Advancements in semen sample analysis: evaluating the efficacy and progress in human DNA removal techniques

Lei Hua1,†, Yinuo Jiang2,3,†, Dongliang Lu2, Bin Wang2, Baoquan Han2,*, Junqing Hou4,*, Zhongyi Sun1,2,*

Abstract
Microorganism types and quantities in male semen correlate with male semen parameters and male infertility. There has been a growing body of evidence suggesting that identifying microorganisms in male semen through metagenomic sequencing is useful for assessing male fertility and planning appropriate treatment interventions. Nevertheless, non-targeted metagenomic sequencing of semen samples is challenging due to the large number of sperm cells and high amounts of intracellular and extracellular human DNA, which can obscure microbial DNA. In this study, we assessed and refined three methods for DNA extraction from semen. Using low-speed centrifugation, propidium monoazide (PMA) photochemical induction, and the HostZERO Microbial DNA Kit, metagenomic sequencing was performed to detect and evaluate its efficacy. Our findings indicate that low-speed centrifugation effectively reduces sperm and sperm-associated DNA but may inadvertently remove microbial DNA. Increasing sample quantities is a potential strategy for enhancing the microbial DNA ratio in semen samples. PMA and the HostZERO Microbial DNA Kit utilization and optimization show promise in reducing total semen DNA and augmenting microbial DNA representation. Preliminary results suggest that traditional disinfection methods during semen sample collection do not increase microbial contamination significantly. In particular, due to the limited sample size, further investigation with a larger cohort is necessary for conclusive validation of our findings.

Keywords
Microbiome sequencing; Host DNA depletion; Human semen; Male infertility

1. Background
Presently, 9% of couples suffer from infertility challenges, with male factors accounting for around 50% of cases [1]. Varicocele, hypogonadism, cryptorchidism and obstructive diseases [2], are among the myriad conditions contributing to abnormal semen parameters in approximately 85% of male infertility cases [3]. The abnormalities manifest as a decline in sperm concentration, diminished motility, and increased morphological abnormalities among sperm. There is a correlation between male infertility, abnormal semen parameters, and the presence, types and quantity of microorganisms in semen [4].

Male semen was traditionally considered sterile. Researches have challenged this notion since several decades ago [4–6], showing that the semen of at least some healthy males and most males with reduced fertility is not entirely sterile [7]. In male semen, bacterial RNA sequencing has identified a number of potentially beneficial or harmful bacteria, including Lactobacillus, Anaerococcus, Pseudomonas and Klebsiella [4, 8]. Elevated levels of harmful bacteria and diminished levels of beneficial bacteria in semen are associated with suboptimal semen parameters. During bacterial DNA sequencing, human DNA sequencing capabilities, reduces accuracy, and decreases microbial data volume at identical sequencing depths. To address this issue, methods have been developed to eliminate host DNA.

Several pre-extraction techniques are employed to eradicate intracellular host cell DNA, including filtration, density gradient centrifugation, and detergents such as sterile water, soap, Triton X-100, Tween 20 and ionic surfactants. Extracellular DNA can also be removed with DNA enzymes or propidium monoazide (PMA). These methods have been formulated into commercially available kits, including QIAamp DNA Microbiome, MolYsis™ Complete/UltraDeep Microbiome Prep, and HostZERO Microbial DNA Kit. High-throughput targeted capture techniques have also been developed using viral and bacterial DNA as probes for sequence-specific hybridization with target DNA. NEBNext Microbiome DNA Enrichment Kit and LOOXETER Enrichment Kit serve this purpose.
The human genome is nearly 1000 times larger than microbial genomes (3.2 Gb vs. 3.6 Mb) [9], and sperm counts in semen are approximately 5000 times higher than bacteria counts (49,000,000/mL vs. 10,000/mL) [10, 11], suggesting human DNA significantly outnumber bacterial DNA. Thus, enhancing bacterial DNA requires efficient methods of removing human DNA. In this study, we aim to evaluate the efficacy of these methods on human semen samples. Human sperm, which contains a limited volume and resembles certain microbial entities, was tested on a small group of volunteers using centrifugation, PMA photochemical induction, and the HostZERO Microbial DNA Kit. The 5 μm filter technique was not used. The processed samples were then subjected to metagenomic sequencing analysis to detect and assess the potential in regional bias. No statistically significant differences in regional diversity were found, indicating a lack of significant variation in microbial DNA abundance across different regions.

2. Materials and methods

2.1 Test materials

2.1.1 Volunteer recruitment

Fundamental information was collected through a questionnaire. This questionnaire was distributed by the Department of Reproductive Medicine, Qingdao Municipal Hospital. Respondents were required to provide fundamental personal information and pertinent medical histories as exclusion criteria. Volunteers were informed of the experimental objectives before semen collection. Data security protocols ensure confidentiality of respondents’ personal information and questionnaire outcomes. Semen samples from adult males meeting the criteria are earmarked for use in subsequent experiments. Using the aforementioned criteria, 16 volunteers were identified as meeting the inclusion criteria. All volunteers exhibited healthy semen quality (i.e., no conditions affecting semen quality, such as oligospermia or azoospermia). Among them, 15 took part in routine medical examinations at the hospital. A distinctive case involved a patient who underwent kidney stone surgery at the hospital. Notably, the average age of these volunteers is 29.72 ± 5.21 (mean ± standard deviation). Volunteers are primarily from Qingdao City, with no statistically significant differences in regional diversity, potentially leading to latent regional bias.

2.1.2 Main reagents

Sperm capacitate solution Dulbecco’s Modified Eagle Medium/Nutrient Mixture F-12 medium (DMEM/F12, SH30023.0, Hyclone, UT, USA) with 10% Fetal Bovine Serum (FBS, 10099-141, Gibco, Grand Island, NE, USA), sterile water, 0.2 mmol/L PMA solution, HostZERO Microbial DNA Kit (D4310, Zymo Research, CA, USA), etc.

2.2 Experimental design

This experiment employs three methods—low-speed centrifugation, propidium monoazide (PMA) photochemical induction, and the HostZERO Microbial DNA Kit—to eliminate human DNA from human sperm cells’ intracellular and extracellular components. Microbial DNA is then enriched in the semen using these methods. Next, metagenomic sequencing of the samples is performed, with a sequencing depth of 1 Gb per sample. Samples with Qubit concentrations below 0.2 ng/µL are excluded from the analysis. Based on the remaining data obtained from the samples, the NonHostData/CleanData ratio is calculated, which indicates the amount of microbial DNA in total DNA. Comparative analyses are conducted based on these ratios.

2.3 Test methods

2.3.1 Human semen sample collection standards

Participants are required to observe a 3–7-day abstinence period before sample procurement. They are instructed to thoroughly clean their glans and coronal sulcus with soapy water, then change into freshly laundered clothing on the day of sampling. Rigorous hand hygiene is strictly recommended prior to formal sampling, with a 2–3-time washing regimen with soapy water, followed by a 2–3-time disinfection with 75% alcohol in the glans and coronal sulcus. Manually collect seminal fluid through masturbation and direct deposit ejaculates into a sterile receptacle subsequently [11].

2.3.2 Sample pretreatment

The semen samples are incubated undisturbed at 37 °C for 30 minutes to allow natural liquefaction. Aseptic pipette tips to ensure sterility during agitation and homogenization. 1 mL of liquefied semen is carefully aspirated and mixed with 500 μL of sperm capacitation solution. It is then subdivided into 4 or 6 aliquots for further examination or processing. Fig. 1 shows the first experiment used 4 aliquots, whereas each of the following three experiments employed 6 aliquots each. In particular, aliquots are processed with the HostZERO Microbial DNA Kit, optimizing treatment efficiency. Using the kit strategically enhances the overall effectiveness of subsequent analyses or treatments.

2.3.3 Sample centrifugation and hypotonic depletion of human DNA from sperm cells

Due to the similar size of sperm cells and certain microorganisms, we avoided filtration methods and instead employed low-speed centrifugation. Recognizing the potential impact of centrifugation speed and duration on sperm removal efficiency, we conducted comparisons comparing speeds and durations. The accompanying schematic diagram shows details of the specific time and speed parameters for each trial. Two centrifugation steps were taken. Initially, the precipitate was centrifuged to remove sperm. We performed a second centrifugation using the supernatant from the initial centrifugation. This additional centrifugation promotes the sedimentation of both microorganisms and residual sperm cells. 1 mL of sterile water is added to the resultant supernatant, followed by brief vortexing. Subsequently, the mixture is allowed to stand for 5 minutes in a water bath, leveraging osmotic pressure to induce swelling and rupture of the remaining sperm cells and liberate human DNA. Despite rigid cell walls, microorganisms remain impervious to osmotic pressure and do not swell or rupture.
2.3.4 Depletion of extracellular human DNA by PMA treatment

Propidium monoazide (PMA) intercalates DNA through cell membrane impermeability. Covalently modifies DNA without light, preventing subsequent amplification. Exposure to light can render residual PMA inactive. PMA stands out as a preferable alternative to DNA nucleases, obviating the necessity for washing steps that might lead to inadvertent DNA loss. The protocol involves adding PMA to the sample to achieve a final concentration of $10^{12} \mu M$, followed by a brief vortexing step. The sample then undergoes a 5-minute incubation period in darkness at room temperature. It is then placed on ice and positioned 20 cm away from a standard desktop 650 W halogen lamp for 25 minutes. The sample undergoes a brief cycle of centrifugation and vortexing every 5 minutes. Ultimately, the treated sample is frozen at $-20^\circ C$ for preservation.

2.3.5 HostZERO Microbial DNA Kit treatment removes extracellular human DNA

HostZERO Microbial DNA Kit utilizes an innovative approach to address host DNA contamination by selectively lysing eukaryotic cells and subsequently degrading this DNA before the overall DNA purification process. Zymo Research’s impartial purification technology seamlessly complements this method, which distinctively captures DNA from viable microbial cells in biological samples. As per the manufacturer’s guidelines, the protocol is outlined as follows: first, eliminating eukaryotic host DNA from the sample; second, unbiased lysis of the remaining microbial cells; and finally, isolating microbial DNA.

2.3.6 Metagenomic sequencing

Non-targeted metagenomic sequencing involves the comprehensive sequencing of total DNA from all organisms within a sample without specific primers targeting particular sequences. It captures the entire genetic spectrum of bacteria, fungi, viruses, parasites, and other entities present in the sample. Non-targeted metagenomic sequencing offers superior taxonomic resolution and facilitates gene functional analysis compared to targeted high-throughput sequencing. However, its efficacy may be compromised in samples containing a substantial host DNA background. Therefore, methods for host DNA removal were developed to optimize metagenomic sequencing efficiency. In preparation for metagenomic sequencing of human semen samples, three distinct methods—centrifugation, PMA treatment, and HostZERO Microbial DNA Kit treatment—were employed to mitigate host DNA in human semen. The processed samples were then transported on dry ice to the sequencing facility, where metagenomic sequencing was conducted, which generated 1 Gb of data for each sample. The dataset enabled a comprehensive assessment and evaluation of the efficacy of these methods within the specific context of human semen metagenomic sequencing.

2.3.7 Data processing and analysis

Metagenomic sequencing data were analyzed using Prism 8 software, with an emphasis on comparing total DNA concen-
trations between treatment groups and the DNA/Total DNA ratio. The DNA/Total DNA indicates the amount of microbial DNA relative to total DNA. Inter-group comparisons were conducted using t-tests, and all reported p-values reflect a two-sided analysis. Significance levels were established at p < 0.05 to determine statistical significance.

3. Results and analysis

We conducted four experiments, labeled 1st, 2nd, 3rd and 4th, respectively, as shown in Fig. 1. Additional clarification is provided in the detailed description of each experiment. 16 volunteers participated in this study, and each volunteer’s semen was assigned as a distinct sample for the corresponding experiment. The experimental design encompassed an initial phase involving 5 groups, followed by a second round with 4 samples, a third round with 4 samples, and a final round with 3 samples.

The 1st experiment evaluated the efficacy of removing sperm at varying centrifugation speeds. The 2nd experiment assessed whether sample volume influenced the ultimate fraction of microbial DNA. This experiment also compared PMA and HostZERO reagent kits’ effects on semen, in comparison with the 1st experiment. In the 3rd experiment, procedural refinements were systematically compared to the H1 group in the 2nd experiment to gauge their impact. Lastly, the 4th experiment sought to determine the presence of detectable microbial contamination during sampling.

3.1 1st experiment

In the 1st experiment, we concentrated on centrifugation to eliminate host DNA, evaluating the efficacy of different centrifugation speeds on sperm removal. Semen samples were collected from 5 participants and divided into four groups after pre-processing: 1 Control group (no centrifugation); 2 E1.5 (1500 rpm × 3 minutes); 3 E2 (2000 rpm × 3 minutes); 4 E3 (3000 rpm × 3 minutes), with 5 samples allocated to each group. As illustrated in the graph, centrifugation significantly reduced the total DNA amount compared to the control. Furthermore, the total DNA amount decreased gradually with increasing centrifugation speed as the Fig. 2A shows, but no statistically significant differences were observed at different speeds as the Fig. 2B shows. Since semen contains a negligible amount of bacterial DNA compared to human DNA, the decline in total DNA amount following centrifugation primarily signifies host DNA removal. Centrifugation effectively eliminates both sperm and internal DNA from human semen.

To evaluate the amount of microbial DNA in total DNA, we infer that low-speed centrifugation effectively reduces the total DNA amount in semen without increasing the amount of microbial DNA. This is likely attributed to the concurrent removal of most microorganisms associated with sperm cells during centrifugation.

3.2 2nd experiment

In the 2nd experiment, we aimed to explore whether sample volume would influence the amount of microbial DNA. To address this, samples were divided into two groups: H1 group with a sample volume of 1 mL; H2 group with a sample volume of 2 mL, featuring 2 samples in each group. Following the pre-processing of each sample, low-speed centrifugation (1500 × 10 minutes, and (1600 rpm, 1700 rpm, 1800 rpm, 1900 rpm, 2000 rpm) × 3 minutes) was used to predominantly remove sperm, followed by high-speed centrifugation at 60,000g for 20 minutes to isolate microbial DNA. Both groups’ samples were now standardized in volumes. After remove extracellular host DNA with PMA and the HostZERO Microbial DNA Kit, samples from the same participants in each group were combined and submitted for analysis.

There was no significant difference in total DNA concentration between the two groups as the Fig. 3A shows, possibly due to the limited sample size. Furthermore, the H2 group had marginally higher microbial bacterial DNA than the H1 group as the Fig. 3B shows, although the difference was not statistically significant. This suggests that increasing the sample concentration might enhance the bacterial DNA amount. Nevertheless, further studies with a larger sample size are necessary to substantiate and validate these preliminary findings.

3.3 3rd experiment

In the 3rd experiment, we modified specific steps, based on the findings in the 2nd experiment. However, we maintained a sample volume of 1 mL to facilitate a direct comparison with the H1 group in the 2nd experiment. Following pre-processing, for each 1 mL sample, low-speed centrifugation at 2000 rpm for 10 minutes was used to eliminate the majority of sperm. Microbial DNA was isolated using high-speed centrifugation at 60,000g for 20 minutes was. Following the removal of extracellular host DNA using PMA and the HostZERO Microbial DNA Kit, samples from the same participant were amalgamated and sent for analysis. Improvements implemented include: (1) Addition of 20 µL of proteinase K to each sample, followed by 1-minute vortexing, followed by incubation at 55 °C for 15 minutes. (2) Addition of 100 µL of DNA/RNA Shield™ (2X concentrate) to each sample, accompanied by 1-minute vortexing and 15-minute incubation at room temperature. (3) Addition of 750 µL of ZymoBIOMICS™ Lysis Solution to each sample, followed by 10-minute vortexing and centrifugation at 12,000g for 1 minute.

The results indicated a notably lower total DNA concentration than the H1 group in the 2nd experiment as the Fig. 4A shows. However, t-tests revealed no statistically significant differences between the groups due to the limited sample size (only 2 samples in each group). Although the amount of
**FIGURE 2. The results of the 1st experiment.** (A) The result of the concentration of total DNA in the 1st experiment. (B) The result of the amount of microbial DNA in total DNA in the 1st experiment. **indicates statistical significance, $p$-value $< 0.001$.

**FIGURE 3. The results of the 2nd experiment.** (A) The result of the concentration of total DNA in the 2nd experiment. (B) The result of the amount of bacterial DNA in total DNA in the 2nd experiment.
microbial DNA in total DNA was higher than the H1 group in the 2nd experiment as the Fig. 4B shows, to t-tests did not reveal any significant differences. Certain modifications to the HostZERO Microbial DNA Kit may improve results when applied to semen. For these findings to be validated, further research with a larger sample size is needed.

3.4 4th experiment

In the 4th experiment, we aimed to examine potential microbial contamination during sampling by collaborating with volunteers and local healthcare professionals. To minimize the risk of microbial contamination, we collected samples from a patient with healthy semen who had kidney stones in the operating room while maintaining sterility throughout the procedure. Simultaneously, two specimens obtained through conventional methods served as the control group (Group C). Following pre-processing, low-speed centrifugation at 8000g for 6 minutes was used to predominantly eliminate sperm, followed by high-speed centrifugation at 60,000g for 20 minutes to isolate microbial DNA. The HostZERO Microbial DNA Kit was used to eliminate extracellular host DNA, incorporating the enhancements implemented in the 3rd experiment, before sending the samples for analysis.

The results showed that the amount of microbial DNA in total DNA was similar in Groups A and C as the Fig. 5 shows. Thus, semen obtained under completely sterile conditions on the operating table and semen acquired through routine methods after disinfection did not differ significantly in microbial content. Furthermore, routinely obtained semen did not show any noticeable microbial contamination.

4. Discussion

Microorganism types and quantities in male semen correlate with male semen parameters [4], and an in-depth exploration of microorganisms in male semen is imperative to preserve and enhance male reproductive abilities. Non-targeted metagenomic sequencing for detecting microbial genes in semen, however, presents significant challenges due to the presence of sperm and the extensive human genome relative to microorganisms. Consequently, non-targeted metagenomic sequencing requires effective and reliable methods for removing host DNA. Presently, host DNA removal strategies are primarily focused on two aspects [12].

Methods such as filters with varying pore sizes, differential centrifugation, density gradient centrifugation, and flow cytometry are used to eliminate host cells before DNA extraction. The remaining host cells are then lysed with sterile water, saponin, Triton X-100, Tween 20, or zwitterionic detergents. To degrade host DNA, DNA enzymes or PMA are used. Kits have been developed based on these methodologies.

As an alternative approach, a highly multiplexed sequence capture method is used, in which known viral and bacterial DNA are utilized as probes for sequence-specific hybridization with target DNA, thus effectively isolating microbial DNA.
from the host background. In this study, we explored the impacts of low-speed centrifugation, PMA photochemical induction, and the HostZERO Microbial DNA Kit on human semen, implementing some enhancements.

Human DNA was initially removed from sperm cells using low-speed centrifugation. Centrifugation significantly reduced the semen DNA content compared to non-centrifugation. Additionally, as centrifugation speed increased, semen DNA content decreased consistently, indicating centrifugation’s effectiveness in removing sperm and their DNA.

Unexpectedly, centrifugation did not increase the amount of microbial DNA. This phenomenon may be related to microbial attachment to the sperm surface [4] or their presence inside sperm [13]. While centrifuging, microbes attached to sperm were simultaneously removed, resulting in microbial DNA loss. We further eliminated human DNA using PMA and the HostZERO Microbial DNA Kit and investigated the relationship between the initial sample volume and the final amount of microbial DNA. Doubling the initial sample volume hinted at a potential increase in the final amount of microbial DNA. However, no significant differences were observed between both groups, possibly due to the small sample size.

Following our considerations, we modified the experimental steps of the HostZERO Microbial DNA Kit based on our considerations, primarily extending their duration. In response to this adjustment, the amount of microbial DNA almost doubled. However, due to the limited number of experimental samples, a statistically significant difference could not be demonstrated. Further research with a larger sample size is necessary to validate these findings.

Finally, we examined the possibility of microbial contamination during sampling. Comparing a sample collected on a sterile operating table with one obtained through routine methods after disinfection showed a similar amount of microbial DNA. Therefore, sampling after routine disinfection may not cause noticeable microbial contamination. For confirmation of this observation, larger sample sizes are needed.

5. Conclusions

In summary, we studied the impacts of low-speed centrifugation, PMA photochemical induction, and the HostZERO Microbial DNA Kit on human semen samples and incorporated some improvements. Low-speed centrifugation has been found to be an effective method of removing sperm and their DNA, as well as microbial DNA. Semen samples could potentially contain more microbial DNA if the sample volume is increased. Along with refined methods, PMA and the HostZERO Microbial DNA Kit may effectively reduce semen DNA content and increase the amount of microbial DNA.
Obtaining semen samples after masturbation does not seem to result in significant microbial contamination. As such, our study is limited, particularly with a small sample size, which may lead to bias and affect both statistical differences and the comprehensive presentation of results. Ultimately, this constraint limits the ability to provide a detailed control and draw precise conclusions, resulting in only a broad overview of the findings. Reducing sequencing depth may impact metagenomic data comprehensiveness, potentially omitting low-abundance microbial species or rare variations.

AVAILABILITY OF DATA AND MATERIALS
The data presented in this study are available on reasonable request from the corresponding author.

AUTHOR CONTRIBUTIONS
ZYS, JQH and BQH—designed the research; provided project oversight. LH and YNJ—performed the research; wrote the paper. DLL and BW—generated and analyzed the data. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE
The collection of human samples was reviewed and approved by the Institutional Research Human or Animal Ethics Committee of Qingdao Municipal Hospital (ref. 2021-064), and all experiments adhered to the regulations of this review board. Each participant signed an informed consent form prior to participation in this study.

ACKNOWLEDGMENT
We are grateful for the technical support provided by Novogene Co., Ltd. (Beijing, China).

FUNDING
This research was funded by the National Natural Science Foundation of China, grant number 82271637, Shenzhen Science and Technology Program, grant number JCYJ20220531103004008 and JCYJ20210324105608022, International Cooperative Research Program of Shenzhen, grant number GJHZ20220913143006011, and Medicine-Engineering Interdisciplinary Research Foundation of Shenzhen University (2023YG018).

CONFLICT OF INTEREST
The authors declare no conflict of interest.

REFERENCES