris-Cl pH 7.6, 5 mM $MgCl_2$, 10% DMF) supplemented with 3.4% PEG8000 (VWR Life Science, Cat.No.97061), 1 mM ATP (NEB, Cat.No. P0756), and 1U/μl RNase inhibitor (TaKaRa, Cat.No. 2313B). The reaction was incubated at 55°C for 30 min. 20 μl tagmentation product was mixed with 20.4 μl Q5 High-Fidelity 2X Master Mix (NEB, Cat.No. M0492L), 0.4 μl SuperScript II reverse transcriptase, and incubated at 42°C for 15 min to fill the gaps, followed by 70°C for 15 min to inactivate SuperScript II reverse transcriptase. Then index PCR was performed by adding 4 μl 10 μM unique dual index primers and 4 μl Q5 High-Fidelity 2X Master Mix, with the following thermo profile: 98°C 30 s, 18 cycles of [98°C 20 s, 60°C 20 s, 72°C 2 min], 72°C 5 min. The PCR product was then purified with 0.8x VAHTS DNA Clean Beads (Vazyme, Cat. No. N411). These libraries were sequenced on Illumina NextSeq 500 with 2x75 paired-end mode for metagenomic analysis.

For preparing MINERVA-e libraries through SARS-CoV-2 enrichment, 1 μL metagenomic library was first quantified for N gene using quantitative PCR (F: GGGGAACTTCTCCTGCTAGAAT, R: CAGACATTTTGCTCTCAAGCTG) after 1:200 dilution. Then 8~16 libraries were pooled together based on qPCR results to obtain relatively uniform amount of data. Specifically, Ct values were divided into 4 groups and each group corresponded to a pooling volume: Ct above 28 (50ul), Ct of 24-28 (20ul), Ct of 20-24 (8ul), Ct below 20 (3ul). Pooled library was further processed with TargetSeq One Cov Kit (iGeneTech, Cat.No.502002-V1) following manufacturer's instruction. The iGeneTech Blocker was replaced by the IDT xGen Universal Blockers (NXT). These MINERVA libraries were sequenced on Illumina NextSeq 500 with 2x75 paired-end mode for deep SARS-CoV-2 analysis.

Data Processing

For metagenomic RNA-seq data, raw reads were quality controlled using BBmap (version 38.68) (Bushnell, 2014) and then mapped to the human genome reference (GRCh38) using STAR (version 2.6.1d) (Dobin et al., 2013) with default parameters. All unmapped reads were collected using samtools (version 1.3) (Li et al., 2009) for microbial taxonomy assignment by Centrifuge (version 1.0.4) (Kim et al., 2016). Custom reference was built from all complete bacterial, viral and any assembled fungal genomes downloaded from NCBI RefSeq database (viral and fungal genomes were downloaded on February 4th, 2020, and bacterial genomes were downloaded on November 14th, 2018). There were 11,174 bacterial, 8,997 viral, and 308 fungal genomes respectively. Bacterial Shannon diversity (entropy) was calculated at species level, and the species abundance was measured based on total reads assigned at the specific clade normalized by genome size and sequencing depth. Bacterial genus composition was analyzed based on reads proportion directly assigned by Centrifuge. For dsDL sequencing data, sub-sampling was performed for each sample to obtain ~12M pair-end nonhuman reads, which is the median of MINERVA datasets. The same workflow as above was performed for the removal of human reads and microbial taxonomy assignment.

For metagenome assembly, nonhuman reads from samples of the same sample type were merged first for co-assembly using MEGAHIT (version 1.2.9) (Li et al., 2015) with default parameters. In total there were approximately 200M, 150M and 50M read pairs for pharyngeal (n = 68), sputum (n = 59) and stool (n = 33) samples respectively. Contigs longer than 200nt were kept, resulting in 273,434; 266,932 and 58,836 contigs for each sample type, and the N50 values were 800bp, 689bp and 699bp respectively. Centrifuge was used to assign taxonomies to contigs using the database mentioned above. After taxonomy assignment at contigs level, we mapped reads of each sample back to the co-assembled contigs to assess the microbial composition of each individual sample.

Reads mapping to the contigs was done using Bowtie2 (version 2.3.5.1) (Langmead and Salzberg, 2012) with ''–very-sensitive'' mode. Pileup from BBmap (version 38.68) package were used to calculate the coverage of mapped reads for each individual samples. The overall mapping ratio were 56.3%, 57.4% and 46.2% for each sample type. For species identified with high abundance by metagenomics analysis, we mapped reads back to their genome reference to check the genome coverage of each sample by using Bowtie2 (version 2.3.5.1) with ''–very-sensitive'' mode. Bedtools (version 2.29.0) (Quinlan and Hall, 2010) was used to calculate the genome coverage, and the genome coverage was then normalized by dividing the sequencing depth of corresponding samples and then multiply a scaling factor of 1,000,000 for comparisons. We calculated the relative abundance of taxa as the ratio of reads assigned to that taxa for the metagenomic strategy, then compared it with the ratio of reads mapped to contigs which were assigned as that taxa for the metagenome assembly strategy.

For SARS-CoV-2 genome analysis, raw reads were trimmed to remove sequencing adaptors and low-quality bases with Cutadapt v1.15. BWA 0.7.15-r1140 was used to align reads to the SARS-CoV-2 reference genome (NC_045512.2). Then we depleted duplicates from the properly-paired alignment with Picard Tools v2.17.6 (Broad Institute, 2019). The resulting bam was further filtered for mutation calling by two criterions: 1) The mapping position was before 29856 in SARS-CoV-2 genome; 2) The soft-clip bases of each read accounted for less than 50%. We used mpileup function in samtools v1.10 to call SNP and InDel with parameter -C 50 -Q 20 -q 15 -E -d 0. We called mutation if the depth $\geq$ 10 and strand bias $\geq$ 0.2 or supported by independently distributed reads, which were checked in IGV (Integrative Genome Browser). The strand bias is defined as the value that minimum of positive strand depth and negative strand depth divided by the maximum. For cluster heatmap analysis, we chose samples of high coverage and two asymptomatic family, and included all mutations which variant allele frequency $\geq$ 0.2 and strand bias $\geq$ 0.2 occurred in at least one sample. The frequency of mutated allele was simply calculated as observed mutant reads divided by total reads for each site.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed in R, version 3.5.1. Permutational multivariate analysis of variance (PERMANOVA) was applied to assess the effect of meta factors, including age, gender and disease status on the Bray-Curtis distance among samples by using adonis2 function in R package vegan (Oksanen et al., 2020). To account for potential effect of repeated-measures of multiple samples from the same subject, permutation was constrained within the same time point during PERMANOVA analysis. PCoA ordination analysis was performed based on Bray-Curtis distance to explore the difference among samples. The Bray-Curtis distance was calculated by vegdist function and PCoA axes were calculated using cmdscale functions from R package vegan. The p values were calculated by PERMANOVA analysis. Generalized Estimating Equations (GEE) was applied to assess associations of microbes with disease status by using gee function in R (Carey, 2006; Carl and Kühn, 2007). Only microbes with relative abundance > 0.01 and prevalence > 0.01 were used in this model. P values were calculated from the estimated robust z-score based on standard Gaussian distribution. The Benjamini-Hochberg procedure was applied for the correction of the p values. Results were filtered based on both significance (BH-adjusted $p < 0.05$) and effect size (absolute coefficient > 0.1). Kruskall-Wallis test and Wilcoxon rank-sum test were used for other multi-group and two-group comparisons respectively if not specifically stated.

ADDITIONAL RESOURCES

Detailed Protocol

A detailed bench protocol is available as Methods S1.

Molecular Cell, Volume *80*

Supplemental information

MINERVA: A Facile Strategy for SARS-CoV-2 Whole-Genome Deep Sequencing of Clinical Samples

Chen Chen, Jizhou Li, Lin Di, Qiuyu Jing, Pengcheng Du, Chuan Song, Jiarui Li, Qiong Li, Yunlong Cao, X. Sunney Xie, Angela R. Wu, Hui Zeng, Yanyi Huang, and Jianbin Wang

Molecular Cell, Volume *80*

Supplemental Information

MINERVA: A Facile Strategy for SARS-CoV-2 Whole-Genome Deep Sequencing of Clinical Samples

Chen Chen, Jizhou Li, Lin Di, Qiuyu Jing, Pengcheng Du, Chuan Song, Jiarui Li, Qiong Li, Yunlong Cao, X. Sunney Xie, Angela R. Wu, Hui Zeng, Yanyi Huang, and Jianbin Wang

Supporting Information

MINERVA: A facile strategy for SARS-CoV-2 whole genome deep sequencing of clinical samples

Chen Chen,[1,11] Jizhou Li,[2,11] Lin Di,[3,4,11] Qiuyu Jing,[5,11] Pengcheng Du,[1,11] Chuan Song,[1] Jiarui Li,[1] Qiong Li,[2] Yunlong Cao,[3] X. Sunney Xie,[3] Angela R. Wu,[5,6,*] Hui Zeng,[1,*] Yanyi Huang,[3,7,8,9,*] and Jianbin Wang[2,9,10,12,*]

Lead Contact: Jianbin Wang (jianbinwang@tsinghua.edu.cn)

Supporting Materials:

Figures S1-S7

Table S1-S4

Figure S1, related to Figure 1. **Comparison between MINERVA and the conventional dsDL strategy**. Cost of highly multiplexed PCR approach is also shown as reference. MINERVA has advantages of less hands-on time and lower cost compared to dsDL, as MINERVA is a simpler procedure and requires much lower sequencing depth (up to 8Gbp for each MINERVA-m or -e sample while ~100Gbp for each dsDL sample). Though highly multiplexed PCR costs even lower, it is generally accepted that it cannot give metagenomics data and is prone to generate bias for samples with SNV. (B-E) Effect of N10 primer during reverse transcription and Tn5 amount on ribosomal rate, insert size, and gene body coverage evenness.

Figure S2, related to Figure 3. **Quality assessment of MINERVA-m data.** (A) Amount of sequencing data for different libraries. (B and C) Comparison between MINERVA-m and dsDL libraries on bacterial species richness (B) and shannon index (C). (D-F) Comparison of microbial composition of pharyngeal (D), sputum (E) and stool (F) samples profiled by metagenomic and assembly strategies.

Figure S3, related to Figure 3. **Metagenomics analysis of COVID-19 samples using MINERVA.** (A) Microbial composition of non-template controls (NTCs) based on bacterial genus, fungal genus and viral family. (B) Kruskal-Wallis test was used for the comparison of species richness between NTCs and patient samples. (C) PERMANOVA analysis to assess factors associated with microbial compositions in all samples (PERMANOVA test, *p<0.05, **p<0.01, ***p<0.001). (D) PCoA analysis based on Bray-Curtis distance of sputum samples, including Healthy controls (n=8), mild (n=8), moderate (n=33), severe (n=9), critical (n=1) patient samples and non-template controls (NTC, n=2). Patient samples were significantly different from healthy controls ($p < 0.001$ by PERMANOVA test). (E) Wilcoxon rank-sum test was used for the comparison of alpha diversity between Healthy controls (n=8) and Patient sputum samples (n=51). Decreased Shannon index was observed in patient samples. (F) Multivariate analysis (by MaAsLin linear model) to determine taxa associated with disease using sputum samples. Patients were taken as random effect in this model. Relative abundance was transformed using arcsin square-root transformation first before running this model. Results were filtered based on both significance (FDR-adjusted $p < 0.05$) and effect size (absolute coefficient >= 0.1). *Streptococcus* and

Veillonella were found most significantly associated with disease (top panel) and they are enriched in patient and healthy groups respectively (bottom panel). (G) PCoA analysis based on Bray-Curtis distance of stool samples, including Healthy controls (n=8), mild (n=15), moderate (n=10 patient samples and non-template controls (NTC, n=2). There is no difference between patient samples and healthy controls (p=0.08). (H) Comparison of alpha diversity between Healthy controls (n=8) and Patient stools samples (n=25). Wilcoxon rank-sum test was applied. No difference of alpha diversity between healthy controls and patient samples was observed, in terms of both species richness (left panel) and Shannon index (right panel). (I) Comparison of abundance of taxa (shown as read count, in log10-scale) identified in (F) between patient samples and NTC controls. Kruskal-Wallis test was applied here.

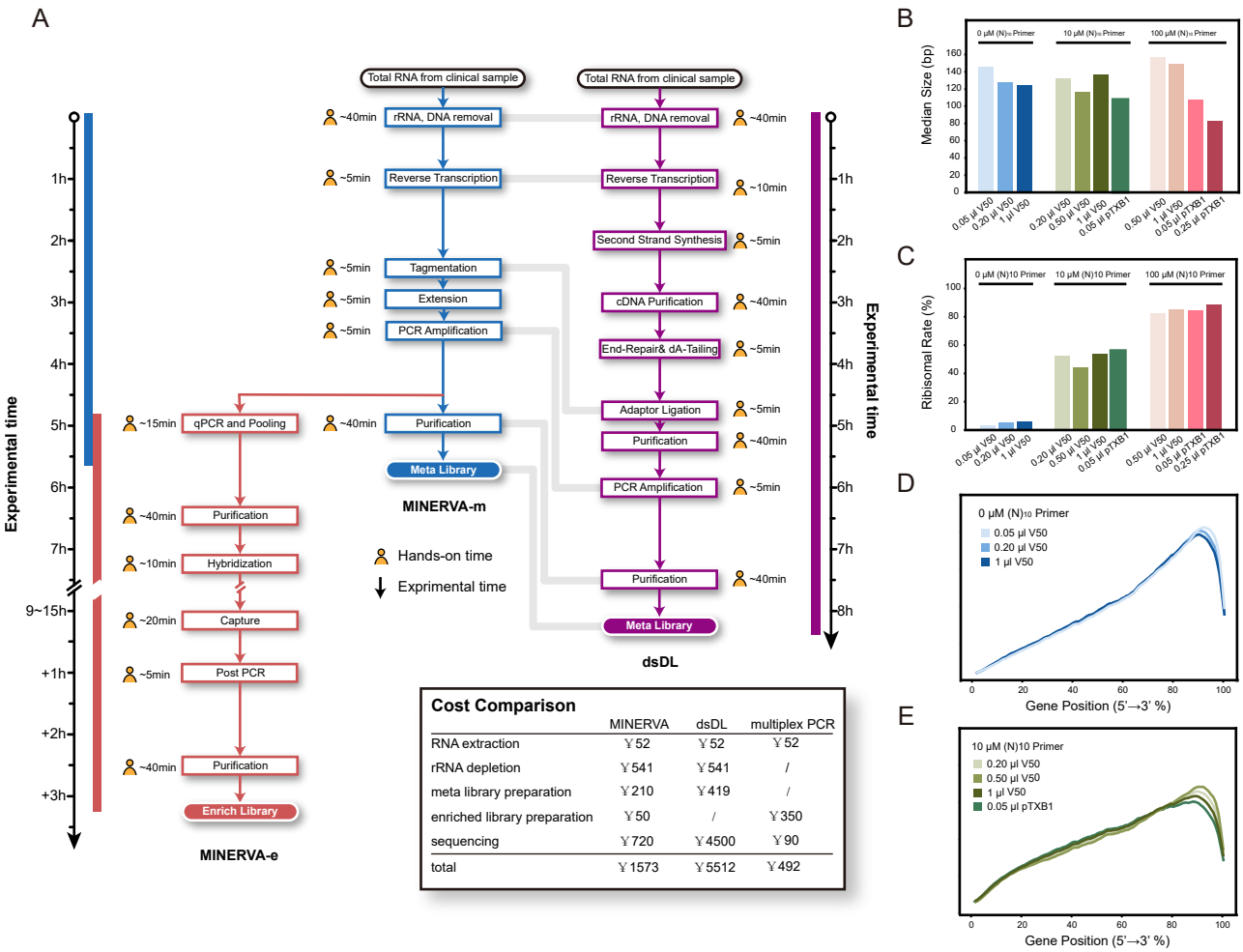
Figure S4, related to Figure 4. **Abundance of identified potential pathogens profiled by metagenomics assembly workflow**. The abundance was calculated as the coverage of contigs assigned to certain species, and then normalized by sequencing depth.

Figure S5, related to Figure 5. **SARS-CoV-2 genome sequencing results of MINERVA-m and MINERVA-e libraries**. (A) SARS-CoV-2 mapping ratio statistics of MINERVA-e libraries. (B and C) SARS-CoV-2 genome coverage and depth statistics of MINERVA-m libraries. (D and E) SARS-CoV-2 genome depth statistics of MINERVA-e libraries of 3T3 (D) and healthy donor (E) samples.

Figure S6, related to Figure 6. **Longitudinal SARS-CoV-2 mutation analysis of individual patients**.

Figure S7, related Figure 7. **Evaluation of microbial profiles by merged MINERVA (MINERVA-m+e) datasets**. (A and B) Correlation of alpha diversity, including species richness (A) and Shannon index (B), between dsDL and merged (m+e) datasets. (C) Comparison of bacterial composition of stool and sputum samples between dsDL and MINERVA-m+e datasets. Genus with relative abundance over 1% are shown here.

Figure S1



Cost Comparison

	MINERVA	dsDL	multiplex PCR
RNA extraction	¥52	¥52	¥52
rRNA depletion	¥541	¥541	/
meta library preparation	¥210	¥419	/
enriched library preparation	¥50	/	¥350
sequencing	¥720	¥4500	¥90
total	¥1573	¥5512	¥492

Figure S2

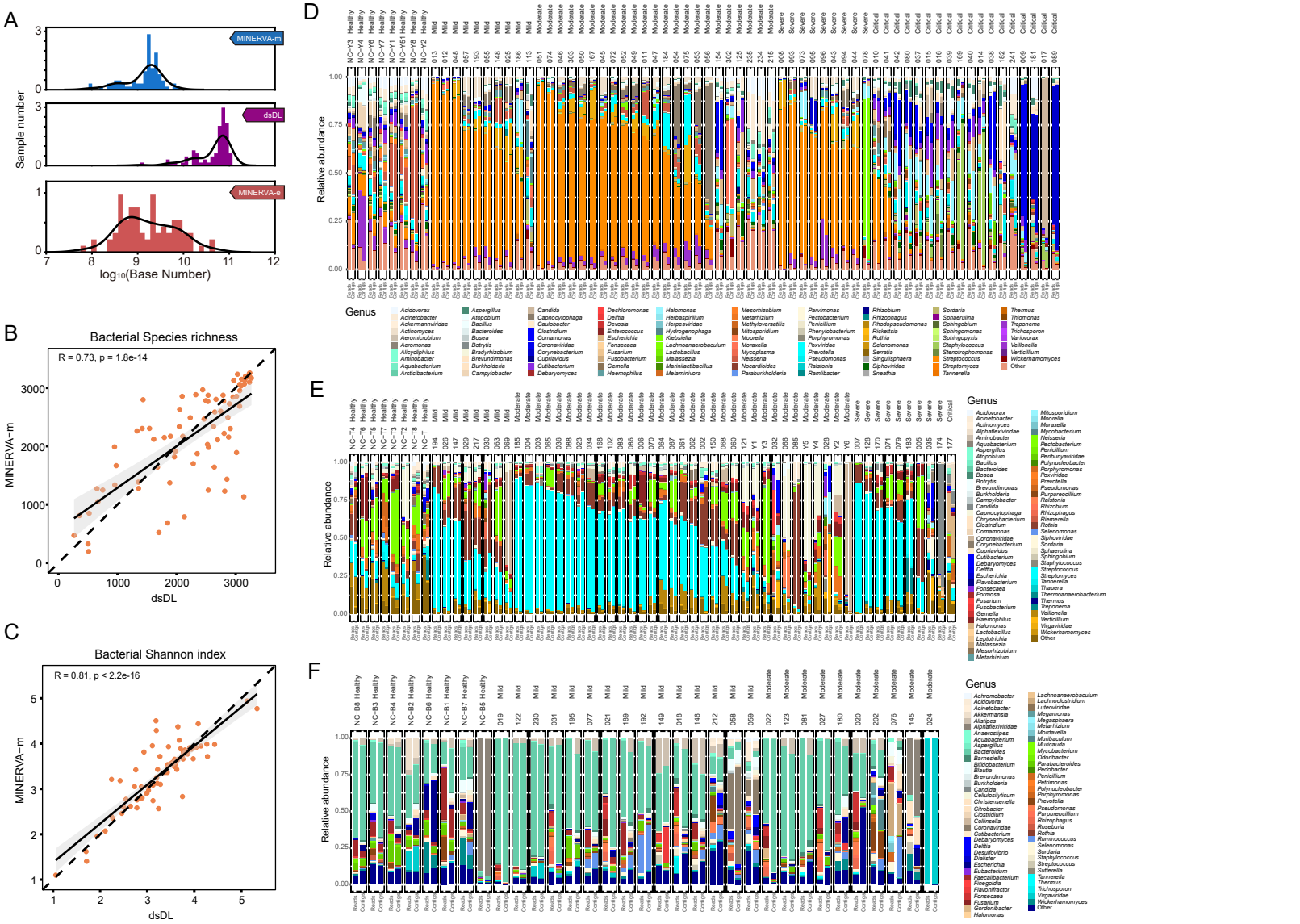

A
MINERVA-m
dsDL
MINERVA-e
Sample number
log10(Base Number)
B
Bacterial Species richness
R = 0.73, p = 1.8e-14
MINERVA−m
dsDL
C
Bacterial Shannon index
R = 0.81, p < 2.2e-16
MINERVA−m
dsDL
D
Relative abundance
Genus
E
Relative abundance
Genus
F
Relative abundance
Genus

Figure S3

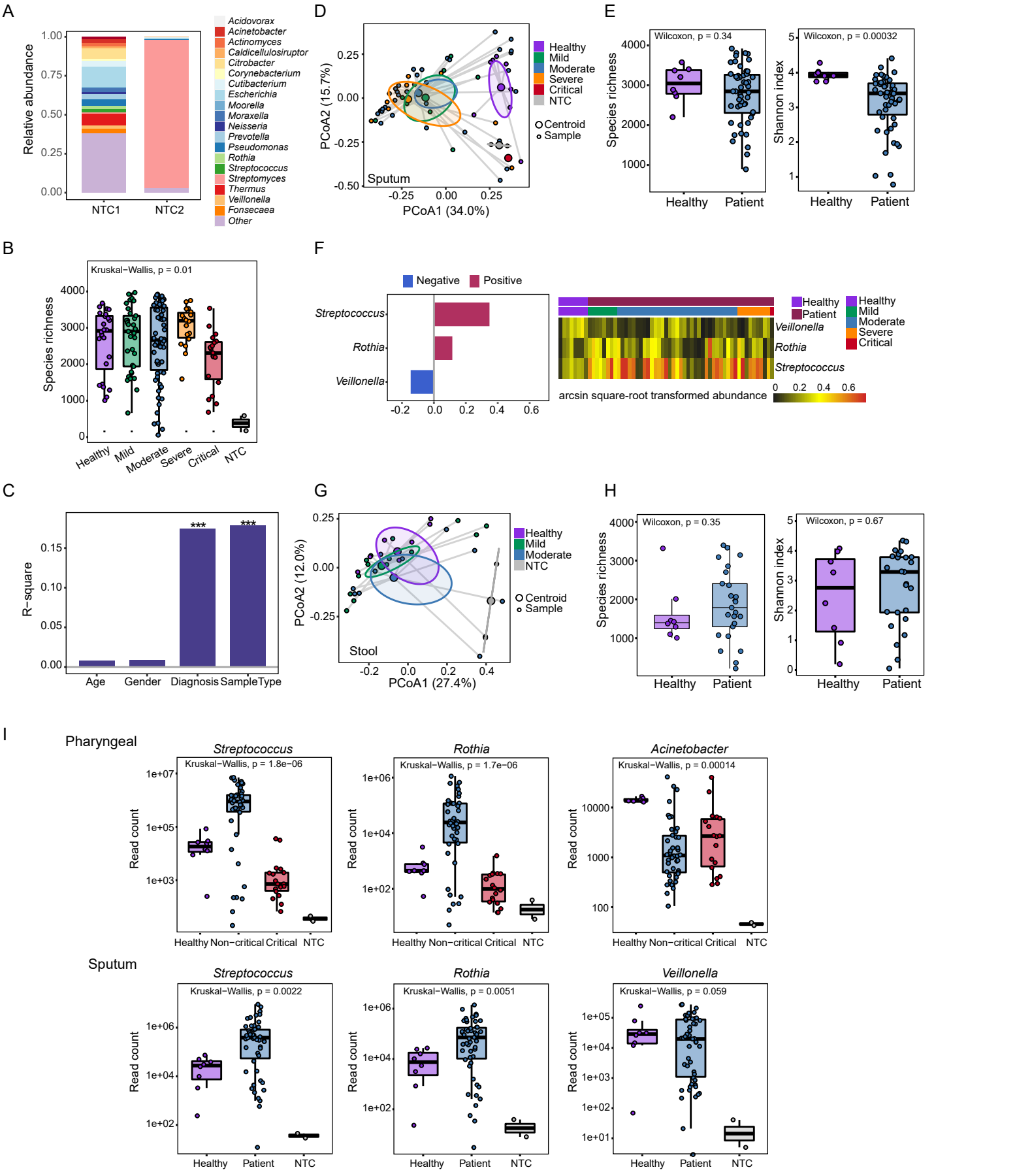

A
Relative abundance
NTC1
NTC2
Acidovorax
Acinetobacter
Actinomyces
Caldicellulosiruptor
Citrobacter
Corynebacterium
Cutibacterium
Escherichia
Moorella
Moraxella
Neisseria
Prevotella
Pseudomonas
Rothia
Streptococcus
Streptomyces
Thermus
Veillonella
Fonsecaea
Other
B
Kruskal−Wallis, p = 0.01
Species richness
Healthy
Mild
Moderate
Severe
Critical
NTC
C
R−square
Age
Gender
Diagnosis
SampleType
D
PCoA2 (15.7%)
PCoA1 (34.0%)
Sputum
Healthy
Mild
Moderate
Severe
Critical
NTC
Centroid
Sample
E
Wilcoxon, p = 0.34
Species richness
Wilcoxon, p = 0.00032
Shannon index
Healthy
Patient
F
Negative
Positive
Streptococcus
Rothia
Veillonella
arcsin square-root transformed abundance
G
PCoA2 (12.0%)
PCoA1 (27.4%)
Stool
H
Wilcoxon, p = 0.35
Species richness
Wilcoxon, p = 0.67
Shannon index
I
Pharyngeal
Streptococcus
Kruskal−Wallis, p = 1.8e−06
Rothia
Kruskal−Wallis, p = 1.7e−06
Acinetobacter
Kruskal−Wallis, p = 0.00014
Read count
Healthy
Non−critical
Critical
NTC
Sputum
Streptococcus
Kruskal−Wallis, p = 0.0022
Rothia
Kruskal−Wallis, p = 0.0051
Veillonella
Kruskal−Wallis, p = 0.059

Figure S4

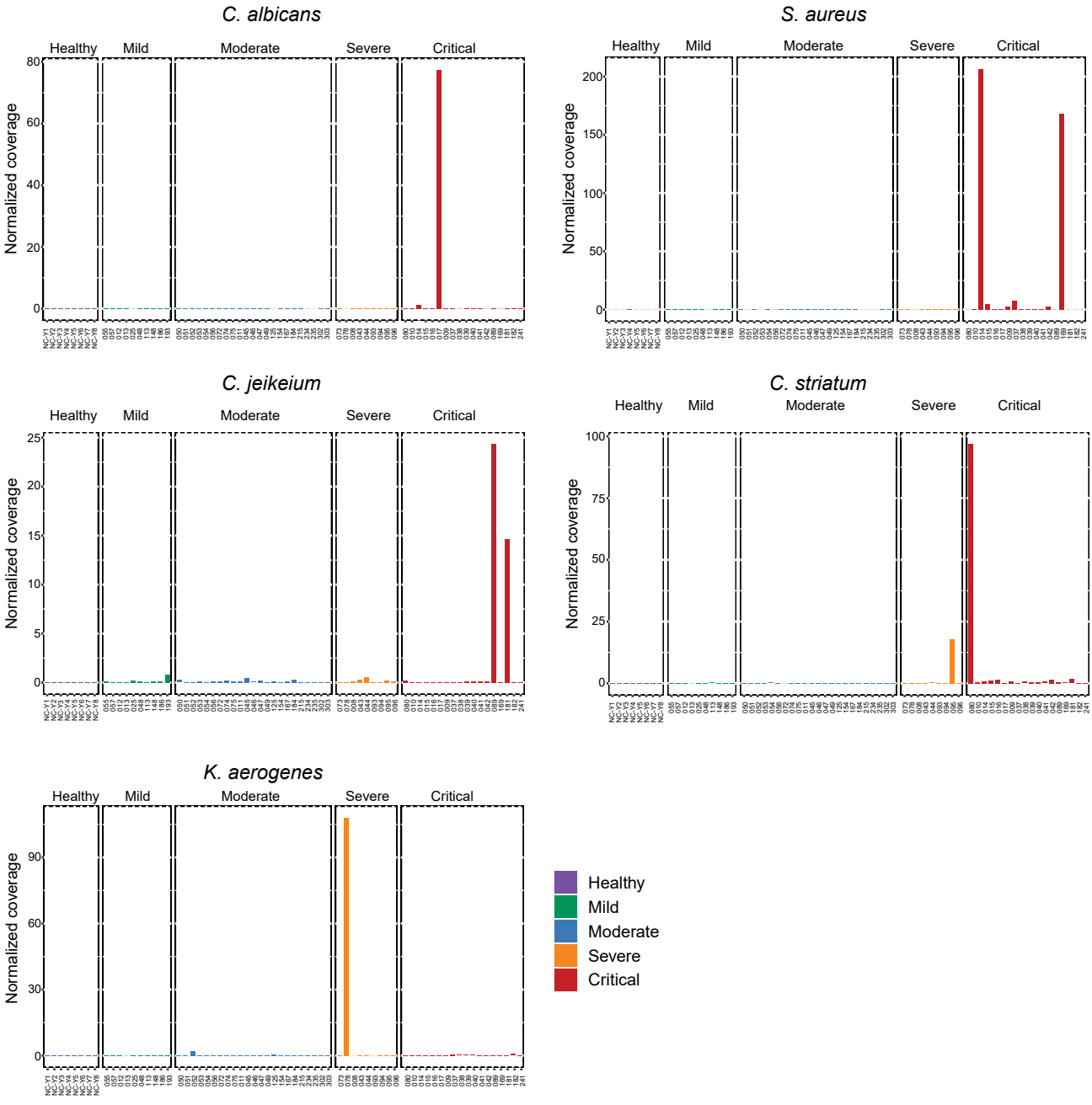

C. albicans
S. aureus
C. jeikeium
C. striatum
K. aerogenes
Healthy
Mild
Moderate
Severe
Critical
Normalized coverage

Figure S5

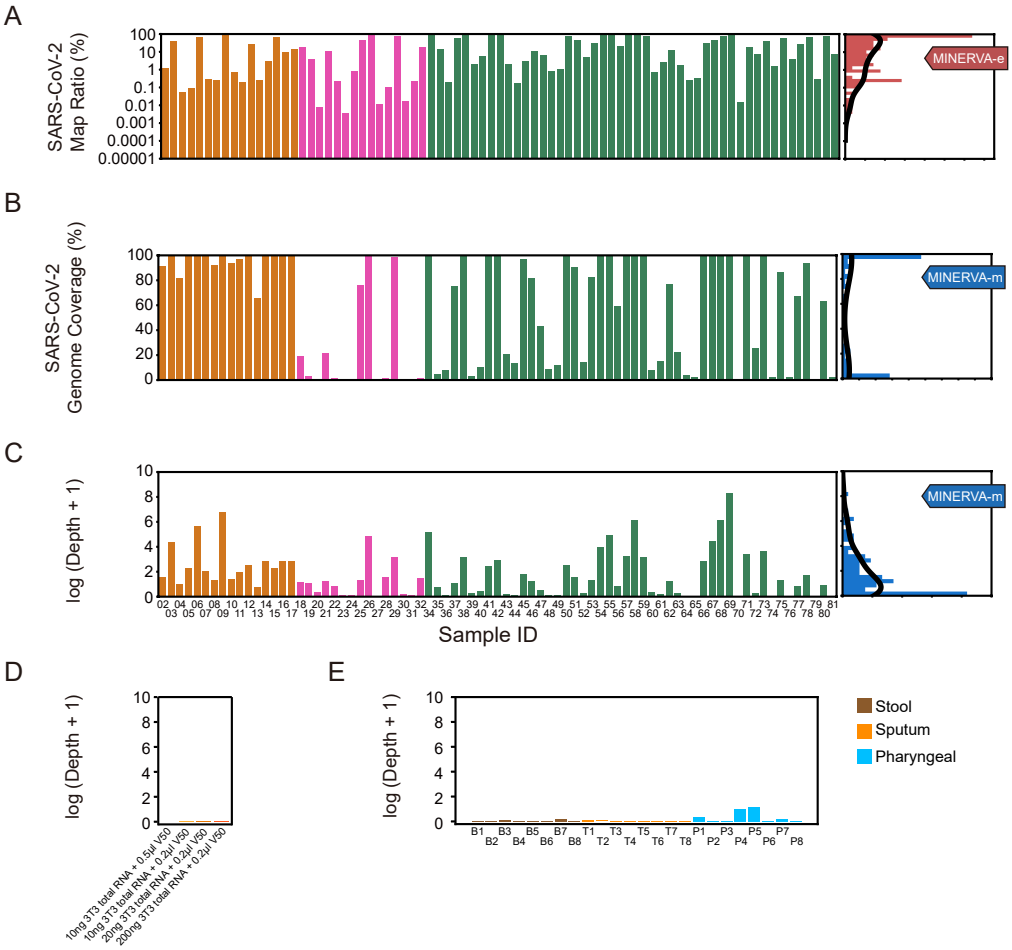

A
SARS-CoV-2
Map Ratio (%)
100
10
1
0.1
0.01
0.001
0.0001
0.00001
MINERVA-e
B
SARS-CoV-2
Genome Coverage (%)
100
80
60
40
20
0
MINERVA-m
C
log (Depth + 1)
10
8
6
4
2
0
MINERVA-m
Sample ID
D
log (Depth + 1)
10ng 3T3 total RNA + 0.5μl V5D
10ng 3T3 total RNA + 0.2μl V5D
20ng 3T3 total RNA + 0.2μl V5D
200ng 3T3 total RNA + 0.2μl V5D
E
log (Depth + 1)
Stool
Sputum
Pharyngeal

Figure S6

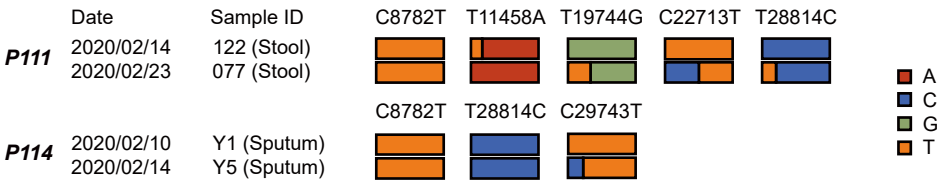

Date
Sample ID
C8782T
T11458A
T19744G
C22713T
T28814C
P111
2020/02/14
122 (Stool)
2020/02/23
077 (Stool)
C8782T
T28814C
C29743T
P114
2020/02/10
Y1 (Sputum)
2020/02/14
Y5 (Sputum)
A
C
G
T

Figure S7

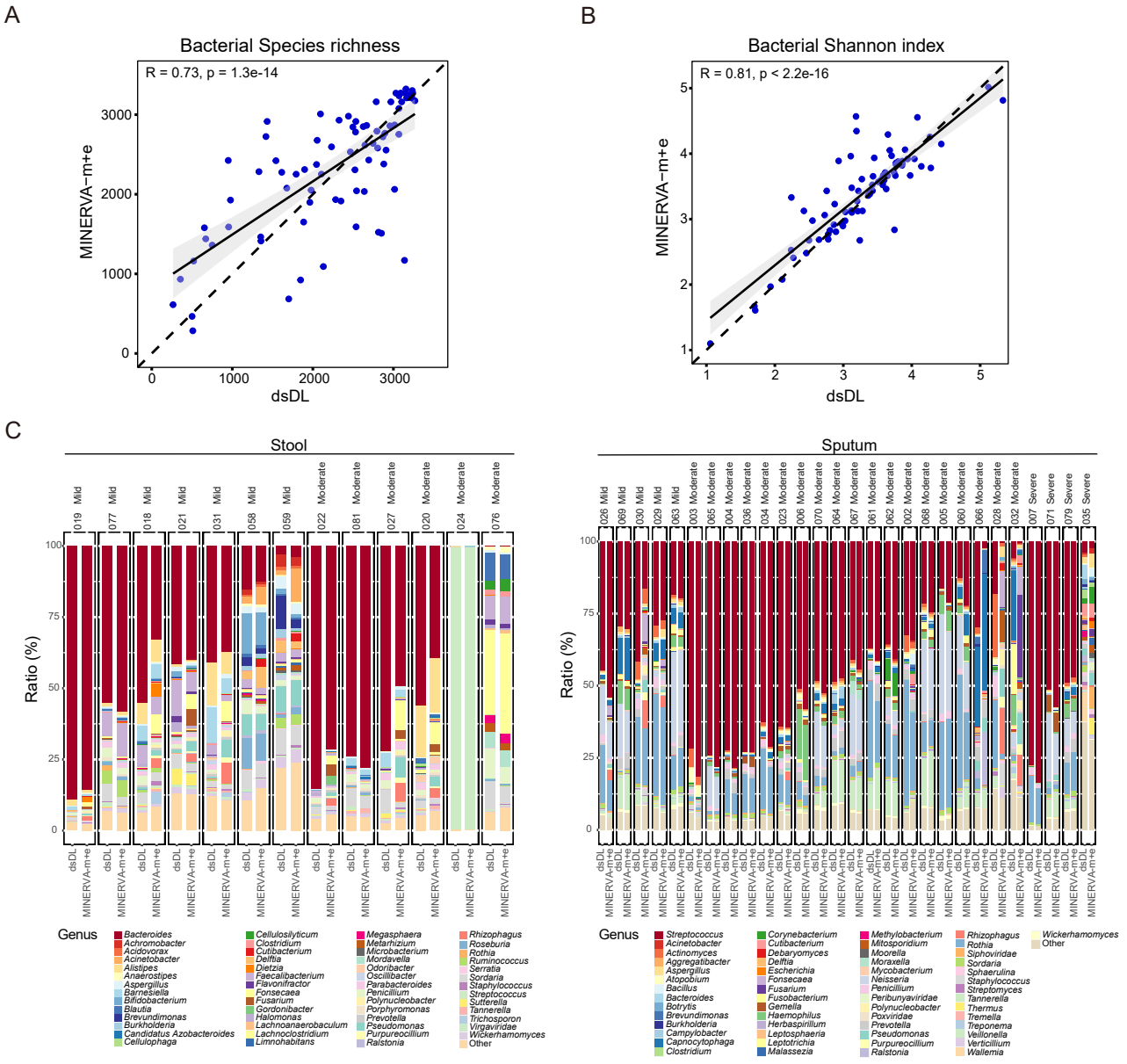

A
Bacterial Species richness
R = 0.73, p = 1.3e-14
MINERVA−m+e
dsDL
B
Bacterial Shannon index
R = 0.81, p < 2.2e-16
MINERVA−m+e
dsDL
C
Stool
Sputum
Ratio (%)
Genus
Bacteroides
Achromobacter
Acidovorax
Acinetobacter
Alistipes
Anaerostipes
Aspergillus
Barnesiella
Bifidobacterium
Blautia
Brevundimonas
Burkholderia
Candidatus Azobacteroides
Cellulophaga
Cellulosilyticum
Clostridium
Cutibacterium
Delftia
Dietzia
Faecalibacterium
Flavonifractor
Fonsecaea
Fusarium
Gordonibacter
Halomonas
Lachnoanaerobaculum
Lachnoclostridium
Limnohabitans
Megasphaera
Metarhizium
Microbacterium
Mordavella
Odoribacter
Oscillibacter
Parabacteroides
Penicillium
Polynucleobacter
Porphyromonas
Prevotella
Pseudomonas
Purpureocillium
Ralstonia
Rhizophagus
Roseburia
Rothia
Ruminococcus
Serratia
Sordaria
Staphylococcus
Streptococcus
Sutterella
Tannerella
Trichosporon
Virgaviridae
Wickerhamomyces
Other
Streptococcus
Acinetobacter
Actinomyces
Aggregatibacter
Aspergillus
Atopobium
Bacillus
Bacteroides
Botrytis
Brevundimonas
Burkholderia
Campylobacter
Capnocytophaga
Clostridium
Corynebacterium
Cutibacterium
Debaryomyces
Delftia
Escherichia
Fonsecaea
Fusarium
Fusobacterium
Gemella
Haemophilus
Herbaspirillum
Leptosphaeria
Leptotrichia
Malassezia
Methylobacterium
Mitosporidium
Moorella
Moraxella
Mycobacterium
Neisseria
Penicillium
Peribunyaviridae
Polynucleobacter
Poxviridae
Prevotella
Pseudomonas
Purpureocillium
Ralstonia
Rhizophagus
Rothia
Siphoviridae
Sordaria
Sphaerulina
Staphylococcus
Streptomyces
Tannerella
Thermus
Tremella
Treponema
Veillonella
Verticillium
Wallemia
Wickerhamomyces
Other

Table S1, related to Figure2

QC of All Sequencing Libraries

	MINERVA-e					MINERVA-m					dsDL			
Sample ID	Input (μl)	Total Reads	Mapping Ratio	Dedup Coverage	Dedup Depth	Input (μl)	Total Reads	Mapping Ratio	Dedup Coverage	Dedup Depth	Input (μl)	Mapping Ratio	Dedup Coverage	Dedup Depth
002	10.8	8,195,484	1.1638%	65.03%	4.9783	10.8	5,094,918	0.1307%	90.87%	4.0136	10.0	0.0004%	68.96%	1.4060
003	10.8	207,709,272	40.9510%	100.00%	946.9857	10.8	78,496,044	0.0640%	99.98%	77.9333	10.0	0.0196%	99.97%	323.3180
004	10.8	20,676,516	0.0524%	98.40%	5.2755	10.8	10,065,834	0.0126%	81.80%	1.7732	10.0	0.0001%	76.27%	1.7278
005	10.8	111,115,446	0.0926%	100.00%	61.0830	10.8	63,641,086	0.0092%	99.85%	8.8701	10.0	0.0001%	56.29%	0.9247
006	10.8	356,396,564	70.7456%	100.00%	3143.9166	10.8	78,571,822	0.2261%	100.00%	295.0117	10.0	0.0945%	99.97%	1469.0300
007	10.8	9,942,786	0.2915%	99.64%	10.7657	10.8	52,865,726	0.0101%	98.97%	6.5938	10.0	0.0002%	94.25%	3.8018
008	10.8	10,763,536	0.2497%	99.54%	8.5537	10.8	47,103,600	0.0067%	91.82%	2.7499	10.0	0.0008%	66.42%	2.2585
009	10.8	541,628,130	96.7444%	100.00%	3350.7269	10.8	21,166,616	5.9250%	100.00%	884.5224	10.0	4.7686%	99.95%	57.2028
010	10.8	5,668,752	0.7054%	54.96%	2.7434	10.8	4,089,390	0.1066%	93.61%	3.2721	10.0	0.0002%	62.59%	1.2704
011	10.8	11,176,696	0.2093%	99.99%	11.2174	10.8	41,913,790	0.0200%	96.90%	6.1212	10.0	0.0001%	87.87%	2.5211
012	10.8	16,151,064	27.1568%	99.84%	34.2213	10.8	40,226,610	0.0427%	99.72%	11.6552	10.0	0.0036%	99.86%	11.9888
013	10.8	14,881,342	0.2670%	96.25%	4.8453	10.8	7,335,290	0.0130%	65.48%	1.1660	10.0	0.0006%	99.96%	12.2455
014	10.8	13,884,580	3.0261%	99.99%	19.2685	10.8	21,795,080	0.1882%	99.55%	16.0357	10.0	0.0010%	62.14%	1.1220
015	10.8	12,387,588	65.9296%	99.90%	33.5811	10.8	3,105,472	0.3064%	99.49%	8.9043	10.0	0.0704%	99.87%	13.4546
016	10.8	13,025,918	9.2768%	99.84%	21.6711	10.8	19,495,766	0.3039%	99.55%	15.9261	10.0	0.0037%	84.08%	2.2149
017	10.8	16,384,556	13.2056%	99.86%	26.7111	10.8	32,164,850	0.1227%	99.88%	16.2674	10.0	0.0058%	31.20%	0.8448
018	2.7	112,794,866	19.1814%	77.76%	6.9234	2.7	35,616,834	0.0253%	18.55%	2.3147	10.0	0.0001%	78.13%	2.0642
019	2.7	93,872,216	4.1077%	64.13%	4.6651	2.7	35,775,154	0.0244%	2.94%	1.9927	10.0	0.0000%	53.13%	0.9280
020	2.7	82,493,546	0.0076%	54.44%	2.2178	2.7	36,822,898	0.0030%	0.12%	0.4650	10.0	0.0000%	13.77%	0.1814
021	2.7	92,874,254	11.0905%	75.64%	6.0904	2.7	22,813,548	0.0549%	21.45%	2.4093	10.0	0.0001%	61.54%	1.2370
022	2.7	106,391,154	0.2108%	68.52%	3.4031	2.7	37,797,862	0.0108%	1.27%	1.3720	10.0	0.0000%	23.47%	0.3527
023	2.7	135,954,444	0.0034%	67.87%	1.9198	2.7	36,422,158	0.0013%	0.13%	0.1155	10.0	0.0004%	0.76%	0.0201
024	2.7	68,556,412	0.7855%	58.16%	2.1196	2.7	32,646,446	0.0018%	0.12%	0.1614	10.0	0.0000%	22.52%	0.3292
025	2.7	201,614,222	47.7643%	99.81%	32.1785	2.7	34,297,778	0.0092%	75.47%	2.9363	10.0	0.0013%	99.99%	29.3693
026	2.7	642,140,604	92.8531%	100.00%	1621.9374	2.7	41,504,770	0.1649%	99.89%	124.7145	10.0	0.0694%	99.99%	1111.6309
027	2.7	28,519,004	0.0112%	92.50%	4.5196	2.7	34,802,762	0.0001%	0.11%	0.0351	10.0	0.0000%	10.32%	0.1326
028	2.7	9,792,364	0.1043%	76.38%	4.0190	2.7	22,201,210	0.0817%	1.61%	3.6903	10.0	0.0002%	39.17%	0.9250
029	2.7	175,344,692	75.5914%	100.00%	327.4946	2.7	33,195,792	0.0341%	98.64%	22.5668	10.0	0.0184%	99.98%	235.8581
030	2.7	25,325,652	0.0177%	94.24%	5.2230	2.7	36,375,652	0.0034%	0.19%	0.2462	10.0	0.0000%	21.75%	0.4325
031	2.7	51,558,410	0.2134%	98.55%	10.2587	2.7	48,321,376	0.0011%	0.47%	0.0996	10.0	0.0001%	12.83%	0.3015
032	2.7	1,638,196	17.5565%	82.31%	3.1014	2.7	2,855,228	2.6550%	0.89%	3.5194	10.0	0.0508%	49.21%	7.7024
034	5.4	146,373,482	90.6535%	100.00%	7359.4100	5.4	10,605,224	0.7548%	100.00%	176.2427	10.0	0.1662%	99.98%	842.7465
035	5.4	749,392	13.5239%	33.49%	1.1914	5.4	4,854,666	0.0509%	4.67%	1.2298	10.0	0.0041%	32.24%	0.5876
036	5.4	66,509,586	0.1866%	99.96%	23.6832	5.4	111,047,572	0.0001%	7.79%	0.1038	10.0	0.0003%	57.51%	1.0422
037	5.4	9,834,452	60.1041%	98.86%	10.6667	5.4	4,406,210	0.0336%	74.81%	2.0089	10.0	0.0201%	56.10%	1.2234
038	5.4	43,022,752	97.0263%	100.00%	252.6044	5.4	1,123,198	0.9702%	99.99%	23.4852	10.0	0.5773%	99.85%	17.2460
039	5.4	1,340,114	2.0089%	4.79%	0.4514	5.4	3,976,390	0.0153%	2.83%	0.3394	10.0	0.0009%	24.95%	0.4494
040	5.4	3,628,686	5.6932%	24.91%	0.8514	5.4	4,912,264	0.0177%	9.88%	0.5260	10.0	0.0008%	19.53%	0.3346
041	5.4	29,918,078	83.4096%	99.86%	48.3122	5.4	5,950,044	0.1372%	98.88%	10.4956	10.0	0.1244%	99.91%	25.1505
042	5.4	63,305,808	92.6539%	99.99%	242.1007	5.4	2,034,818	0.4434%	99.61%	17.8696	10.0	0.1145%	99.97%	75.3447
043	5.4	28,502,096	2.0792%	98.90%	8.6775	5.4	21,291,192	0.0006%	20.84%	0.2769	10.0	0.0008%	43.21%	0.8251
044	5.4	23,678,350	0.1632%	73.65%	2.0966	5.4	15,802,662	0.0008%	13.19%	0.1653	10.0	0.0004%	31.84%	0.5936
045	5.4	28,763,804	3.0793%	99.93%	15.9794	5.4	19,682,236	0.0196%	96.32%	4.9701	10.0	0.0012%	97.70%	5.9600
046	5.4	48,033,536	10.5088%	100.00%	158.1837	5.4	31,002,106	0.0034%	81.51%	2.4673	10.0	0.0041%	99.96%	41.5512
047	5.4	32,374,174	6.1377%	99.89%	17.6006	5.4	20,075,024	0.0017%	42.98%	0.7470	10.0	0.0014%	84.73%	2.6273
048	5.4	33,545,522	0.8612%	95.79%	6.1428	5.4	24,579,728	0.0003%	8.20%	0.1238	10.0	0.0008%	54.43%	1.1477
049	5.4	41,228,324	1.0147%	99.88%	18.1782	5.4	29,981,390	0.0003%	11.99%	0.1741	10.0	0.0005%	91.95%	3.3415
050	5.4	12,177,692	71.4756%	99.90%	121.7697	5.4	22,754,546	0.0264%	99.44%	11.6945	10.0	0.0421%	99.96%	51.3386
051	5.4	6,105,636	46.9134%	99.78%	28.5245	5.4	21,784,580	0.0109%	90.56%	3.7560	10.0	0.0058%	99.95%	25.2443
052	5.4	3,830,746	4.8117%	90.86%	4.5674	5.4	24,464,202	0.0008%	14.20%	0.2901	10.0	0.0005%	89.58%	2.9335
053	5.4	6,556,106	32.5097%	100.00%	72.9775	5.4	25,666,594	0.0055%	82.54%	2.7256	10.0	0.0033%	99.96%	47.4457
054	5.4	45,398,712	88.7216%	100.00%	1074.7572	5.4	28,016,362	0.0849%	100.00%	52.9310	10.0	0.0896%	100.00%	843.4279
055	5.4	103,807,862	94.1763%	100.00%	3139.9488	5.4	25,661,784	0.2476%	100.00%	139.5156	10.0	0.1555%	99.99%	1233.6985
056	5.4	5,072,904	21.2207%	99.86%	38.2663	5.4	25,257,896	0.0029%	59.39%	1.3265	10.0	0.0020%	99.96%	37.2734
057	5.4	26,651,568	81.5262%	100.00%	764.1384	5.4	23,771,526	0.0475%	99.90%	25.3226	10.0	0.0392%	100.00%	773.4338
058	5.4	119,250,340	89.4189%	100.00%	1499.0878	5.4	21,737,454	1.3970%	99.99%	481.7897	10.0	0.8035%	99.96%	18.0897
059	5.4	13,876,992	73.0525%	99.98%	94.7249	5.4	6,897,404	0.2907%	99.98%	23.3441	10.0	0.0756%	60.08%	1.0516
060	5.4	5,064,692	0.7576%	81.74%	2.8944	5.4	28,676,654	0.0022%	7.97%	0.4627	10.0	0.0004%	46.13%	0.9281
061	5.4	3,237,272	2.4584%	94.59%	4.1871	5.4	12,599,798	0.0007%	14.80%	0.1898	10.0	0.0003%	38.29%	0.6626
062	5.4	6,685,592	12.8404%	99.96%	46.9370	5.4	18,066,044	0.0082%	76.60%	2.4364	10.0	0.0048%	99.89%	11.2049
063	5.4	6,482,172	1.7973%	96.84%	5.4459	5.4	23,768,308	0.0006%	21.74%	0.2900	10.0	0.0005%	77.05%	2.1138
064	5.4	9,466,060	0.2669%	96.99%	4.5785	5.4	22,213,606	0.0001%	3.55%	0.0437	10.0	0.0001%	49.20%	0.8110
065	5.4	5,783,354	0.3323%	76.84%	1.9841	5.4	24,375,006	0.0001%	2.16%	0.0297	10.0	0.0000%	47.11%	0.6932
066	5.4	5,970,510	29.3245%	99.91%	30.3217	5.4	22,078,330	0.0951%	99.92%	16.9899	10.0	0.0290%	92.86%	4.1057
067	5.4	8,697,652	47.3466%	100.00%	768.4558	5.4	23,501,026	0.1764%	100.00%	88.7774	10.0	0.1034%	99.97%	490.0341
068	5.4	25,258,898	76.1765%	100.00%	4894.2700	5.4	28,494,342	0.8557%	100.00%	475.1583	10.0	0.5174%	99.96%	177.1438
069	5.4	141,070,280	96.2193%	100.00%	7846.8553	5.4	28,135,628	13.5652%	100.00%	4153.0717	10.0	8.8958%	100.00%	6361.1653
070	5.4	5,686,808	0.0154%	77.96%	1.8157	5.4	24,052,452	0.0000%	0.54%	0.0102	10.0	0.0010%	99.68%	9.4708
071	5.4	9,016,134	18.4892%	100.00%	506.1078	5.4	30,946,146	0.0458%	99.95%	30.1117	10.0	0.0199%	99.96%	26.0630
072	5.4	7,900,462	7.0146%	99.99%	23.4415	5.4	10,547,490	0.0016%	25.25%	0.3908	10.0	0.0018%	99.96%	18.4789
073	5.4	8,211,094	39.8226%	99.99%	87.6919	5.4	20,473,706	0.1506%	99.99%	38.6144	10.0	0.0774%	99.96%	16.5545
074	5.4	13,940,912	1.6234%	54.47%	1.7018	5.4	10,982,782	0.0001%	1.71%	0.0198	10.0	0.0023%	99.95%	43.3575
075	5.4	53,210,982	59.1457%	100.00%	148.0972	5.4	12,675,448	0.0103%	86.27%	2.8130	10.0	0.0057%	99.62%	8.0070
076	5.4	5,350,422	4.1008%	38.45%	0.6652	5.4	4,988,872	0.0003%	2.41%	0.0256	10.0	0.0005%	99.45%	10.3293
077	5.4	80,704,356	27.2584%	99.68%	20.6439	5.4	41,708,526	0.0016%	67.15%	1.3075	10.0	0.0012%	99.59%	10.0565
078	5.4	87,431,520	63.3773%	99.92%	55.5415	5.4	21,631,534	0.0194%	93.61%	4.5826	10.0	0.0063%	89.19%	2.5408
079	5.4	27,533,572	0.3070%	45.27%	0.7658	5.4	16,031,304	0.0000%	0.12%	0.0042	10.0	0.0003%	98.62%	6.1102
080	5.4	3,872,932	72.5768%	99.34%	14.4106	5.4	1,250,890	0.0831%	63.00%	1.5866	10.0	0.1389%	97.92%	5.2081
081	5.4	4,648,190	7.6058%	98.09%	9.4652	5.4	3,511,362	0.0003%	2.11%	0.0210	10.0	0.0007%	99.94%	13.4037
083	5.4	10,026,796	83.5902%	100.00%	178.1880	5.4	7,883,708	0.0169%	68.29%	3.7261	/	/	/	/
085	5.4	18,520,938	32.5616%	99.96%	56.0855	5.4	36,049,244	0.0032%	79.45%	3.0384	/	/	/	/
086	5.4	34,368,266	51.9656%	100.00%	133.5960	5.4	45,682,850	0.0066%	97.18%	8.8622	/	/	/	/
088	5.4	12,826,026	0.7087%	51.79%	1.0182	5.4	32,102,908	0.0015%	25.98%	0.4293	/	/	/	/
089	5.4	43,399,718	0.1826%	87.32%	2.7735	5.4	180,096,594	0.0020%	98.93%	6.8831	/	/	/	/
093	5.4	35,612,300	5.3625%	99.86%	34.4695	5.4	96,879,110	0.0005%	50.67%	1.3602	/	/	/	/
094	5.4	28,681,322	26.5076%	98.44%	13.6589	5.4	58,002,768	0.0024%	81.96%	3.6088	/	/	/	/
095	5.4	23,106,298	3.0988%	99.50%	11.2516	5.4	73,228,566	0.0003%	23.60%	0.4939	/	/	/	/
096	5.4	15,588,394	0.2395%	34.38%	0.6021	5.4	35,386,008	0.0002%	0.63%	0.0609	/	/	/	/
102	5.4	5,346,588	0.5901%	64.35%	1.3610	5.4	77,690,126	0.0002%	3.72%	0.1436	/	/	/	/
113	5.4	16,728,222	8.2095%	98.26%	8.0897	5.4	57,925,030	0.0009%	20.78%	0.6608	/	/	/	/
121	5.4	31,507,858	0.1795%	15.50%	1.4324	5.4	122,096,044	0.0007%	39.08%	0.7721	/	/	/	/
122	5.4	39,411,066	8.5777%	99.39%	10.5704	5.4	93,269,750	0.0009%	46.95%	1.0274	/	/	/	/
123	5.4	36,100,522	74.6819%	100.00%	213.5300	5.4	36,077,758	0.0052%	94.85%	6.0883	/	/	/	/
125	5.4	248,648	59.4748%	3.68%	1.4559	5.4	295,912	0.0703%	12.55%	0.2656	/	/	/	/
128	5.4	44,498,056	36.2687%	99.84%	39.7181	5.4	47,908,876	0.0022%	32.81%	0.8910	/	/	/	/
145	5.4	116,748,088	87.2807%	91.26%	14.9366	5.4	7,389,184	0.0390%	23.30%	0.7610	/	/	/	/
146	5.4	15,901,590	77.2499%	100.00%	424.3760	5.4	3,819,436	0.0114%	46.82%	1.0621	/	/	/	/
147	5.4	17,456,218	78.8375%	38.98%	3.6053	5.4	6,994,844	0.0164%	19.16%	0.7468	/	/	/	/
148	5.4	139,565,182	46.5267%	95.02%	14.5469	5.4	131,964,738	0.0027%	28.84%	0.9558	/	/	/	/
149	5.4	245,791,438	69.0462%	96.67%	22.6291	5.4	118,696,898	0.0152%	42.45%	2.6673	/	/	/	/
150	5.4	154,011,282	69.8185%	90.38%	14.1609	5.4	99,944,572	0.0187%	25.71%	2.5604	/	/	/	/
154	5.4	252,038,586	100.4621%	100.00%	420.7170	5.4	2,200,888	1.2509%	99.87%	50.0489	/	/	/	/
167	5.4	108,884,390	23.3966%	99.81%	22.4000	5.4	72,888,692	0.0008%	59.01%	1.5816	/	/	/	/
168	5.4	25,285,700	0.9654%	98.68%	6.1981	5.4	6,631,934	0.0010%	17.79%	0.2290	/	/	/	/
169	5.4	6,380,892	45.4035%	29.26%	1.2985	5.4	1,315,150	0.0056%	6.48%	0.1197	/	/	/	/
170	5.4	102,753,144	3.0334%	83.14%	6.2282	5.4	94,394,542	0.0032%	30.69%	1.3714	/	/	/	/
174	5.4	39,535,420	48.6137%	75.42%	9.3010	5.4	23,470,828	0.0415%	45.00%	3.7544	/	/	/	/
177	5.4	86,654,130	90.5498%	92.93%	23.5223	5.4	8,160,156	0.1322%	70.11%	5.8358	/	/	/	/
180	5.4	68,021,724	0.0562%	91.73%	3.9670	5.4	56,507,160	0.0005%	21.98%	0.3774	/	/	/	/
181	5.4	23,772,212	0.0114%	95.38%	5.2273	5.4	17,507,066	0.0010%	8.36%	0.2010	/	/	/	/
182	5.4	41,601,964	0.1233%	84.18%	2.9324	5.4	28,846,054	0.0006%	12.65%	0.2351	/	/	/	/
183	5.4	75,925,168	5.4340%	98.46%	8.5665	5.4	62,049,032	0.0005%	36.81%	0.6766	/	/	/	/
184	5.4	51,000,162	1.0074%	86.07%	3.3773	5.4	56,335,394	0.0004%	24.62%	0.4155	/	/	/	/
185	5.4	72,999,296	13.3787%	98.48%	11.9794	5.4	64,798,486	0.0008%	53.78%	1.1309	/	/	/	/
186	5.4	17,335,644	89.1650%	53.28%	3.8052	5.4	5,380,772	0.0186%	25.79%	0.8651	/	/	/	/
189	5.4	374,145,632	95.0024%	100.00%	1582.2800	5.4	111,729,018	0.0242%	99.38%	75.1353	/	/	/	/
192	5.4	60,506,220	12.7664%	98.28%	11.9391	5.4	94,939,084	0.0004%	20.52%	0.4943	/	/	/	/
193	5.4	72,845,794	13.4477%	99.83%	24.8392	5.4	119,377,918	0.0005%	36.36%	0.9976	/	/	/	/

194	5.4	113,929,174	3.5935%	94.84%	6.6863	5.4	74,882,784	0.0003%	10.12%	0.2392	/	/	/	/
195	5.4	231,325,932	56.4214%	99.93%	128.3160	5.4	79,420,208	0.0013%	73.32%	2.4829	/	/	/	/
202	5.4	95,943,060	1.7786%	99.32%	17.7605	5.4	89,627,594	0.0003%	5.01%	0.1965	/	/	/	/
212	5.4	46,499,962	3.9862%	23.11%	8.1436	5.4	38,353,226	0.0004%	5.08%	0.1772	/	/	/	/
215	5.4	2,730,220	5.1572%	10.83%	8.0875	5.4	6,576,196	0.0157%	1.33%	0.7010	/	/	/	/
217	5.4	161,866,574	0.1522%	36.96%	8.0366	5.4	100,830,192	0.0001%	0.24%	0.0928	/	/	/	/
230	5.4	185,858,560	49.0328%	99.85%	166.8540	5.4	129,451,404	0.0022%	94.36%	6.8372	/	/	/	/
234	5.4	715,080	5.5852%	27.33%	0.4704	5.4	2,587,668	0.0027%	0.39%	0.0659	/	/	/	/
235	5.4	645,750	89.4572%	42.31%	0.7548	5.4	785,898	0.0015%	0.93%	0.0190	/	/	/	/
241	5.4	9,439,472	86.2488%	79.06%	12.7547	5.4	7,703,008	0.0373%	48.17%	2.7466	/	/	/	/
302	5.4	24,776,174	0.2073%	26.62%	0.8079	5.4	54,812,846	0.0009%	0.50%	0.2893	/	/	/	/
303	5.4	144,586,436	0.0590%	76.90%	2.5174	5.4	144,176,290	0.0002%	0.12%	0.1328	/	/	/	/
Protocol Optimization														
02						2.7	16,408,970	0.1012%	88.32%	4.8521				
03						2.7	47,083,344	0.0520%	99.96%	40.6153				
04						2.7	57,543,194	0.0027%	44.43%	0.7633				
05						2.7	59,021,696	0.0019%	41.02%	0.6959				
06						2.7	43,005,642	0.1520%	100.00%	114.7798				
07	2.7	6,807,324	0.2123%	92.42%	3.2703	2.7	59,245,398	0.0048%	93.44%	4.4251				
08	2.7	7,133,858	0.2009%	95.80%	4.2571	2.7	53,691,494	0.0057%	91.36%	4.0402				
09	2.7	534,740,522	97.9777%	100.00%	2185.2300	2.7	17,555,514	6.0440%	100.00%	378.9593	Input volume: 2.7 µl vs 10.8 µl			
10						2.7	17,409,862	0.1325%	84.16%	5.2398				
11	2.7	58	0.0000%	0.00%	0.0000	2.7	1,618	0.0618%	0.10%	0.0010				
12	2.7	17,074,722	30.2408%	99.61%	15.5067	2.7	47,310,582	0.0206%	98.08%	9.1497				
13						2.7	48,064,580	0.0035%	41.59%	0.7463				
14	2.7	12,514,502	4.9961%	98.33%	7.4607	2.7	28,491,880	0.1188%	99.49%	18.5844				
15						2.7	18,079,484	0.2210%	98.69%	10.7718				
16	2.7	8,061,366	18.2108%	94.00%	4.9191	2.7	15,778,084	0.1176%	98.85%	12.2731				
17	2.7	17,317,738	12.2331%	99.49%	11.8779	2.7	45,806,156	0.0700%	99.44%	17.8944				
50	2.7	5,302,774	71.9360%	99.74%	56.2274	2.7	9,440,676	0.0260%	94.59%	4.86135	Input volume: 2.7 µl vs 5.4 µl			
		6,874,918	71.0528%	99.90%	68.0711		13,313,870	0.0267%	97.44%	6.94325				
51	2.7	2,651,528	45.9457%	99.27%	12.73	2.7	9,956,018	0.0106%	66.31%	1.68468				
		3,454,108	47.6123%	99.75%	16.0504		11,828,562	0.0112%	75.74%	2.11945				
52	2.7	1,812,456	5.1702%	78.61%	2.70916	2.7	11,905,088	0.0011%	7.45%	0.166271				
		2,018,290	4.4788%	68.52%	1.99716		12,559,114	0.0006%	8.10%	0.129552				
53	2.7	2,661,912	32.8762%	99.97%	33.6088	2.7	11,393,230	0.0058%	56.44%	1.28014				
		3,894,194	32.2177%	99.91%	40.2593		14,273,364	0.0054%	64.91%	1.47253				
54	2.7	16,862,842	88.9164%	100.00%	554.182	2.7	10,453,800	0.0886%	99.98%	21.1365				
		28,535,870	88.4747%	99.98%	623.171		17,562,562	0.0827%	100.00%	32.3643				
55	2.7	50,308,172	94.4399%	100.00%	1763.58	2.7	12,242,678	0.2470%	99.97%	68.163				
		53,499,690	93.7940%	100.00%	1910.21		13,419,106	0.2482%	100.00%	75.0258				
56	2.7	2,407,594	21.5466%	99.75%	18.4627	2.7	11,510,576	0.0029%	34.62%	0.621309				
		2,665,310	20.8997%	99.82%	20.063		13,747,320	0.0029%	39.51%	0.716082				
57	2.7	12,058,338	80.8714%	100.00%	375.207	2.7	10,977,398	0.0477%	99.15%	11.8002				
		14,593,230	81.9503%	100.00%	438.47		12,794,128	0.0474%	99.35%	13.7123				
42	5.4	63,305,808	92.6539%	99.99%	242.1007	5.4	2,034,818	0.4434%	99.61%	17.8696	With post-added carrier RNA			
43	5.4	28,502,096	2.0792%	98.90%	8.6775	5.4	21,291,192	0.0006%	20.84%	0.2769				
44	5.4	23,678,350	0.1632%	73.65%	2.0966	5.4	15,802,662	0.0008%	13.19%	0.1653				
45	5.4	28,763,804	3.0793%	99.93%	15.9794	5.4	19,682,236	0.0196%	96.32%	4.9701				
46	5.4	48,033,536	10.5088%	100.00%	158.1837	5.4	31,002,106	0.0034%	81.51%	2.4673				
47	5.4	32,374,174	6.1377%	99.89%	17.6006	5.4	20,075,024	0.0017%	42.98%	0.747				
48	5.4	33,545,522	0.8612%	95.79%	6.1428	5.4	24,579,728	0.0003%	8.20%	0.1238				
49	5.4	41,228,324	1.0147%	99.88%	18.1782	5.4	29,981,390	0.0003%	11.99%	0.1741				
Y1	5.4	1,267,440	56.3684%	93.93%	5.5805	5.4	751,372	0.0527%	25.64%	0.5249	RNA extraction with carrier RNA			
Y2	5.4	75,641,774	86.5550%	100.00%	4395.5348	5.4	5,811,906	0.4477%	99.98%	58.3571				
Y3	5.4	40,132,820	38.8485%	100.00%	145.8907	5.4	29,519,046	0.0144%	98.34%	9.2615				
Y4	5.4	21,566,432	11.2625%	99.85%	32.3432	5.4	22,669,654	0.0036%	70.52%	1.6935				
Y5	5.4	9,021,160	2.7007%	86.56%	3.0329	5.4	8,699,170	0.0011%	7.99%	0.1337				
Y6	5.4	18,074,214	97.7771%	100.00%	968.249	5.4	18,398	30.6175%	99.58%	13.0292				
Negative Control														
Control	/	149,388	0.1787%	8.26%	0.1239	/	836,522	0.0004%	0.10%	0.0020	Non-template control			
NC-B1	5.4	28,506,496	0.0097%	3.49%	0.0423	?	22,207,000	0.0000%	0.49%	0.0072	Normal human stool			
NC-B2	5.4	30,122,844	0.0003%	4.03%	0.0517	?	29,775,692	0.0001%	0.10%	0.0060				
NC-B3	5.4	71,722,722	0.0142%	8.11%	0.1009	?	24,740,524	0.0001%	0.10%	0.0061				
NC-B4	5.4	16,824,960	0.0092%	4.66%	0.0607	?	14,371,532	0.0000%	0.00%	0.0000				
NC-B5	5.4	15,111,452	0.0232%	3.33%	0.0596	?	17,902,254	0.0000%	0.35%	0.0070				
NC-B6	5.4	31,748,618	0.0166%	1.45%	0.0243	?	29,428,006	0.0001%	0.44%	0.0132				
NC-B7	5.4	25,058,140	0.0096%	16.25%	0.2426	?	29,213,710	0.0002%	0.10%	0.0162				
NC-B8	5.4	26,839,054	0.0003%	0.34%	0.0067	?	24,241,724	0.0001%	0.10%	0.0061				
NC-T1	5.4	490,128	31.5238%	4.55%	0.1510	?	3,029,996	0.1411%	2.69%	0.3196	Normal human sputum			
NC-T2	5.4	40,241,182	0.3646%	2.48%	0.0822	?	30,330,186	0.0169%	1.25%	0.3536				
NC-T3	5.4	57,892,572	0.0009%	1.69%	0.0662	?	44,494,444	0.0093%	1.22%	0.2657				
NC-T4	5.4	10,104,442	0.0061%	0.10%	0.0700	?	8,905,284	0.0428%	1.41%	0.2725				
NC-T5	5.4	9,157,200	0.0117%	0.24%	0.0611	?	9,345,238	0.0620%	1.54%	0.3442				
NC-T6	5.4	12,384,160	0.0023%	3.58%	0.0569	?	34,626,836	0.0143%	1.43%	0.2997				
NC-T7	5.4	22,771,792	0.0021%	1.56%	0.0383	?	80,571,702	0.0003%	0.51%	0.0706				
NC-T8	5.4	2,326,752	0.0114%	1.40%	0.0532	?	12,203,758	0.0406%	1.72%	0.2943				
NC-P1	5.4	1,983,706	22.1787%	17.89%	0.4204	?	21,048,464	0.0862%	15.73%	0.7194	Normal human pharyngeal			
NC-P2	5.4	741,534	1.0643%	0.56%	0.0472	?	6,789,294	0.2767%	2.64%	0.5605				
NC-P3	5.4	719,086	0.0152%	0.22%	0.0255	?	10,691,944	0.1372%	1.82%	0.4917				
NC-P4	5.4	25,966,836	24.9029%	37.63%	1.8097	?	52,143,516	0.0357%	31.96%	1.1184				
NC-P5	5.4	13,752,522	66.3305%	44.61%	2.2808	?	18,214,848	0.0910%	34.26%	1.1053				
NC-P6	5.4	3,162,132	0.0043%	2.01%	0.0312	?	29,562,108	0.0497%	3.04%	0.5153				
NC-P7	5.4	1,037,626	35.8005%	10.33%	0.2497	?	11,182,330	0.1426%	8.13%	0.5826				
NC-P8	5.4	5,959,374	1.1421%	2.22%	0.0585	?	66,771,636	0.0197%	3.19%	0.4318				

Table S2, related to Figure 2

Metagenomic Profiles of MINERVA-m Libraries

Sample ID	Human Ratio	Nonhuman Reads	Viral Ratio (to Nonhuman)	Fungal Ratio (to Nonhuman)	Bacterial Ratio (to Nonhuman)
002	0.189022659	494070	0.011172506	0.04548141	0.523413686
003	0.003152193	8023735	0.005855752	0.018087462	0.708748606
004	0.002247542	787922	0.003186864	0.018070317	0.759708448
005	0.003376301	2361502	0.004108402	0.036782522	0.800813635
006	0.001556586	4297298	0.02418799	0.023355141	0.683550454
007	0.002152436	3490992	0.002329997	0.021135826	0.804028769
008	0.002496948	4366334	0.003062752	0.022149932	0.81782406
009	0.235250651	1513285	0.417012658	0.063244531	0.168218809
010	0.100165824	292498	0.010495798	0.050437952	0.34255277
011	0.005796999	4936618	0.002005219	0.018073102	0.691540443
012	0.056740197	2872201	0.004992687	0.025371483	0.68052793
013	0.002099317	995977	0.003084409	0.014818615	0.849455359
014	0.07402999	1012355	0.014750754	0.104524599	0.300168419
015	0.137521869	132205	0.029219772	0.136674105	0.296289853
016	0.128452996	919591	0.01303623	0.157629859	0.300164965
017	0.075050746	3373926	0.004681786	0.519567116	0.127690708
018	0.004962925	273574	0.002408855	0.300960618	0.380547859
019	0.002035641	1126424	0.001543824	0.092432335	0.740606557
020	0.002011489	226583	0.001288711	0.275642921	0.425058367
021	0.021512964	384049	0.00883481	0.227106437	0.434988764
022	0.001262345	467215	0.001988378	0.178305491	0.5252036
023	0.000967573	2612059	0.002827272	0.087083791	0.6101451
024	0.003675512	14482545	0.989738061	0.004180481	0.002953141
025	0.002920678	3314711	0.002423741	0.078022187	0.613572948
026	0.004236899	1924866	0.021695536	0.104623906	0.691714644
027	0.00113097	309059	0.005015224	0.283327779	0.508368952
028	0.862835618	1187883	0.007710355	0.312298433	0.17533545
029	0.007660008	1480193	0.006239051	0.120162709	0.598969864
030	0.318459812	1834920	0.004199638	0.242711399	0.380481438
031	0.004161361	1035181	0.008846762	0.172939805	0.437442341
032	0.661264484	338880	0.011101275	0.389633499	0.219154273
034	0.016808897	835001	0.050602335	0.027046674	0.694854258
035	0.787935399	619841	0.020623031	0.075046665	0.115150498
036	0.001534898	5513710	0.003862916	0.041223967	0.747251306
037	0.219956548	184865	0.008086982	0.195093717	0.4835745
038	0.260678457	37991	0.144455266	0.144165723	0.352662473
039	0.411262577	263095	0.010650145	0.096839545	0.363526483
040	0.337620057	286824	0.011466962	0.097931135	0.36846289
041	0.159757375	263758	0.022672298	0.079171058	0.479014855
042	0.523014344	71365	0.069740069	0.099110208	0.239879493
043	0.002240742	1419615	0.0016589	0.034086707	0.649658534
044	0.005322417	1579626	0.001700403	0.027012723	0.650570451
045	0.002754412	1952079	0.002996293	0.026872888	0.653770672
046	0.000702947	2672772	0.002713288	0.024649315	0.725376126
047	0.004772987	1278475	0.001493967	0.032775768	0.6517296
048	0.001687305	2318133	0.003838865	0.024468829	0.707022418
049	0.002643559	2696632	0.001065032	0.02367138	0.646335132
050	0.007682803	3220837	0.003400979	0.02947681	0.6994865
051	0.016305143	4544027	0.005112866	0.021789923	0.897927543
052	0.014083711	3064879	0.00211656	0.03112423	0.640877177
053	0.005910151	2732709	0.001599146	0.030647244	0.600823944
054	0.002928867	3265342	0.00610962	0.02664897	0.606791877
055	0.001797654	3837017	0.014074214	0.024665254	0.635461871
056	0.001543522	3581067	0.000865385	0.022282465	0.665598549
057	0.003176819	3346679	0.004133052	0.024325906	0.670528306
058	0.022796132	638484	0.236040371	0.078437361	0.403385206
059	0.057016346	270289	0.033819356	0.060857083	0.284902456
060	0.017012024	2339720	0.001854495	0.025785564	0.59654916
061	0.000922177	939891	0.001568267	0.043892324	0.639443297
062	0.005408816	1399198	0.002913097	0.033985898	0.715117517
063	0.001247926	1141437	0.00111088	0.056198459	0.620442477
064	0.001482335	1121648	0.001077878	0.036808339	0.582269125
065	0.001927175	1167864	0.004348965	0.042675346	0.793624086
066	0.079589675	2729856	0.010843063	0.022070761	0.693472843
067	0.002689423	1612208	0.014605436	0.040224959	0.627299951
068	0.008419602	2602295	0.047991869	0.025559362	0.540959422
069	0.089599293	3831180	0.493960085	0.01486435	0.312107236
070	0.001069733	1187881	0.00220645	0.044801626	0.621201114
071	0.068376486	2455028	0.004898926	0.029558115	0.731783507
072	0.00499718	1036507	0.001826326	0.032783184	0.670467252
073	0.032560807	1503578	0.012516145	0.050292037	0.68134144
074	0.001041278	1496732	0.001725092	0.025977931	0.811353001
075	0.000682444	849815	0.00227932	0.02841795	0.671447315
076	0.001104006	104031	0.001278465	0.10538205	0.783900953
077	0.001377956	664662	0.000926787	0.0765216	0.695597161
078	0.01292382	1170985	0.006317758	0.076883991	0.691229179
079	0.005025328	1076912	0.00219238	0.043353589	0.688387723

080	0.497480386	65519	0.018467925	0.16599765	0.312825287
081	0.00023917	107785	0.000129888	0.078610196	0.764392077
Y1	0.245092935	73250	0.003740614	0.036641638	0.306197952
Y2	0.022313967	205752	0.070434309	0.040475913	0.50662448
Y3	0.014267506	736811	0.004001026	0.082238186	0.543242433
Y4	0.039013972	989139	0.001116122	0.051536741	0.547061636
Y5	0.047463699	811926	0.001019797	0.034744546	0.592098295
Y6	0.499126321	4143	0.639150374	0.020516534	0.09534154
083	0.002700039	424059	0.003442917	0.03267234	0.613619803
085	0.002132024	1683305	0.001428143	0.067047861	0.747913777
086	0.001808648	1998054	0.004591968	0.041100991	0.740861859
088	0.001610477	2370305	0.004237429	0.033140039	0.714821088
089	0.125489899	6824623	0.002909758	0.083338523	0.612114252
093	0.00148497	11315945	0.002669154	0.019454495	0.787331504
094	0.031525493	1801276	0.002588165	0.084962549	0.576490221
095	0.002028582	9966114	0.002931634	0.025248959	0.789525988
096	0.003554301	2728927	0.001659993	0.023487986	0.643969956
102	0.012532408	7315003	0.002668625	0.051052611	0.625627631
113	0.267526409	1265738	0.006691748	0.129785943	0.412912467
121	0.004616899	9471265	0.000884676	0.027631578	0.569689688
122	0.001011443	3642880	0.00183893	0.036619927	0.71837612
123	0.001337782	727333	0.001729607	0.037014682	0.570275238
125	0.450101208	70914	0.013213188	0.041092027	0.358659785
128	0.045390128	2512033	0.008107776	0.039144788	0.66629061
145	0.211780262	837066	0.160462855	0.097835774	0.180954668
146	0.063735532	130696	0.003687948	0.088923915	0.742624105
147	0.153438777	271297	0.00517514	0.053671806	0.519342271
148	0.038597722	10604568	0.002198298	0.031893237	0.683701684
149	0.13523961	4498775	0.006449533	0.043098177	0.496394241
150	0.169917709	3665622	0.006883688	0.072267681	0.504646415
154	0.240664459	221335	0.061987485	0.063406149	0.223615786
167	0.0012147	9913749	0.002558467	0.027010468	0.785593926
168	0.00142026	1155851	0.002136088	0.048642948	0.728684753
169	0.060191017	107793	0.012366295	0.076405704	0.342304231
170	0.001586528	6999442	0.0019743	0.07589005	0.740438309
174	0.191731766	2646320	0.005560552	0.598146483	0.185157124
177	0.502751222	343623	0.024652017	0.108156322	0.332268213
180	0.002114844	1001601	0.000532148	0.117777438	0.601394168
181	0.033779088	578560	0.002986726	0.122637237	0.586087873
182	0.012390545	624906	0.002301146	0.198493853	0.453301777
183	0.006847255	2486586	0.001180735	0.054931541	0.598985115
184	0.000957281	6152701	0.001215726	0.026229619	0.544070807
185	0.001434217	10373292	0.001358103	0.020667499	0.795980582
186	0.146166134	343916	0.006681864	0.062704265	0.470010119
189	0.002723003	1873722	0.007107244	0.062748369	0.729391553
192	0.005586001	1901991	0.002362261	0.05567692	0.735655952
193	0.003591491	13047108	0.001417939	0.026802951	0.583769522
194	0.003134144	15221187	0.002546451	0.015259585	0.746436398
195	0.00363156	1308850	0.009598503	0.049409787	0.60937235
202	0.010700493	2811087	0.000547475	0.046687989	0.364012925
212	0.029842354	1346975	0.002861226	0.035996214	0.438361514
215	0.619402321	787932	0.005943406	0.035761969	0.235500018
217	0.005447566	4264261	0.000804125	0.046439231	0.50423438
230	0.001840298	4858865	0.040871891	0.028553376	0.573844921
234	0.583608584	181154	0.003930358	0.026099341	0.239602769
235	0.245952364	19768	0.00252934	0.02028531	0.222177256
241	0.18030581	1320663	0.003024239	0.029583626	0.326096817
302	0.109655167	3954842	0.001874411	0.041846172	0.338902287
303	0.002610408	13874159	0.00169113	0.020673109	0.688516544

Table S3, related to Figure 3

Metagenomics Co-assembly Statistics

Sample	n Samples	TotalReads	n Contigs	n Bases	min ContigLength	max ContigLength	ave ContigLength	N50	UniqMapReads	MultiMapReads	UniqMapRatio	MultiMapRatio
PharyngealSwab	69	199359275	273434	178540833	200	86844	652	800	67454202	44781114	0.338354972	0.224625185
Sputum	59	146500324	266932	158442390	200	43382	593	689	50974274	33141029	0.347946493	0.226218128
Faeces	33	49540839	58836	34845039	200	32392	592	699	19330467	3547114	0.390192564	0.071599797

Table S4, related to Figure 3

Mapping Stats of Metagenomics Co-Assembly

SampleType	SampleID	TotalReads	UniqMapReads	MultiMapReads	ave_Coverage	perc_CoveredScaffolds (%)	perc_CoveredBases (%)	UniqMapRatio	MultiMapRatio
PharyngealSwab	008	4366334	1783005	1363718	3.084	12.27	5.93	0.408352865	0.312325626
PharyngealSwab	009	1513285	706582	16461	0.659	5.76	2.47	0.466919318	0.01087766
PharyngealSwab	010	292498	94369	18209	0.112	7.27	2.46	0.322631266	0.062253417
PharyngealSwab	011	4936618	1894918	1068398	3.053	49.12	27.87	0.383849429	0.216423065
PharyngealSwab	012	2872201	1229347	709968	1.781	20.6	9.95	0.428015658	0.247186043
PharyngealSwab	013	995977	398471	285583	0.689	16.53	7.41	0.400080524	0.286736541
PharyngealSwab	014	1012355	411937	30936	0.398	6.5	3.18	0.406909632	0.03055845
PharyngealSwab	015	132205	29039	2209	0.03	2.83	0.98	0.219651299	0.016708899
PharyngealSwab	016	919591	152510	14876	0.203	5.29	2.28	0.165845468	0.016176757
PharyngealSwab	017	3373926	2333580	71904	2.129	7.95	5.09	0.69165121	0.021311671
PharyngealSwab	025	3314711	1147369	751685	2.009	42.72	22.93	0.346144506	0.226772409
PharyngealSwab	037	184865	35204	8377	0.043	3.28	1.02	0.190430855	0.045314148
PharyngealSwab	038	37991	13528	1118	0.014	1.45	0.42	0.356084336	0.029428022
PharyngealSwab	039	263095	53347	8918	0.057	4.12	1.34	0.202767061	0.033896501
PharyngealSwab	040	286824	59756	10008	0.065	4.57	1.53	0.20833682	0.034892478
PharyngealSwab	041	263758	69386	30269	0.093	8.6	2.51	0.263066902	0.1147605
PharyngealSwab	042	71365	15896	2267	0.022	2.26	0.69	0.222742241	0.031766272
PharyngealSwab	043	1419615	547757	310884	0.826	28.94	13.89	0.38584898	0.218991769
PharyngealSwab	044	1579626	595784	351913	0.952	37.62	18.8	0.37716776	0.222782481
PharyngealSwab	045	1952079	712994	510938	1.192	36.65	18.98	0.365248538	0.261740432
PharyngealSwab	046	2672772	967220	731743	1.674	38.65	19.89	0.361878978	0.273776813
PharyngealSwab	047	1278475	427931	264408	0.691	28.04	13.33	0.334719881	0.206815151
PharyngealSwab	048	2318133	965852	690827	1.576	14.18	7.53	0.416650813	0.29801008
PharyngealSwab	049	2696632	987688	664946	1.62	43.07	22.99	0.366267255	0.246583887
PharyngealSwab	050	3220837	1128779	1018225	1.97	32.8	15.7	0.350461386	0.316136768
PharyngealSwab	051	4544027	1514724	1755654	2.984	22.84	9.44	0.33334397	0.386365222
PharyngealSwab	052	3064879	1079983	833390	1.748	37.8	17.76	0.35237378	0.271916118
PharyngealSwab	053	2732709	1086150	586272	1.529	40.51	20.17	0.397462738	0.21453876
PharyngealSwab	054	3265342	1392790	745293	1.959	35.92	18.71	0.426537251	0.228243473
PharyngealSwab	055	3837017	1481664	1037941	2.307	41.85	22.06	0.38614997	0.270507272
PharyngealSwab	056	3581067	1744283	644864	2.177	37.1	18.75	0.487084715	0.180075938
PharyngealSwab	057	3346679	1183773	973519	2.025	34.33	17.5	0.353715728	0.290891059
PharyngealSwab	072	1036507	290346	285954	0.552	22.4	9.19	0.280119671	0.275882363
PharyngealSwab	080	65519	13976	2993	0.017	1.83	0.6	0.213312169	0.045681405
PharyngealSwab	073	1503578	506083	320679	0.779	16.79	8.06	0.336585797	0.213277263
PharyngealSwab	074	1496732	435717	548667	0.899	24.6	10.15	0.291112237	0.366576648
PharyngealSwab	075	849815	308377	194762	0.487	22.15	10.18	0.362875449	0.229181645
PharyngealSwab	078	1170985	697111	52100	0.665	6.59	3.68	0.595320179	0.044492457
PharyngealSwab	089	6824623	3475088	240955	3.47	6.88	5.89	0.50919853	0.035306712
PharyngealSwab	093	11315945	3723595	4628299	7.239	38.03	20.98	0.329057361	0.409006848
PharyngealSwab	094	1801276	504405	375116	0.748	20.11	8.13	0.280026492	0.208250152
PharyngealSwab	095	9966114	4341011	3493745	6.604	30.68	17.76	0.435577097	0.350562416
PharyngealSwab	096	2728927	1009847	881358	1.65	25.08	13.03	0.370052772	0.322968698
PharyngealSwab	113	1265738	257605	85145	0.276	4.04	1.73	0.203521582	0.067269056
PharyngealSwab	125	70914	14436	765	0.013	1.71	0.49	0.203570522	0.010787715
PharyngealSwab	148	10604568	3733956	3290346	6.143	50.06	29.35	0.352108261	0.310276288
PharyngealSwab	154	221335	53814	2635	0.056	2.81	0.91	0.243133711	0.011905031
PharyngealSwab	167	9913749	3650945	3243359	6.44	42.72	24.44	0.368270873	0.327157668
PharyngealSwab	169	107793	46761	5015	0.045	2.31	1.05	0.433803679	0.046524357
PharyngealSwab	181	578560	198491	13459	0.211	4.77	3.33	0.343077641	0.023262929
PharyngealSwab	182	624906	76894	15046	0.106	2.26	0.68	0.123048907	0.024077221
PharyngealSwab	184	6152701	2514518	1565590	3.73	44.07	25.25	0.408685226	0.254455726
PharyngealSwab	186	343916	101216	44941	0.123	7.37	2.61	0.294304423	0.130674351
PharyngealSwab	193	13047108	4609877	4271399	7.492	47.5	27.86	0.35332558	0.327382819
PharyngealSwab	215	787932	230204	43440	0.201	6.81	4.21	0.292162268	0.055131661
PharyngealSwab	234	181154	69388	13051	0.055	4.23	2.02	0.38303322	0.072043676
PharyngealSwab	235	19768	4724	546	0.005	1.13	0.29	0.238972076	0.027620397
PharyngealSwab	241	1320663	575223	98715	0.523	10.11	7	0.435556232	0.074746548
PharyngealSwab	302	3954842	2470321	175206	2.172	11.38	9	0.624632033	0.044301643
PharyngealSwab	303	13874159	4517530	5327147	8.503	45.94	26.32	0.325607484	0.383961795
PharyngealSwab	NC-P1	2661398	761	398	0.275	13.4	6.75	0.00028594	0.000149545
PharyngealSwab	NC-P2	2330412	170	446	0.228	8.77	5.92	7.29E-05	0.000191382
PharyngealSwab	NC-P3	2161861	624	355	0.262	12.93	6.96	0.00028864	0.00016421
PharyngealSwab	NC-P4	4137917	3656	445	0.669	23.82	10.31	0.000883536	0.000107542
PharyngealSwab	NC-P5	2511179	798	280	0.312	17.32	8.26	0.000317779	0.000111501
PharyngealSwab	NC-P6	2830752	1927	327	0.397	22.62	9.81	0.000680738	0.000115517
PharyngealSwab	NC-P7	2242409	309	358	0.223	12.73	6.35	0.000137798	0.00015965
PharyngealSwab	NC-P8	3485836	3461	244	0.582	22.23	9.43	0.000992875	7.00E-05
Sputum	002	494070	140944	75071	0.227	13.21	5.32	0.285271318	0.151944057
Sputum	003	8023735	3356679	1723799	5.828	35.7	24.18	0.418343702	0.214837479
Sputum	004	787922	281106	203174	0.571	17.49	8.52	0.356768817	0.25786055
Sputum	005	2361502	931296	528019	1.64	17.29	9.45	0.394365959	0.22359456
Sputum	006	4297298	1795018	912860	3.062	28.97	17.61	0.417708523	0.212426506
Sputum	007	3490992	1254702	1119251	2.656	7.31	3.83	0.359411308	0.320611162
Sputum	023	2612059	714059	635651	1.643	33.87	17.88	0.273370165	0.243352466
Sputum	026	1924866	589166	504226	1.299	31.7	16.72	0.306081566	0.261953819
Sputum	028	1187883	200546	108134	0.354	1.76	0.53	0.168826391	0.091030851
Sputum	029	1480193	461446	297838	0.922	30.59	15.56	0.311747184	0.201215652
Sputum	030	1834920	321218	180026	0.675	19.28	7.72	0.175058313	0.098111089
Sputum	032	338880	28743	10370	0.047	1.9	0.53	0.084817635	0.030600803
Sputum	034	835001	316709	247425	0.598	21.54	10.63	0.379291761	0.29631701
Sputum	035	619841	27798	6465	0.036	2.71	0.74	0.044846985	0.010430094
Sputum	036	5513710	1593407	1807549	3.766	22.93	13.15	0.28898999	0.327828087
Sputum	060	2339720	933470	601315	1.412	38.56	19.19	0.398966543	0.257002975
Sputum	061	939891	330917	239703	0.591	20.02	10.23	0.352080188	0.255032764
Sputum	062	1399198	526859	389053	0.953	28.34	14.13	0.376543563	0.278054285
Sputum	063	1141437	359801	176564	0.614	31.08	14	0.315217572	0.154685716

Sputum	064	1121648	335683	232574	0.658	27.69	13.21	0.2992766	0.207350256
Sputum	065	1167864	330975	412004	0.787	13.71	6.13	0.283402006	0.352784228
Sputum	066	2729856	1375266	277463	1.729	19.76	14.57	0.503787013	0.10164016
Sputum	067	1612208	545788	367380	1.027	31.21	15.62	0.338534482	0.227873823
Sputum	068	2602295	1134835	613140	1.778	42.34	22.58	0.436090067	0.235615101
Sputum	069	3831180	2491631	454648	2.924	37	20.08	0.650356026	0.118670488
Sputum	070	1187881	387497	262701	0.737	28.28	14.08	0.326208602	0.22115094
Sputum	071	2455028	818077	835395	1.677	24.94	13.97	0.33322512	0.340279215
Sputum	079	1076912	338675	233563	0.66	28.09	13.56	0.314487163	0.216882159
Sputum	Y1	73250	27204	4874	0.035	3.9	1.15	0.371385666	0.066539249
Sputum	Y2	205752	81647	18825	0.104	8.73	3.31	0.396822388	0.091493643
Sputum	Y3	736811	235949	84960	0.345	13.81	5.74	0.320230018	0.115307725
Sputum	Y4	989139	472515	73394	0.555	12.51	6.06	0.477703336	0.074199885
Sputum	Y5	811926	356758	90023	0.482	14.48	7.07	0.439397186	0.110875868
Sputum	Y6	4143	2371	28	0.002	0.07	0.04	0.572290611	0.006758388
Sputum	083	424059	122192	96275	0.209	17.21	5.82	0.288148583	0.227032088
Sputum	085	1683305	692954	332272	1.121	7.82	4.2	0.411662771	0.197392629
Sputum	086	1998054	654106	581458	1.293	19.38	10.09	0.327371533	0.291012155
Sputum	088	2370305	799537	638923	1.669	28.16	15.63	0.337313974	0.269553074
Sputum	102	7315003	2357669	1791322	4.195	39.03	23.94	0.322305951	0.24488329
Sputum	121	9471265	4566208	1657183	6.127	49.56	33.47	0.482111735	0.174969553
Sputum	128	2512033	1034440	732061	1.632	21.56	11.45	0.411793953	0.291421729
Sputum	147	271297	93885	62114	0.144	10.22	3.33	0.346059853	0.228952034
Sputum	150	3665622	1199211	757308	1.759	31.83	14.34	0.327150754	0.206597407
Sputum	168	1155851	366957	157751	0.651	23.35	12.04	0.317477772	0.136480394
Sputum	170	6999442	2409604	2003490	4.79	29.39	18.78	0.344256585	0.286235674
Sputum	174	2646320	1707227	78442	1.816	8.78	5.54	0.645132486	0.029641918
Sputum	177	343623	76152	9862	0.106	4.4	1.58	0.221614968	0.028700058
Sputum	183	2486586	858556	351550	1.293	19.53	10.62	0.345275008	0.141378581
Sputum	185	10373292	3598872	3603825	7.792	41.64	27.68	0.346936344	0.34741382
Sputum	194	15221187	5907981	5822870	11.38	37.91	25.07	0.388141937	0.382550323
Sputum	217	4264261	1406016	733694	2.036	42.67	20.77	0.329720906	0.172056542
Sputum	NC-T1	638736	114	100	0.051	8.33	2.64	0.000178477	0.000156559
Sputum	NC-T2	1565794	3548	409	0.294	20.71	7.85	0.002265943	0.000261209
Sputum	NC-T3	1649872	4471	459	0.322	25.62	8.52	0.002709907	0.000278203
Sputum	NC-T4	765516	1161	113	0.108	15.6	4.65	0.001516624	0.000147613
Sputum	NC-T5	894809	832	157	0.086	15	3.84	0.000929807	0.000175456
Sputum	NC-T6	1679014	3487	510	0.337	26.34	9.17	0.002076814	0.00030375
Sputum	NC-T7	2798816	9028	1199	0.595	31.11	11.29	0.00322565	0.000428395
Sputum	NC-T8	1077251	1311	217	0.141	16.55	5.51	0.001216987	0.000201439
Stool	018	273574	31308	4239	0.219	9.42	4.34	0.1144407	0.015494894
Stool	019	1126424	498848	200522	3.608	20.41	15.96	0.442859882	0.178016449
Stool	020	226583	28809	1543	0.202	10.71	4.34	0.127145461	0.006809867
Stool	021	384049	84715	16479	0.548	15.86	8.68	0.220583832	0.042908587
Stool	022	467215	146894	42618	1.048	14.92	11.15	0.314403433	0.091217106
Stool	024	14482545	9653002	1290951	56.071	1.88	0.66	0.666526636	0.089138408
Stool	027	309059	49144	8974	0.347	14.78	7.74	0.15901171	0.029036527
Stool	031	1035181	317176	41209	1.964	32.85	25.21	0.306396659	0.039808497
Stool	058	638484	177979	10781	0.862	7.94	2.79	0.278752482	0.01688531
Stool	059	270289	48843	6183	0.207	3.8	1.1	0.180706577	0.022875515
Stool	076	104031	4211	2274	0.052	1.07	0.25	0.040478319	0.021858869
Stool	077	664662	113871	36679	0.792	18.79	11.39	0.171321664	0.05518444
Stool	081	107785	11233	4263	0.11	2.82	1.2	0.104216728	0.039550958
Stool	122	3642880	1711348	290582	8.92	33.04	29.73	0.469778856	0.079767107
Stool	123	727333	257616	73089	1.471	20.68	15	0.354192646	0.100489047
Stool	145	837066	449395	46805	2.163	4.04	2.09	0.536869255	0.055915543
Stool	146	130696	13406	8797	0.081	2.71	0.86	0.102573912	0.067308869
Stool	149	4498775	1977149	341414	9.221	33.89	27.56	0.439486082	0.075890437
Stool	180	1001601	119940	37646	0.815	7.17	2.1	0.119748283	0.037585825
Stool	189	1873722	220255	167417	1.312	11.88	5.86	0.117549455	0.089349968
Stool	192	1901991	443243	158658	2.373	22.87	15.51	0.233041586	0.083416799
Stool	195	1308850	192020	108721	1.105	15.14	8.04	0.146708943	0.08306605
Stool	202	2811087	238038	118490	1.212	17.42	6.47	0.084678276	0.042150954
Stool	212	1346975	212864	31493	1.005	20.97	10.81	0.158031144	0.023380538
Stool	230	4858865	2325941	497281	12.012	44.62	35.58	0.478700478	0.102345095
Stool	NC-B1	274121	329	0	0.121	19.35	4.5	0.0012002	0
Stool	NC-B2	494231	498	3	0.373	27.44	11.57	0.001007626	6.07E-06
Stool	NC-B3	1395918	499	0	0.905	52.17	26.62	0.000357471	0
Stool	NC-B4	243923	397	1	0.174	24.49	8.5	0.001627563	4.10E-06
Stool	NC-B5	1204783	297	1	1.998	13.38	3.07	0.000246517	8.30E-07
Stool	NC-B6	347233	433	0	0.22	23.68	7.26	0.001247001	0
Stool	NC-B7	291237	335	0	0.074	9.35	1.9	0.001150266	0
Stool	NC-B8	259671	431	1	0.179	21.16	7.22	0.001659793	3.85E-06