Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Evaluation and Validation of a Reproducible mNGS Workflow for Gut Microbiome: Optimizing Preservation and Automated Extraction

Yingying Zheng
Nankai University
Hailing Wanghu
Guangdong Zijian Chuanghe Pharmaceutical Biotechnology Co., Ltd
Zhongkui Xia
Guangdong Zijian Chuanghe Pharmaceutical Biotechnology Co., Ltd
Yanjuan Huang
Guangdong Zijian Chuanghe Pharmaceutical Biotechnology Co., Ltd
Xiaodi Luo
Guangdong Zijian Chuanghe Pharmaceutical Biotechnology Co., Ltd
Xianjin Wang
Guangdong Zijian Chuanghe Pharmaceutical Biotechnology Co., Ltd
Lingling Qiu
Guangdong Zijian Chuanghe Pharmaceutical Biotechnology Co., Ltd
Nengquan Chen
Guangdong Zijian Chuanghe Pharmaceutical Biotechnology Co., Ltd
Xiao Han
Fuzhou University
Jinglan Lai
Mengchao Hepatobiliary Hospital of Fujian Medical University
Zihe Rao
linc0612@sina.cn

Nankai University

Article

Keywords: Shotgun metagenomics, Fecal microbiome, Inhibitor removal

Posted Date: May 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9206006/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on October 11th, 2026.
See the published version at https://doi.org/10.1038/s41598-026-70557-8.

Evaluation and Validation of a Reproducible mNGS Workflow for Gut Microbiome: Optimizing Preservation and Automated Extraction

Yingying Zheng[1 2+], Hailing Wanghu[2+], Zhongkui Xia[2], Yanjuan Huang[2], Xiaodi Luo[2], Xianjin Wang[2], Lingling Qiu[2], Nengquan Chen[2], Xiao Han[3], Jinglan Lai[4], Zihe Rao[1],*

[1]College of Life Sciences, Nankai University, Tianjin 300071, China.

[2]Guangdong Zijian Chuanghe Pharmaceutical Biotechnology Co., Ltd., Guangzhou, Guangdong 510535, China.

[3]College of Biological Science and Engineering, Fuzhou University, Fuzhou, Fujian 350108, China.

[4]Department of Infectious Diseases, Mengchao Hepatobiliary Hospital of Fujian Medical University, Fuzhou, Fujian 350025, China.

*Correspondence to: Zihe Rao, email: linc0612@sina.cn

Note: The authors wish it to be known that the first two authors be regarded as the joint First Authors.

Abstract

Technical variability in pre-analytical and analytical steps significantly hinders the reproducibility of gut metagenomic next-generation sequencing (mNGS). To address this, we developed and validated a practical mNGS workflow. First, by evaluating three commercial preservation buffers at 37°C, we identified Buffer C as the most effective for maintaining microbiome integrity under simulated transport stress. Second, we introduced a cost-effective, self-prepared inhibitor removal buffer (Self-IB) for automated extraction, which yielded species abundances highly concordant with the manual reference method ($R^2 > 0.998$). Finally, the optimized workflow demonstrated high intra- and inter-batch reproducibility, with most species-level coefficients of variation below 10%. This integrated, end-to-end pipeline (Buffer C +

Self-IB + automated workflow) facilitates high data fidelity and robustness, offering a practical strategy for industrial and clinical applications requiring longitudinal consistency, such as dynamic monitoring following fecal microbiota transplantation (FMT).

Keywords Shotgun metagenomics，Fecal microbiome，Inhibitor removal

Introduction

The gut microbiota, comprising a complex community of microorganisms, plays a fundamental role in host physiology, immune regulation, and disease susceptibility [1–2]. Accurate characterization of its composition and function is therefore critical for advancing microbiome research and its clinical translation [3]. Due to the historical challenges in cultivating the majority of gut microbes – as evidenced by the fact that a significant proportion of species lack cultured representatives [4] culture-independent molecular techniques, particularly NGS, have become indispensable.

Two principal NGS-based approaches are widely employed: 16S rRNA gene amplicon sequencing and mNGS. While 16S sequencing is cost-effective for broad compositional surveys, its limited taxonomic resolution and inability to access functional genetic content restrict its utility for strain-level tracking or functional insight into specific pathogens [5–7]. In contrast, mNGS provides high-resolution taxonomic and functional profiling, making it the preferred method for in-depth analysis [8].

However, the reproducibility and accuracy of mNGS in gut microbiome studies are compromised by significant pre-analytical and analytical variabilities. Two key bottlenecks hinder its reliable application, especially in longitudinal or high-throughput settings: (i) shifts in microbial community structure during sample storage and transport, and (ii) suboptimal inhibitor removal and DNA recovery efficiency in automated nucleic acid extraction workflows. While previous studies have often addressed these challenges in isolation—evaluating either preservation buffers or extraction methods—research that systematically integrates both components and validates the long-term stability of the complete workflow remains relatively scarce. This gap limits the establishment of a standardized, end-to-end pipeline for applications demanding exceptional consistency, such as the dynamic monitoring of FMT.

To address this, we developed and evaluated a practical mNGS workflow, focusing on the

optimization of preservation and automated extraction steps.. We first evaluated three commercial preservation buffers under simulated transport stress (37°C) to identify the optimal solution for maintaining community integrity. We then optimized an automated extraction protocol by developing and incorporating a cost-effective, Self-IB, benchmarking its performance against a manual reference. Finally, we assessed the intra- and inter-batch reproducibility of the complete pipeline. The optimized workflow—integrating the selected preservation buffer, the Self-IB pretreatment, and automated processing—demonstrated high fidelity and robustness, providing a reliable framework for industrial-scale and clinical gut microbiome studies requiring stringent reproducibility.

Result

Optimization of Automated Nucleic Acid Extraction Buffers for Robust Fecal mNGS Profiling

As the foundational steps of the workflow, sample collection and nucleic acid extraction are critical determinants of data quality in gut microbiota profiling. While previous studies have optimized manual fecal DNA extraction protocols [9-10], the low throughput of manual methods remains a bottleneck for industrial-scale applications. Automated extraction systems, capable of processing 96 samples simultaneously, offer a high-throughput solution. However, they require careful optimization to match the performance of manual “gold-standard” methods.

To this end, we evaluated an automated extraction workflow against manual extraction (the reference method) by testing four different sample pretreatment conditions: (1) no pretreatment, (2) PBS pretreatment, (3) pretreatment with the commercial QIAGEN inhibitor removal buffer (QIA-IB), and (4) pretreatment with Self-IB. Each condition was tested with four technical replicates.

Regarding nucleic acid yield and purity, the untreated group exhibited an abnormally high DNA concentration (480 ± 30 ng/μL) but significantly lower purity (A260/A280 = 1.7 ± 0.03; A260/A230 = 0.48 ± 0.02) compared to the manual reference (A260/A280 = 2.1 ± 0.03; A260/A230 = 2.15 ± 0.05) (Fig. 2A-C). The characteristically high concentration coupled with low purity strongly indicates the co-extraction of PCR inhibitors, which are known to interfere with spectrophotometric measurements, leading to an overestimation of DNA yield [10].PBS

pretreatment enhanced purity metrics (A260/A280 = 1.82 ± 0.07; A260/A230 = 0.81 ± 0.07) but reduced DNA yield by approximately 60% (Fig. 2A-C). In contrast, both the QIA-IB and Self-IB pretreatments substantially improved both yield and purity, with QIA-IB showing the closest performance to the manual standard (Fig. 2A-C). The primary function of inhibitor removal buffers is to dissociate these inhibitory substances from DNA, thereby facilitating accurate yield quantification and ensuring downstream compatibility.

We next assessed the biological fidelity of the extracted DNA by comparing the microbial community profiles. The relative abundance of the top 10 species detected by the automated workflow without pretreatment diverged most significantly from the manual reference, followed by the PBS-pretreated group (Fig. 2D). The two IB-treated groups, however, showed community profiles highly similar to the manual extraction. A detailed correlation analysis of the top 30 species, performed by linear regression, showed that both Self-IB (slope = 0.996) and QIA-IB (slope = 1.000) yielded relative abundances nearly identical to the manual method, resulting in a coefficient of determination (R^2) > 0.998 (Fig. 2EF). The 0.4% higher slope for QIA-IB indicates a marginally closer match, although this difference is biologically negligible. Therefore, given its comparable performance and significantly lower cost (as detailed in the Methods section), Self-IB presents a highly cost-effective alternative for large-scale applications.

In summary, incorporating an inhibitor removal buffer (IB) in the pretreatment step is essential for accurate and reliable automated nucleic acid extraction from fecal samples. While the commercial QIA-IB showed optimal performance, the self-prepared Self-IB serves as a highly effective and cost-efficient alternative, making the optimized automated workflow a robust solution for high-throughput gut microbiome studies.

Performance Evaluation of Fecal Preservation Buffers under Simulated Transport Conditions (37°C)

Studies have shown that without an effective preservation buffer, fecal samples stored at room temperature for 24 hours can exhibit up to an 80% decrease in similarity to fresh samples; even at -20°C, the microbial community structure begins to change after three days [11]. In the clinical-scale application of gut microbiome analysis, most samples require transportation under variable conditions. Given the consistently warm climate in southern China, the absence of a

reliable preservation buffer would severely compromise the biological authenticity of sequencing results.

To address this, we evaluated the performance of three mainstream commercial preservation buffers (A, B, and C) under simulated transport stress. Aliquots of the same fecal sample were preserved with each buffer and incubated at 37°C for 0, 1, 3, 5, 7, and 9 days, followed by standardized automated nucleic acid extraction and metagenomic sequencing.

At the nucleic acid level, all buffers showed time-dependent DNA degradation (Fig. 3A), but with distinct patterns. Preservation Buffer A yielded the highest initial DNA concentration, but underwent the most rapid degradation. Buffer B maintained a relatively stable concentration profile, while Buffer C, after an initial decrease, demonstrated the most stable concentration over the remaining period. In terms of purity, Buffer A's A260/230 ratio fell below 1.0 after Day 3 (minimum 0.53), indicating significant accumulation of organic impurities over time (Fig. 3B). In contrast, both Buffer B and C maintained stable A260/280 and A260/230 ratios throughout the incubation, with Buffer C showing the highest and most stable purity metrics (Fig. 3BC).

Analysis of microbial diversity revealed that the Shannon index fluctuated with storage time for all buffers, showing a general upward trend (Fig. 3D). Based on the coefficient of variation (CV) of the Shannon index, Preservation Buffer B exhibited the largest fluctuation (CV = 14.13%), followed by Buffer A. Preservation Buffer C demonstrated the highest stability (CV = 2.54%).

Changes in the relative abundance of the top 10 bacterial species mirrored the diversity trends but highlighted a key finding concerning *Faecalibacterium prausnitzii* (Fig. 3E). In Preservation Buffer C, its relative abundance remained highly stable (Day 0 = 8.9%, Day 9 = 10.1%). In contrast, Buffer A showed a marked increase that plateaued after Day 5 (Day 0 = 4.1%, Day 5 = 14.9%, Day 9 = 14.5%), coinciding with a stabilization of its Shannon index. *Faecalibacterium prausnitzii* is an obligate anaerobe whose DNA is sensitive to preservation conditions [12]. The substantial increase in its abundance in Buffer A suggests that the reducing agents in this buffer may maintain an anaerobic microenvironment that permits continued microbial activity—or even proliferation—of this taxon during storage. Conversely, the stability observed in Buffer C provides direct evidence supporting its proposed mechanism of action: the potent lytic agents in Buffer C likely achieve rapid "metabolic inactivation" upon contact, effectively freezing the microbial community and preventing such compositional shifts. This clear divergence underscores the

critical difference between preservation strategies that merely inhibit decay and those that halt metabolism entirely.

Taken together, Preservation Buffer C demonstrated the most stable performance in maintaining both nucleic acid integrity and microbial community structure under simulated high-temperature transport and is therefore recommended for use in demanding logistics environments.

Robustness Assessment of Shannon Diversity Index in mNGS-Based Microbiome Profiling

While metagenomic next-generation sequencing (mNGS) has matured as a standard tool in pathogen detection[13-14], its application in gut microbiota research—particularly for longitudinal studies like FMT monitoring—demands exceptional technical reproducibility to ensure reliable interpretation of microbial dynamics. Such applications require highly stable and reproducible results across batches and operators to avoid misjudgment of therapeutic efficacy [15].

To systematically evaluate the robustness of our standardized mNGS workflow, we conducted intra- and inter-operator reproducibility assessments. For the intra-operator (intra-batch) analysis, a single operator (Operator A) processed five fecal samples (S1–S5) with three technical replicates per sample over three consecutive days. Analysis of the top 10 species demonstrated highly consistent relative abundance profiles across all replicates for samples S1–S4 (Fig. 4A-D). Sample S5, which was characterized by a notably higher initial relative abundance of the oxygen-sensitive species Roseburia faecis, exhibited greater variability, particularly on day 2 (Fig. 4E). Quantification of the CV for species- and genus-level relative abundances confirmed excellent stability for S1–S4 (mean CV<10%), while S5 showed a higher mean CV (>13%) (Table.1). This fluctuation was primarily attributable to a significant time-dependent decrease in the relative abundance of the oxygen-sensitive species *Roseburia faecis* (mean CV=22.17%) in S5, suggesting that even minimal oxygen exposure during handling can affect susceptible anaerobes.

For the inter-operator (inter-batch) analysis, a second operator (Operator B) independently processed the same five samples in parallel on the third day. Comparison of the top 30 species' relative abundances between operators (150 data points in total) revealed near-perfect agreement, with a linear regression slope of 1.01 and an R^2 value of 0.9915 (Fig.4F). Furthermore, the Shannon diversity index of all five samples remained stable throughout the reproducibility

assessments, with all CV values below 10% in both intra- and inter-operator tests (Table 2). This confirms that the overall community diversity was robustly preserved despite the species-level fluctuations observed in S5.

Together, these results indicate that, our optimized mNGS workflow demonstrates high reproducibility. Most samples exhibited a mean species-level abundance CV < 10% in intra-operator tests, and all samples showed CV < 10% in inter-operator comparisons. The strong correlation between operators (R^2 > 0.99) confirms that technician-induced variability is negligible.

Limitations and future directions

While this study establishes a robust DNA-based workflow, several limitations should be acknowledged. First, our validation primarily focused on bacterial community analysis; consequently, the assessment depth for fungal and viral communities remains limited. Second, all optimizations were conducted on a specific sequencing platform (Salus Pro) with dedicated reagents, and thus the generalizability of this standardized workflow to other platforms requires further verification. Finally, as the protocol is DNA-based, it does not encompass the detection of RNA viruses (e.g., norovirus, rotavirus), which are relevant for a comprehensive gut pathogen profile.

Furthermore, the scope of our optimization was defined by comparing a limited set of three commercial preservation buffers and one automated extraction platform. While this approach identified a performant and reproducible protocol under the tested conditions, the generalizability of these findings to all available buffers, extraction systems, or sample types (e.g., from diseased populations) requires further validation.

Future work will focus on addressing these limitations. We plan to: 1. validate the workflow's performance on fungal and viral taxa using defined mock communities or clinical samples; 2. test its transferability to other mainstream sequencing platforms; and 3.develop an integrated RNA detection module to enable complementary metatranscriptomic analysis. Furthermore, multi-center validation will be essential to confirm the workflow's robustness and reproducibility across different laboratory settings.

Discussion

The gut microbiota, a complex ecosystem vital to host physiology, interacts with systemic organs via multi-organ axes, underpinning its role in health and disease [16-18]. Consequently, precise characterization of its composition is essential, particularly for interventions like FMT, where evaluating efficacy relies on accurate, reproducible metagenomic tracking.

This study evaluates and integrates key components of a practical mNGS workflow for gut microbiome analysis, addressing major sources of pre-analytical variability sources: sample preservation and automated nucleic acid extraction, while demonstrating high technical reproducibility (Fig.5).

Our findings elucidate a critical principle for sample preservation: rapid metabolic inactivation is superior to merely maintaining an anoxic environment. Buffers A and B, likely reliant on reducing agents (e.g., dithiothreitol), preserve DNA integrity initially but sustain an anaerobic microenvironment. Under prolonged thermal stress (37°C), this permits continued proliferation of obligate anaerobes like *Faecalibacterium prausnitzii*, thereby distorting community structure (Fig. 3E) [19]. In stark contrast, Buffer C's formulation, which includes potent lytic agents, achieves superior stability by irreversibly inactivating microbial enzymes upon contact, effectively “freezing” the in vivo community state. This “metabolic inactivation” mechanism is key to accurate preservation and should be a primary consideration in future preservative design.

In parallel to the preservation buffer optimization, our evaluation of the automated nucleic acid extraction workflow highlights the importance of effective inhibitor removal for maintaining microbial representation fidelity. The pretreatment conditions comparison demonstrated that both the commercial QIA-IB and self-prepared Self-IB buffers achieved species abundances highly consistent with the manual reference method, with correlation coefficients exceeding 0.998. Although QIA-IB exhibited a marginally superior linear regression slope (1.000 versus 0.996 for Self-IB), this minor difference is biologically insignificant. The comparable performance of Self-IB, coupled with its cost-effectiveness, positions it as a practical alternative for large-scale applications, where economic considerations are paramount. This optimization step ensures that the automated extraction process yields community profiles highly comparable to the manual reference method, thereby reducing differential bias.

The high intra- and inter-batch reproducibility (CV<10% for most samples; R^2 >0.99) confirms the robustness of our optimized workflow against technical variations. However, the higher variability in sample S5, primarily driven by fluctuations in the oxygen-sensitive species Roseburia faecis, offers a critical insight. While minimal oxygen exposure during handling is a plausible contributor, this observation also highlights the significant impact of intrinsic sample properties on technical stability. Factors such as the initial microbial composition, endogenous enzyme activity, and concentration of PCR inhibitors can vary significantly between individual fecal samples, potentially influencing their resilience even under standardized processing. Therefore, the variability in S5 may not be solely a technical artifact but also a reflection of its unique biogeochemical makeup. Acknowledging that sample-specific factors contribute to reproducibility provides a more realistic assessment of the workflow's capabilities and limitations in handling the inherent heterogeneity of human samples, pointing toward future optimization for personalized applications.

In summary, the integration of a metabolically inactivating preservative (Buffer C) with an inhibitor-removal optimized automated extraction (Self-IB/QIA-IB) forms a robust pipeline for high-fidelity gut microbiome profiling. Future work should focus on multi-center validation of this workflow and investigate strategies to account for intrinsic sample variability, further advancing its utility in precision gut microbiome diagnostics.

Materials & Methods

Sample collection and preparation

Fecal samples used for the evaluation of preservation buffers and automated extraction protocols were self-donated by a healthy adult volunteer, while samples for assessing metagenomic reproducibility were collected from five healthy adult volunteers. For the preservation buffer performance test, 10 g of a homogenized fecal sample was aliquoted into each of three 50 mL centrifuge tubes. Then, 10 mL of one of the three commercial preservation buffers (Preservation Buffer A from Genstone Biotech, Cat. No. TR110-1000; Preservation Buffer B from Shenzhen Zhongke Baikang Biotechnology, Cat. No. BM16010; Preservation Buffer C from Jiangsu Cowin Biotechnology, Cat. No. CWYO41S) was added to each tube, yielding a 1:1 (w/v) mixture. After thorough vortexing, 500 μL aliquots were dispensed into 1.5 mL tubes corresponding to time

points (Days 0, 1, 3, 5, 7, 9). Day 0 aliquots were processed immediately; others were incubated at 37°C and retrieved at the indicated times for nucleic acid extraction (using 600 μL per extraction). The remaining homogenate was stored at -80°C. For extraction optimization, samples preserved with Buffer C (from the -80°C stock) were processed in five parallel workflows: direct automated extraction (600 μL input), or pretreatment for 5 minutes with 1 mL of PBS, QIA-IB, or Self-IB prior to automated extraction (each using 600 μL of pretreated mixture), alongside a manual extraction reference (600 μL input). For reproducibility assessment, fresh samples (0.5–1g) from volunteers were collected directly into tubes containing 3 mL of Preservation Buffer C, vortexed, stored at -80°C, and extracted using the Self-IB+automated extraction protocol.

Nucleic acid extraction

The automated extraction was performed using the Stream SP96 system (DAAN GENE, China) with the corresponding Fecal DNA Automated Extraction Kit (Cat. No: DA1781). The detailed instrument settings and extraction parameters are listed in Supplementary Table 1. The manual extraction was performed using the Microbial DNA Extraction Kit (Cat. No: SL-MO0122; Shenzhen Salus Medical Technology Co., Ltd., China),following the manufacturer's instructions preciselyIn the optimization tests for the automated extraction workflow, 1 mL of each buffer—phosphate-buffered saline (PBS; Gibco®, Cat. No: 10010023), QIAamp InhibitEX Buffer (QIA-IB; QIAamp® Fast DNA Stool Mini Kit, Qiagen N.V., Cat. No: 51604), and self-prepared inhibitor removal buffer (Self-IB)—was added to the samples.

Preparation and Application of the Self-prepared InhibitEX Buffer (Self-IB)

The Self-IB was formulated to efficiently dissociate PCR inhibitors (e.g., humic acids, polysaccharides) from fecal samples. All reagents were of molecular biology grade. Briefly, 1.21 g of Tris-base was dissolved in 80 mL of nuclease-free water, the pH was adjusted to 8.0 ± 0.2 with HCl, and the volume was brought to 100 mL to yield a 100 mM Tris-HCl buffer. To this, 46.88 g of guanidine hydrochloride, 0.372 g of EDTA·2Na, and 1 mL of Triton X-100 were added sequentially with thorough mixing. The solution was stirred for 30 minutes at room temperature, filter-sterilized (0.22 μm), aliquoted, and stored at 4°C in the dark for up to 3 months.A cost analysis revealed that preparing 100 mL of our Self-IB costs approximately $1, which is estimated

to be only about 20% of the per-reaction cost of the commercial QIA-IB based on current list prices.

For automated nucleic acid extraction, 600 µL of a homogenized fecal sample (preserved in Buffer C) was mixed with 1 mL of Self-IB in a 2 mL microcentrifuge tube. The mixture was vortexed vigorously for 1 minute to ensure complete homogenization and inhibitor binding. It was then incubated at room temperature (25°C) for 5 minutes without heating. Following incubation, the sample was centrifuged at 13,000×g for 5 minutes to pellet inhibitors, residual solids, and precipitates. A volume of 600 µL of the resulting clear supernatant was carefully aspirated, avoiding the pellet, and used directly as the input for the subsequent automated extraction workflow on the Stream SP96 system.

Library construction and sequencing

Metagenomic libraries were constructed using the Human Gut Microbiota Detection Kit(Cat. No: SL-M00123; Shenzhen Salus Medical Technology Co., Ltd., China), which employs an enzymatic fragmentation-based (non-PCR) library preparation method to minimize amplification bias. The resulting libraries were sequenced on the Salus Pro platform with a paired-end 150 bp (PE150) strategy, yielding approximately 20 million raw reads per sample.

Bioinformatic analysis

Raw sequencing reads were demultiplexed and subjected to quality control using fastp (v0.23.4) with default parameters to remove low-quality bases and adapter sequences. To eliminate potential host-derived contamination, quality-filtered reads were aligned to the human reference genome (GRCh38) using Bowtie2 (v2.5.4). Reads that failed to align to the host genome were retained as high-quality, non-host clean reads for downstream taxonomic profiling.

Taxonomic profiling was performed using two complementary approaches. For the analysis of bacterial and archaeal composition, clean reads were aligned against the mpa_vJan25_CHOCOPhlAnSGB_202503 database usingMetaPhlAn4 (v4.0.3). This tool utilizes unique clade-specific marker genes to provide efficient and accurate profiling of microbial community relative abundances, with its workflow automatically integrating Bowtie2 alignment and marker gene abundance resolution.

To specifically profile fungal composition, a second analysis was performed using Kraken2 (v2.1.2)with default settings against a custom-built database. This database was constructed to enhance fungal detection sensitivity and specificity, and comprised genomic sequences from the following sources: 1) fungal genomes downloaded from the NCBI RefSeq database; and 2) genomes of culturable bacteria relevant to the gut environment, also sourced from NCBI RefSeq. The database construction followed the standard Kraken2 build protocol, ensuring comprehensive k-mer indexing for accurate taxonomic assignment of eukaryotic microbial reads.

Statistical analysis

The mean of relative abundance of the bacteria of each sample was calculated. The Shannon's index and top 10 bacteria with the highest average relative abundances were used to analyze the CV. The CV value=Standard deviation/Mean, CV value < 15% is consistent with the quality standard.

Ethics statement

This study was conducted under the framework and ethical approval of the parent research program (Approval No. KeShen 2020_100_01) granted by the Medical Ethics Committee of Mengchao Hepatobiliary Hospital of Fujian Medical University. All procedures were performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants prior to sample collection.

Funding information

This work received no specific grant from any funding agency.

Acknowledgement

We gratefully acknowledge Mr. Cheng Chen for his assistance in the metagenomic reproducibility validation experiments, and Mr. Wengen Zhou for his support during the transportation of metagenomic reagents. Besides, the authors acknowledge the use of large language model (LLM) tools for grammar checking and language polishing. All scientific content and interpretation are

the authors.

Data availability

The raw sequencing data (FASTQ files) generated in this study have been deposited in the Genome Sequence Archive[20] in National Genomics Data Center [21], China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA038748) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa. All supplementary materials associated with this article are available through the scientific reports website.

Conflicts of interest

All authors disclosed no relevant relationships.

Figure legend

Figure1.Gut metagenome analytical workflow in this study;

Fecal samples were first collected using preservation tubes containing a preservation solution. Subsequently, nucleic acids were extracted using an automated extraction system with matched reagents. The extracted nucleic acids were then utilized to construct a metagenomic library via an enzymatic digestion method, followed by sequencing on a next-generation sequencing platform. Finally, the resulting data were analyzed for species detection and their relative abundance.

Figure2. Optimization of Automated Nucleic Acid Extraction Buffers for Robust Fecal mNGS Profiling;

(A-C) Nucleic acid concentration and purity extracted from each workflow were detected using Nanodrop. For concentration results: automated extraction yielded the highest concentration, while the PBS+automated extraction workflow had the lowest concentration, followed by IB+automated extraction; the concentration from QIA-IB+automated extraction was close to that from Manual extraction (A). The trends of 260/280 and 260/230 ratios were consistent: Manual extraction resulted in the highest nucleic acid purity, with QIA-IB+automated extraction and IB+automated extraction showing similar purity levels, followed by PBS+automated extraction, and automated extraction had the poorest purity (B, C). (D) Analysis of the relative abundance of

TOP10 species detected by each workflow revealed that automated extraction had the largest difference in TOP10 species relative abundance compared to Manual extraction, followed by PBS+automated extraction; QIA-IB+automated extraction and IB+automated extraction were the closest to Manual extraction. (E) Comparative analysis of TOP30 species detected by Manual extraction and IB+automated extraction showed a slope > 0.9 and R^2 > 0.998. (F) Comparative analysis of TOP30 species detected by Manual extraction and QIA-IB+automated extraction showed a slope > 0.9 and R^2 > 0.998.

Figure3. Performance Evaluation of Fecal Preservation Buffers under Simulated Transport Conditions (37°C)；

(A-C) Nucleic acid concentration and purity extracted from different preservation buffers over Days 0–9 were measured using Nanodrop. For concentration results: Preservation Buffer B exhibited the best DNA concentration retention; Preservation Buffer C showed a decrease on Day1 and Day3 but stabilized thereafter; Preservation Buffer A had the poorest performance, with a significant concentration drop on Day1 (A). Results for the 260/230 ratio indicated that the nucleic acid purity of Preservation Buffer A decreased significantly with prolonged storage time, while Preservation Buffer B and C had comparable purity retention effects (B); results for the 260/280 ratio were similar across all three preservation buffers. (D) Statistical comparison of the Shannon index across the three preservation buffers revealed that Preservation Buffer C had the smallest coefficient of variation (CV, 2.54%), followed by Preservation Buffer A (CV, 6.54%), while Preservation Buffer B had the largest CV (14.13%). (E) Analysis of the relative abundance of top 10 detected species across the three preservation buffers over Days 0–9 showed that Preservation Buffer C maintained the most stable relative abundance of top 10 species, whereas both Preservation Buffer A and B exhibited varying degrees of changes.

Figure4. Robustness Assessment of Shannon Diversity Index in mNGS-Based Microbiome Profiling;

(A-E) Fecal samples from five university volunteers (S1-S5) were used for full-process reproducibility validation. For each sample, operations were conducted consecutively over Days 1–3 with three technical replicates per day, which served as intra-batch repeatability validation.

Figures A–E show the intra-batch repeatability results of the relative abundance of top 10 detected species for the five samples; (F) For the inter-batch repeat experiment, a different operator performed the operations as inter-batch repeatability validation. Comparison of the relative abundance of top 150 detected species between the two batches revealed $R^2 > 0.99$ and slope = 1.011.

Figure5. A Stable and Reproducible Gut Metagenomic Workflow;

After comparing preservation buffers and automated extraction pre-treatment methods, the corresponding optimal workflow was identified, and its stability was evaluated through intra-batch and inter-batch reproducibility tests.

References

[1] Odendaal ML, de Steenhuijsen Piters WAA, Franz E, et al. Host and environmental factors shape upper airway microbiota and respiratory health across the human lifespan. *Cell.* 2024;187(17):4571-4585.e15. doi:10.1016/j.cell.2024.07.008

[2] Kelly TN, Bazzano LA, Ajami NJ, et al. Gut Microbiome Associates With Lifetime Cardiovascular Disease Risk Profile Among Bogalusa Heart Study Participants. *Circ Res*. 2016;119(8):956-964. doi:10.1161/CIRCRESAHA.116.309219

[3] Lupini L, Bassi C, Guerriero P, et al. Microbiota and environmental health monitoring of mouse colonies by metagenomic shotgun sequencing. *World J Microbiol Biotechnol*. 2022;39(1):37. doi:10.1007/s11274-022-03469-0

[4] Xu MQ, Pan F, Peng LH, et al. Advances in the isolation, cultivation, and identification of gut microbes. *Mil Med Res*. 2024;11(1):34. doi:10.1186/s40779-024-00534-7

[5] Wang M, Yuan T, Chen J, et al. A species-level identification pipeline for human gut microbiota based on the V3-V4 regions of 16S rRNA. *Front Microbiol*. 2025;16:1553124. doi:10.3389/fmicb.2025.1553124

[6] Modrego J, Ortega-Hernández A, Sánchez-González S, et al. Analysis of the gut microbiota profile targeted to multiple hypervariable regions of 16S rRNA in a hypertensive heart failure rat model. *Methods Cell Biol*. 2024;188:183-203. doi:10.1016/bs.mcb.2024.04.002

[7] Vasco K, Guevara N, Mosquera J, et al. Characterization of the gut microbiome and resistome of Galapagos marine iguanas (Amblyrhynchus cristatus) from uninhabited islands. Anim Microbiome. 2022;4(1):65.doi:10.1186/s42523-022-00218-4

[8] Angel NZ, Sullivan MJ, Alsheikh-Hussain A, et al. Metagenomics: a new frontier for routine pathology testing of gastrointestinal pathogens. *Gut Pathog*. 2025;17(1):4. doi:10.1186/s13099-024-00673-1

[9] Isokääntä H, Tomnikov N, Vanhatalo S, et al. High-throughput DNA extraction strategy for fecal microbiome studies. *Microbiol Spectr*. 2024;12(6):e0293223. doi:10.1128/spectrum.02932-23

[10] Lim MY, Park YS, Kim JH, Nam YD. Evaluation of fecal DNA extraction protocols for human gut microbiome studies. *BMC Microbiol*. 2020;20(1):212. doi:10.1186/s12866-020-01894-5

[11] Ward AB, Harris PA, Argo CM, et al. Homemade Nucleic Acid Preservation Buffer Proves Effective in Preserving the Equine Faecal Microbiota over Time at Ambient Temperatures. *Animals (Basel)*. 2023;13(19):3107. doi:10.3390/ani13193107

[12] Cheng J, Venkatesh S, Ke K, Barratt MJ, Gordon JI. A human gut Faecalibacterium prausnitzii fatty acid amide hydrolase. *Science*. 2024;386(6720):eado6828. doi:10.1126/science.ado6828

[13] Gu W, Miller S, Chiu CY. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. *Annu Rev Pathol*. 2019;14:319-338. doi:10.1146/annurev-pathmechdis-012418-012751

[14] Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet*. 2019;20(6):341-355. doi:10.1038/s41576-019-0113-7

[15] Goloshchapov OV, Shchukina OB, Kusakin AV, et al. Next-Generation Sequencing-Based Monitoring of Intestinal Bacteria and Bacteriophages Following Fecal Microbiota Transplantation in Inflammatory Bowel Diseases. *Pathogens*. 2023;12(12):1438. doi:10.3390/pathogens12121438

[16] Tang WH, Kitai T, Hazen SL. Gut Microbiota in Cardiovascular Health and Disease. *Circ Res*. 2017;120(7):1183-1196. doi:10.1161/CIRCRESAHA.117.309715

[17] Asadi A, Shadab Mehr N, Mohamadi MH, et al. Obesity and gut-microbiota-brain axis: A narrative review. *J Clin Lab Anal*. 2022;36(5):e24420. doi:10.1002/jcla.24420

[18] Poeta M, Pierri L, Vajro P. Gut-Liver Axis Derangement in Non-Alcoholic Fatty Liver

Disease. Children (Basel). 2017;4(8):66. Published 2017 Aug 2. doi:10.3390/children4080066

[19] Zhai X, Castro-Mejía JL, Gobbi A, et al. The impact of storage buffer and storage conditions on fecal samples for bacteriophage infectivity and metavirome analyses. *Microbiome*. 2023;11(1):193. Published 2023 Aug 28. doi:10.1186/s40168-023-01632-9

[20] Zhang S , Chen X , Jin E , et al. The GSA Family in 2025: A Broadened Sharing Platform for Multi-omics and Multimodal Data. *Genomics Proteomics Bioinformatics*. 2025;23(4):qzaf072. doi:10.1093/gpbjnl/qzaf072

[21] CNCB-NGDC Members and Partners . Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2025. *Nucleic Acids Res*. 2025;53(D1):D30-D44. doi:10.1093/nar/gkae978

Fig1

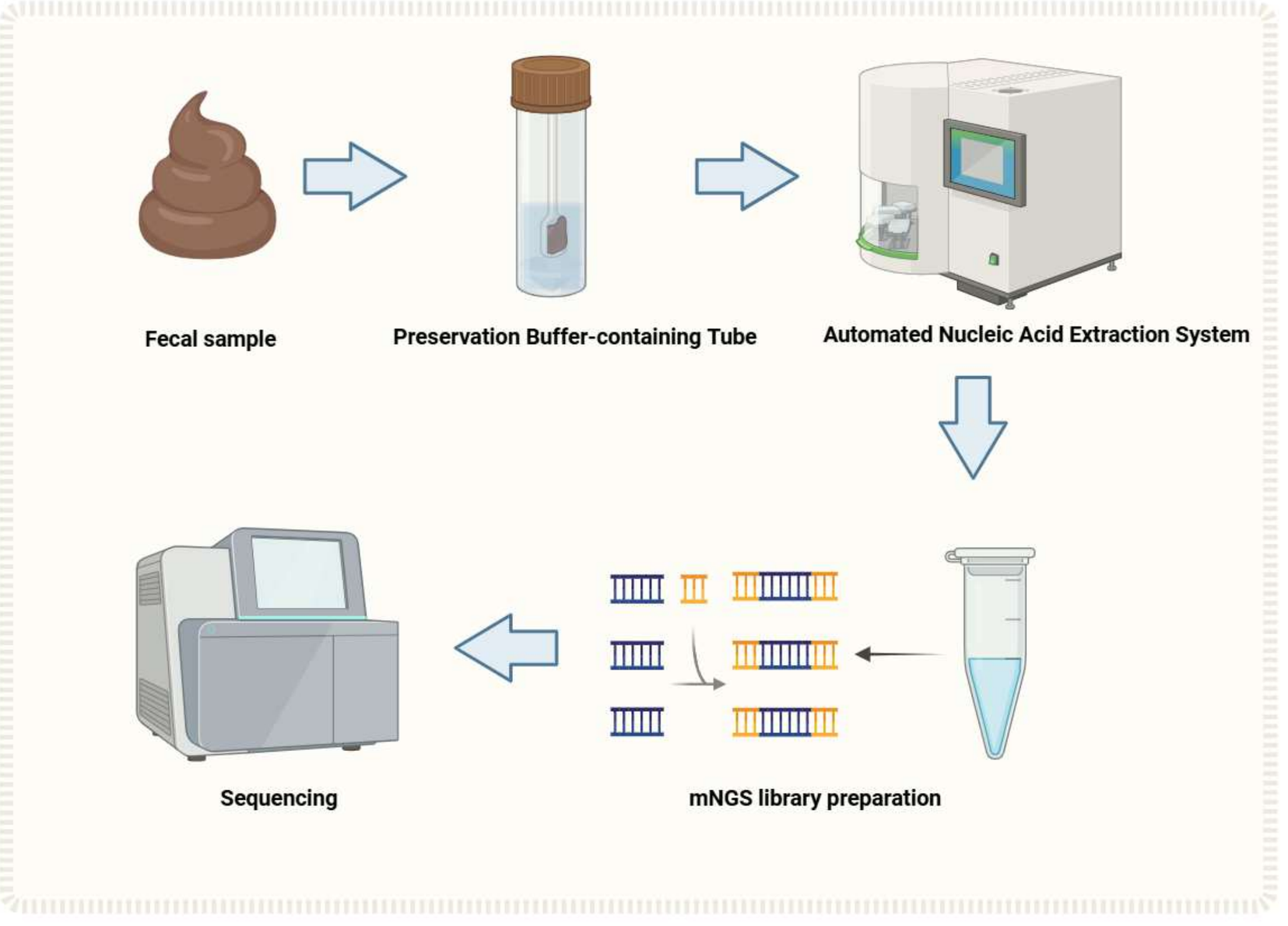

Fecal sample
Preservation Buffer-containing Tube
Automated Nucleic Acid Extraction System
mNGS library preparation
Sequencing

Fig2

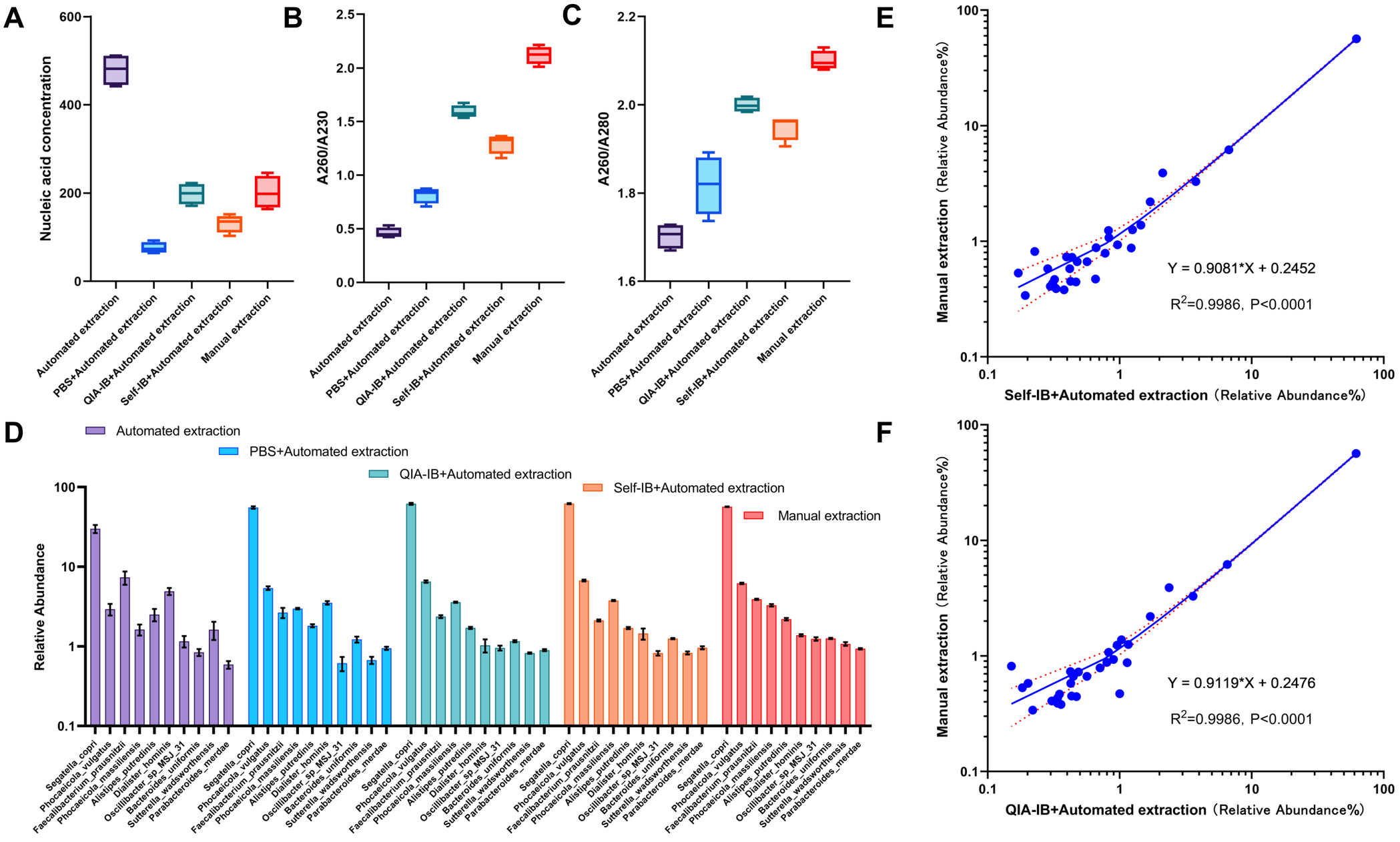
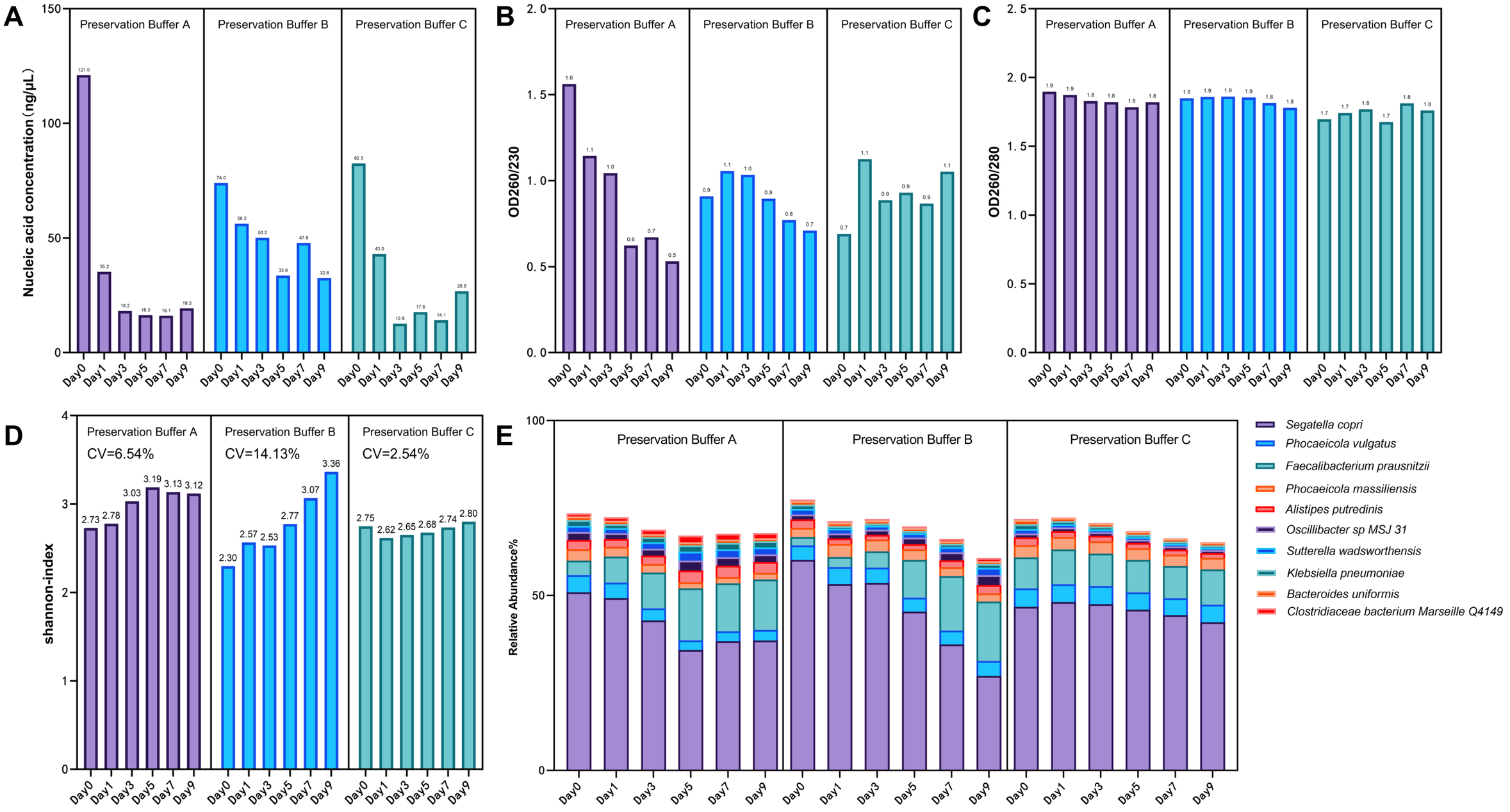

A
Nucleic acid concentration
600
400
200
0
Automated extraction
PBS+Automated extraction
QIA-IB+Automated extraction
Self-IB+Automated extraction
Manual extraction
B
A260/A230
2.5
2.0
1.5
1.0
0.5
0.0
C
A260/A280
2.2
2.0
1.8
1.6
D
Automated extraction
PBS+Automated extraction
QIA-IB+Automated extraction
Self-IB+Automated extraction
Manual extraction
Relative Abundance
100
10
1
0.1
Segatella_copri
Phocaeicola_vulgatus
Faecalibacterium_prausnitzii
Phocaeicola_massiliensis
Alistipes_putredinis
Dialister_hominis
Oscillibacter_sp_MSJ_31
Bacteroides_uniformis
Sutterella_wadsworthensis
Parabacteroides_merdae
E
Manual extraction（Relative Abundance%）
Self-IB+Automated extraction（Relative Abundance%）
Y = 0.9081*X + 0.2452
R2=0.9986, P<0.0001
F
Manual extraction（Relative Abundance%）
QIA-IB+Automated extraction（Relative Abundance%）
Y = 0.9119*X + 0.2476
R2=0.9986, P<0.0001

Fig3


A
Preservation Buffer A
Preservation Buffer B
Preservation Buffer C
Nucleic acid concentration（ng/μL）
121.0
35.2
18.2
16.3
16.1
19.3
74.0
56.2
50.0
33.6
47.8
32.6
82.5
43.0
12.6
17.6
14.1
26.8
B
OD260/230
1.6
1.1
1.0
0.6
0.7
0.5
0.9
1.1
1.0
0.9
0.8
0.7
0.7
1.1
0.9
0.9
0.9
1.1
C
OD260/280
1.9
1.9
1.8
1.8
1.8
1.8
1.8
1.9
1.9
1.9
1.8
1.8
1.7
1.7
1.8
1.7
1.8
1.8
Day0
Day1
Day3
Day5
Day7
Day9
D
shannon-index
Preservation Buffer A
CV=6.54%
2.73
2.78
3.03
3.19
3.13
3.12
Preservation Buffer B
CV=14.13%
2.30
2.57
2.53
2.77
3.07
3.36
Preservation Buffer C
CV=2.54%
2.75
2.62
2.65
2.68
2.74
2.80
E
Relative Abundance%
Preservation Buffer A
Preservation Buffer B
Preservation Buffer C
Segatella copri
Phocaeicola vulgatus
Faecalibacterium prausnitzii
Phocaeicola massiliensis
Alistipes putredinis
Oscillibacter sp MSJ 31
Sutterella wadsworthensis
Klebsiella pneumoniae
Bacteroides uniformis
Clostridiaceae bacterium Marseille Q4149

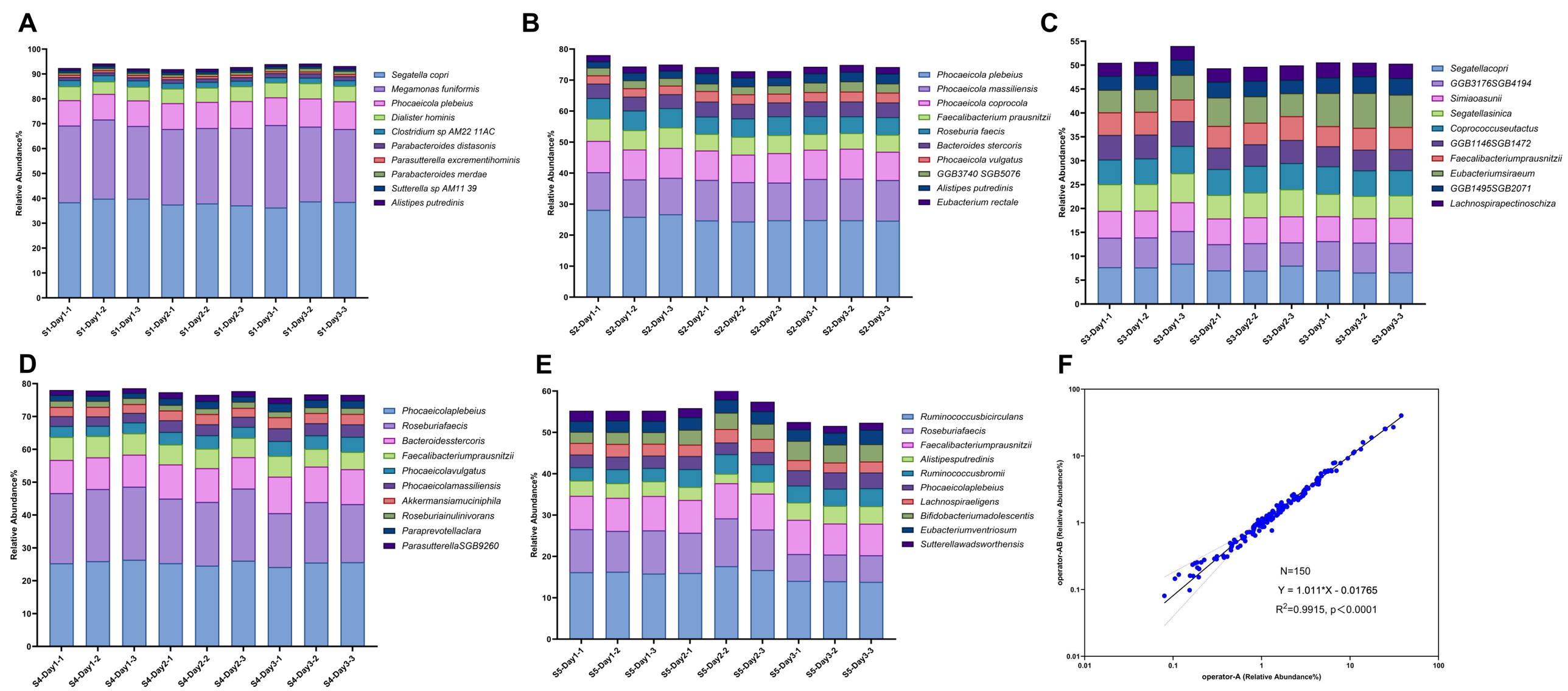

A
Relative Abundance%
S1-Day1-1
S1-Day1-2
S1-Day1-3
S1-Day2-1
S1-Day2-2
S1-Day2-3
S1-Day3-1
S1-Day3-2
S1-Day3-3
Segatella copri
Megamonas funiformis
Phocaeicola plebeius
Dialister hominis
Clostridium sp AM22 11AC
Parabacteroides distasonis
Parasutterella excrementihominis
Parabacteroides merdae
Sutterella sp AM11 39
Alistipes putredinis
B
Relative Abundance%
S2-Day1-1
S2-Day1-2
S2-Day1-3
S2-Day2-1
S2-Day2-2
S2-Day2-3
S2-Day3-1
S2-Day3-2
S2-Day3-3
Phocaeicola plebeius
Phocaeicola massiliensis
Phocaeicola coprocola
Faecalibacterium prausnitzii
Roseburia faecis
Bacteroides stercoris
Phocaeicola vulgatus
GGB3740 SGB5076
Alistipes putredinis
Eubacterium rectale
C
Relative Abundance%
S3-Day1-1
S3-Day1-2
S3-Day1-3
S3-Day2-1
S3-Day2-2
S3-Day2-3
S3-Day3-1
S3-Day3-2
S3-Day3-3
Segatellacopri
GGB3176SGB4194
Simiaoasunii
Segatellasinica
Coprococcuseutactus
GGB1146SGB1472
Faecalibacteriumprausnitzii
Eubacteriumsiraeum
GGB1495SGB2071
Lachnospirapectinoschiza
D
Relative Abundance%
S4-Day1-1
S4-Day1-2
S4-Day1-3
S4-Day2-1
S4-Day2-2
S4-Day2-3
S4-Day3-1
S4-Day3-2
S4-Day3-3
Phocaeicolaplebeius
Roseburiafaecis
Bacteroidesstercoris
Faecalibacteriumprausnitzii
Phocaeicolavulgatus
Phocaeicolamassiliensis
Akkermansiamuciniphila
Roseburiainulinivorans
Paraprevotellaclara
ParasutterellaSGB9260
E
Relative Abundance%
S5-Day1-1
S5-Day1-2
S5-Day1-3
S5-Day2-1
S5-Day2-2
S5-Day2-3
S5-Day3-1
S5-Day3-2
S5-Day3-3
Ruminococcusbicirculans
Roseburiafaecis
Faecalibacteriumprausnitzii
Alistipesputredinis
Ruminococcusbromii
Phocaeicolaplebeius
Lachnospiraeligens
Bifidobacteriumadolescentis
Eubacteriumventriosum
Sutterellawadsworthensis
F
operator-AB (Relative Abundance%)
operator-A (Relative Abundance%)
N=150
Y = 1.011*X - 0.01765
R2=0.9915, p＜0.0001

Fig4

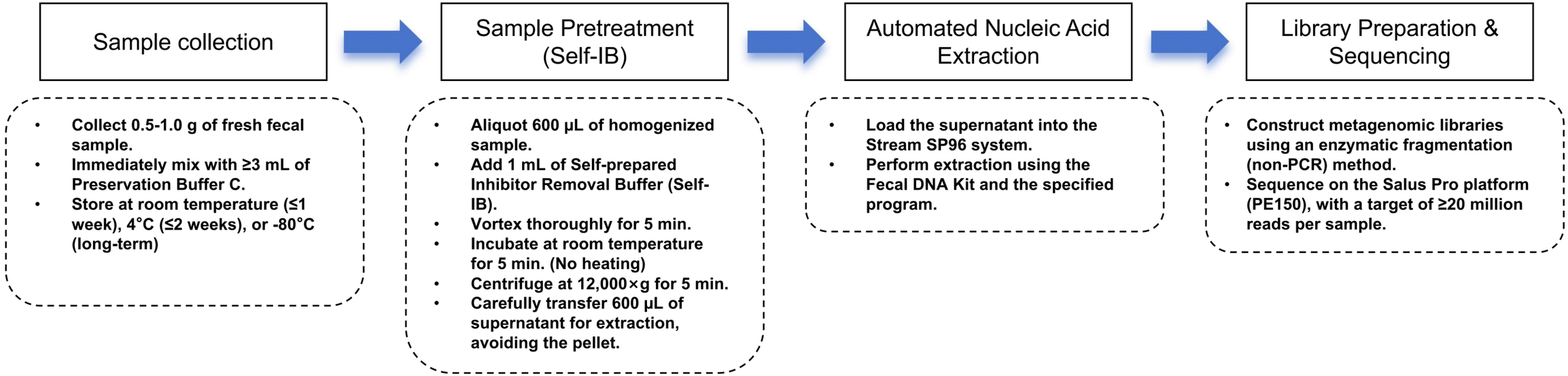

Sample collection
Sample Pretreatment (Self-IB)
Automated Nucleic Acid Extraction
Library Preparation & Sequencing
• Collect 0.5-1.0 g of fresh fecal sample.
• Immediately mix with ≥3 mL of Preservation Buffer C.
• Store at room temperature (≤1 week), 4°C (≤2 weeks), or -80°C (long-term)
• Aliquot 600 µL of homogenized sample.
• Add 1 mL of Self-prepared Inhibitor Removal Buffer (Self-IB).
• Vortex thoroughly for 5 min.
• Incubate at room temperature for 5 min. (No heating)
• Centrifuge at 12,000×g for 5 min.
• Carefully transfer 600 µL of supernatant for extraction, avoiding the pellet.
• Load the supernatant into the Stream SP96 system.
• Perform extraction using the Fecal DNA Kit and the specified program.
• Construct metagenomic libraries using an enzymatic fragmentation (non-PCR) method.
• Sequence on the Salus Pro platform (PE150), with a target of ≥20 million reads per sample.

Fig5

Table 1. The CV value of top 10 most abundant bacteria of mNGS validation*

	CV value of intra-batch (%)		CV value of inter-batch (%)	
Sample	Genus	Species	Genus	Species
S1	3.08-15.42(7.67, 4.13)	3.08-17.24(8.11, 4.86)	1.73-13.80(6.94, 3.53)	1.77-13.01(7.31, 3.27)
S2	2.17-25.69(9.47, 7.26)	3.07-15.87(7.54, 4.33)	1.43-9.94(4.94, 2.94)	2.57-7.70(4.11,1.73)
S3	3.63-15.53(7.87, 3.56)	3.73-17.53(7.93, 4.00)	4.54-11.11(7.99, 1.99)	4.54-10.49(7.86, 1.86)
S4	2.22-17.95(9.01, 3.97)	2.79-17.95(9.31, 4.75)	1.60-7.84(4.31, 2.07)	2.17-7.84(4.80, 2.16)
S5	4.69-21.77(13.30, 6.38)	4.68-22.17(13.68, 5.55)	2.79-10.56(6.58, 2.65)	2.49-10.71(7.14, 2,88)
Total	2.17-25.69(9.46, 2.27)	2.79-22.17(9.31, 2.53)	1.43-13.80(6.15, 1.50)	1.77-13.01(6.24, 1.67)

*The value was presented as lower limit-upper limit (mean, standard deviation).

Table 2. The CV value of Shannon's index

Sample	intra-batch	inter-batch
S1	2.27	1.59
S2	2.03	1.08
S3	1.68	5.94
S4	1.57	1.66
S5	1.76	1.03

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementtable1.xlsx