

# A Refined, Controlled 16S rRNA Gene Sequencing Approach Reveals Limited Detection of Cerebrospinal Fluid Microbiota in Children with Bacterial Meningitis

**Christopher E. Pope**

University of Washington

**Kathryn B. Whitlock**

New Harmony Statistics

**Paul Hodor**

Seattle Children's

**David L. Limbrick**

Washington University in St. Louis

**Patrick J. McDonald**

University of British Columbia

**Jason Hauptman**

University of Washington

**Lucas R. Hoffman**

University of Washington

**Tamara Simon** (✉ [tsimon@chla.usc.edu](mailto:tsimon@chla.usc.edu))

University of Southern California <https://orcid.org/0000-0003-4734-7020>

---

## Research

**Keywords:** bacterial meningitis, high throughput sequencing, cerebrospinal fluid.

**Posted Date:** October 21st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-966673/v1>

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

[Read Full License](#)

---

# Abstract

## Background:

Advances in both laboratory and computational components of high-throughput 16S amplicon sequencing (16S HTS) have markedly increased its sensitivity and specificity. Additionally, these refinements have better delineated the limits of sensitivity, and contributions of contamination to these limits, for 16S HTS that are particularly relevant for samples with low bacterial loads such as human cerebrospinal fluid (CSF). The objectives of this work were to (1) optimize the performance of 16S HTS in CSF samples with low bacterial loads by defining and addressing potential sources of error; and (2) perform refined 16S HTS on CSF samples from children diagnosed with bacterial meningitis and compare results with those from microbiological cultures.

## Results:

Several bench and computational approaches were taken to address potential sources of error for low bacterial load samples. We compared DNA yields and sequencing results after applying three different DNA extraction approaches to an artificially-constructed ("mock") bacterial community. We also compared two post-sequencing computational contaminant removal strategies, decontam R and full contaminant sequence removal. All three extraction techniques followed by decontam R yielded similar results for the mock community.

We then applied these methods to 22 CSF samples from children diagnosed with meningitis, which has low bacterial loads relative to other clinical infection samples. The refined 16S HTS pipelines identified the cultured bacterial genus as the dominant organism for only 3 of these samples.

## Conclusions:

We found that all three DNA extraction techniques followed by decontam R generated similar DNA yields for mock communities at the low bacterial loads representative of CSF samples. However, the limits of detection imposed by reagent contaminants and methodologic bias precluded the accurate detection of bacteria in CSF from children with culture-confirmed meningitis using these approaches despite rigorous controls and sophisticated computational approaches. Although we did not find current DNA-based diagnostics to be useful for pediatric meningitis samples, the utility of these methods for CSF shunt infection remains undefined. Future advances in sample processing methods to minimize or eliminate contamination will be required to improve the sensitivity and specificity of these methods for pediatric meningitis.

## Background

Cerebrospinal fluid (CSF) is generally considered to be sterile. When infection is clinically suspected, the identification of microorganisms in CSF has traditionally relied upon conventional bacterial cultures,

which are tailored to identify specific human pathogens. Our earlier work focused on children with hydrocephalus requiring CSF shunts, in whom CSF infection is generally thought to involve specific microbes adherent to shunt hardware and that are, due to their physiological states, variably detectable by culture. Indirect evidence suggests that CSF shunt infections commonly involve microbes that are not detected by conventional bacterial cultures but that can be identified by molecular microbiological tools. [1] We previously used two molecular microbiological tools – high-throughput 16S amplicon sequencing (16S HTS) and quantitative PCR (qPCR) - to characterize the DNA of culturable and non-culturable microbes to detect [2] and quantify total CSF bacterial loads [3] in CSF from children with clinically-diagnosed shunt infections. These earlier studies detected low CSF levels of DNA from diverse bacteria and fungi that were not detected by conventional bacterial culture.

Since then, advances in both the laboratory and computational components of these molecular microbiological approaches have markedly increased both sensitivity and specificity of both these approaches, especially 16S HTS. These refinements have also better delineated the limits of sensitivity, and contributions of contamination to these limits, for 16S HTS that are particularly relevant for samples with low bacterial loads, such as human CSF. For example, reagents used for 16S HTS are now known to commonly introduce contaminating bacterial DNA, as well as bias for or against specific microbial taxa, which can either obscure the contribution of truly infecting microbes or erroneously identify contaminating bacteria in human infection samples. [4] These experiences have identified specific methodologic refinements, including laboratory and computational controls and systematic bench techniques, that are now in broad use to distinguish true sample microbiota from experimental error, to rigorously exclude contaminants, to delineate and improve the limits of detection, and to define clinical utility.

For sample types with high bacterial loads, such as fecal samples, the contributions of the confounders described above are comparably minor. However, in the context of low bacterial load samples, these details can significantly impact accuracy. [5] There are three known main sources of error in microbiota analysis of low-load samples, each of which requires the inclusion of specific controls to address them: (1) contamination of DNA extraction kits and PCR reagents with bacteria and/or free DNA (reagent contamination); [6] (2) differences in DNA extraction efficiency for different bacterial species (extraction bias); [7] and (3) variation of DNA extraction results, especially from samples with low bacterial load, according to batch of reagents used and from day to day (batch effect). [6, 8, 9]

The objectives of this work were to (1) optimize the performance of 16S HTS in CSF samples with low bacterial loads by defining and addressing potential sources of error (including the contributions of reagent contamination, variable DNA extraction efficiency and bias, and batch effect); and (2) perform refined 16S HTS on CSF samples obtained from children diagnosed with bacterial meningitis and then compare the results with those from microbiological cultures. We hypothesized that, unlike with CSF shunt infection, 16S HTS controlling for contaminants and bias would identify relatively higher abundances of bacteria in the CSF from children with bacterial meningitis. We also hypothesized that the

relative abundances of known pathogenic organisms identified by 16S HTS would be higher than those of organisms not known to be CNS pathogens in meningitis samples.

## Methods

### Objective 1 (Optimize the performance of 16S HTS in CSF samples with low bacterial loads)

We used laboratory techniques to control systematic error associated with reagent contamination, DNA extraction efficiency and bias, and batch effects, comparing three such bench pipelines when extracting an experimentally-constructed suspension of bacteria commonly infecting CSF (a “mock community”). Two computational methods were then used to remove potential contaminant sequences from the resulting 16S HTS results. The refined approach was then applied to bacterial meningitis samples. (Figure 1) More complete details of each of these approaches are provided as follows.

### Mock Community

The mock community was created by mixing equivalent cell numbers of type strains of bacteria known to be frequently isolated from CSF shunt infections.[10] (Table 1), yielding a total final concentration of  $6.00 \times 10^7$  CFU/ $\mu$ l. Serial 1:10 dilutions were made of this mock community, ranging from  $6.00 \times 10^7$  CFU/ $\mu$ l to  $6.00 \times 10^4$  CFU/ $\mu$ l.

Table 1  
Bacterial species and genera present in the  
CSF mock community

Bacterial species and genera	CFU/ $\mu$ l
<i>Acinetobacter baumannii</i>	5.00E+06
<i>Corynebacterium sp.</i>	5.00E+06
<i>Cutibacterium acnes</i>	5.00E+06
<i>Enterococcus faecalis</i>	5.00E+06
<i>Escherichia coli</i>	5.00E+06
<i>Klebsiella pneumoniae</i>	5.00E+06
<i>Proteus mirabilis</i>	5.00E+06
<i>Pseudomonas aeruginosa</i>	5.00E+06
<i>Staphylococcus aureus</i>	5.00E+06
<i>Staphylococcus epidermidis</i>	5.00E+06
<i>Streptococcus pyogenes</i>	5.00E+06
<i>Streptococcus salivarius</i>	5.00E+06
Total CFU/ $\mu$ l	6.00E+07

We used this mock community to optimize and compare our sample processing and sequencing procedures prior to analyzing patient CSF samples and to identify and address biases introduced during DNA extraction, library construction, and taxonomic assignment. A total of 36 mock community samples (including dilutions) were included in our experiments comparing results of three different DNA extraction protocols (**Table 2**).

## Control Samples

In addition to mock communities, each mock community analysis included several control samples to identify and quantify bacterial DNA contamination in DNA extraction kits and PCR reagents (**reagent contamination**). These controls included (1) samples in which the processing reagents were included, with no purposely input (“template”) DNA, from DNA extraction and before the PCR step required to construct sequencing libraries (“no-sample extraction control”), and (2) other samples with only the reagents used during and after PCR and library construction (“no-template library PCR control”). In both types of controls, water was used to make up for the sample volumes of CSF and mock community extraction experiments. A total of 35 no-template extraction controls were included in DNA extraction experiments, and 18 no-template controls were included in library experiments. (**Table 2**). The PCR reaction for a group was repeated if the no-template control yielded a visible EtBr-stained band on a 1%

agarose gel. Otherwise, CSF sample amplicons and no-template controls were analyzed using identical protocols.

<b>Table 2: Number of control samples used for each DNA extraction type and for library PCR</b>					
Controls	Purpose	DNA Extraction Method			Total
		mag mini kit (MAG)	BiOstic Bacteremia DNA isolation Kit (BNC)	BiOstic Bacteremia DNA isolation Kit with carrier RNA (BWC)	
Mock community	Identification of possible biases introduced during DNA extraction, library construction, and taxonomic assignment	8 <sup>1</sup>	16 <sup>1</sup>	12 <sup>1</sup>	36
No-sample DNA extraction controls	Identification of potential contaminants from kit components	11 <sup>2</sup>	13 <sup>2</sup>	11 <sup>2</sup>	35
Sample No-template library PCR controls	Identification of potential contaminants during generation of 16S rRNA libraries	3	3	3	9 <sup>3</sup>
Mock Community No-template library PCR controls	Identification of potential contaminants during generation of 16S rRNA libraries				4 <sup>3</sup>
Extraction control No-template library PCR controls	Identification of potential contaminants during generation of 16S rRNA libraries				5 <sup>3</sup>
Sequencing negative control	Identification of mis-assigned reads during sequencing analysis				1
Total					90
<sup>1</sup> Number of mock community controls per method varied depending on number of kits used to extract the sample set, for required technical replicates, and to address day-to-day variability					
<sup>2</sup> Number of no-sample DNA extraction controls per method depended on number of kits used to extract the sample set and required technical replicates, which differed between methods					
<sup>3</sup> Total number of no-template library PCR control includes three such controls performed for each DNA extraction method (nine in total), four for the mock community analysis, five for the no-sample DNA extraction controls.					

# Dna Extraction

To address methodologic variation in **DNA extraction efficiency and bias**, DNA was extracted and purified from all samples using three extraction approaches: the AGOWA mag Mini DNA isolation kit (AGOWA, LGC Genomics, Berlin, Germany), hereafter MAG; the BiOstic Bacteremia DNA isolation Kit (Qiagen), hereafter BNC; and BNC with the inclusion of carrier RNA (Qiagen) to increase DNA yield by preventing binding to plastic in the kits[11], hereafter BWC. Components of each kit were aliquoted before extraction in an AirClean® Systems PCR Workstation (USA Scientific) decontaminated with LookOut® DNA Erase (Sigma-Aldrich), according to the manufacturer's instructions, followed by 15 min of UV-ray exposure to minimize environmental contamination.

MAG extractions were performed as follows without carrier RNA due to manufacturer concerns that this RNA could displace sample DNA during extraction (personal communication). A 100 µl volume of each sample was aliquoted into a sterile low binding microfuge tube (Axygen, Catalogue Number (CN): 31104081), to which 20 µl of 20mg/mL Proteinase K (Invitrogen, CN: 25530-049) was added. The mixture was vortexed for 20 seconds and incubated at 55°C for 10 minutes. After incubation 250 µl of Lysis buffer was added to the tube and vortexed gently for 15 seconds. The mixture was transferred to a clean 2 ml tube (Sarstedt, CN: 72.693.005) containing 0.3 g of 0.1mm zirconia/silica beads (Biospec Products Bartlesville, OK, USA, (Biospec) CN: 11079101z). Using a Mini-Beadbeater-16 (Biospec, CN: 607) the sample was mechanically disrupted by bead-beating for 3 minutes, followed by a 1-minute incubation on ice and a further 3 minutes of bead-beating. The sample was centrifuged at 4,000 rpm for 10 minutes. The resulting supernatant was transferred to a new low binding microfuge tube. To this, 325 µl of Binding buffer and 10 µl of mag particle suspension (mag-particles) were added, vortexed for 15 seconds, and incubated at room temperature for 30 minutes with continuous mixing on a VWR Tube Rotator (VWR, CN: 10136-084). After the incubation step DNA extraction proceeded according to the manufacturer's instructions as described.[12]

In both BiOstic extraction methods, 100 µl of sample was mixed by gentle vortexing either with or without (depending on the method) 1 µl of added carrier RNA at a stock concentration of 1 µg/µl of RNA. DNA was then extracted from each sample according to the manufacturer's instructions and as described.[12]

## Batch Effects

To control for **batch effects**, all CSF samples were randomized using a random number generator program.[13] Extractions were performed by research staff blinded to the sample identification key.

## Bacterial Quantification

A quantitative PCR (qPCR) assay targeting the 16S rRNA genes was used to measure the total bacterial load in each CSF sample as described earlier.[12]

## Bacterial 16S rRNA gene amplification, sequencing and analyses (16S HTS)

Amplicon library construction was carried out using a one-step PCR amplification targeting the 16S rRNA gene V4 region. The amplicon library primers [14] each contained the specific Illumina adapters, an 8-bp index sequence to allow for multiplex sequencing of the samples, and the 16S rRNA gene specific primer. [15] Library construction, pooling and sequencing proceeded as described for 600 cycles on an Illumina MiSeq desktop sequencer using the Miseq Reagent Kit v3.[12]

## 16s Hts Analysis

Sequencing data were processed using the denoising program DADA2 [16] (version 1.6.0) as described, [17] and aligned to the SILVA 16S reference database (v. 132)[18] to produce a 16S amplicon taxa table for downstream computational analysis. All samples, regardless of sample type or extraction protocol, were run on the same Illumina flow cell to reduce inter-run variation.

## Post-sequencing Contaminant Removal

Two analytic strategies were used to identify and remove contaminant sequences. Using the decontam R package installed from GitHub (<https://github.com/benjjneb/decontam>), [19] contaminant sequences were identified as those *either* with decreasing abundance at higher concentrations (threshold  $p < 0.10$ ) *or* that were more prevalent in kit control samples than in CSF samples (threshold  $p < 0.50$ ). Additionally, all sequences detected in extraction kit controls were considered contaminants and were removed from the CSF samples. Results from each computational approach was compared with the known mock community composition for objective 1 and microbiological culture results for objective 2.

## Objective 2 (Perform refined 16S HTS on CSF samples from bacterial meningitis)

### Study Subjects

Children  $\leq 18$  years old undergoing treatment for conventional culture-confirmed meningitis at St. Louis Children's Hospital in St. Louis, Missouri were eligible for enrollment in this study. Enrollment occurred from 2010 to present. Meningitis was defined as identification of organisms on microbiological culture of CSF fluid obtained from a lumbar puncture in a child without an existing neurosurgical device. All study subjects' CSF underwent routine microbiological processing and testing in the St. Louis Children's Hospital Microbiology Laboratory. For this study, we examined CSF obtained from the subset of 40 children whose conventional microbiological cultures recovered bacterial organisms and for whom there was 300  $\mu$ l of banked CSF available. The study received Institutional Review Board approval from the Seattle Children's Research Institute, St. Louis Children's Hospital and Children's Hospital Los Angeles.

## Clinical Data

For each child we collected culture information, specifically organism recovered in CSF and blood if applicable.

## Csf Specimen Collection

Sterile conditions were standard practice throughout recovery and storage of CSF. CSF samples were obtained from lumbar puncture. The CSF sample analyzed in this study either was left over from this first diagnostic sample or was obtained under sterile conditions. After being collected, CSF samples were stored at 4°C for up to 5 days. CSF was then aliquoted into vials of ~1000 µl for the study and stored at -70°C. After identification for this study, samples were shipped overnight on dry ice for analysis.

## Conventional Microbiological Culture Identification Of Bacteria

All CSF samples were tested using routine aerobic culture techniques in hospital-certified laboratories at SLCH. Conventional cultures are the traditional diagnostic approach used to detect typical pathogens in infectious diseases and were performed in a clinical microbiology laboratory following Clinical and Laboratory Standards Institute guidelines; however, conventional culture approaches do not detect all bacteria present in human disease.[20, 21] During analysis, the laboratory team remained blinded to the samples' culture results.

## Data Analyses

Sequence incidence was calculated for each sample as the ratio of sequence reads divided by total reads. While sequences have been linked to genera to aid in clinical interpretation, multiple sequences may be associated with the same organism or with unique variants; therefore, it is unclear whether adding read counts from unique sequences associated with common genus provides an accurate estimate of the incidence of that genus. For visual clarity, sequences representing less than 10% of reads across all samples, extraction methods and post-processing steps were aggregated to a single 'other' category in figures.

## Results

A bench and computational pipeline to define and address contributions of contaminants and bias to 16S rRNA HTS for pediatric CSF samples

As described above and shown in Figure 1, we used several laboratory methodological techniques to address and control for potential sources of systematic error in defining the microbiota of samples with low bacterial load, specifically due to reagent contamination and DNA extraction efficiency and bias, and batch effects.

To quantify bias introduced by reagent contamination as well as extraction techniques, we compared the final microbiota profiles yielded by three different sample DNA extraction methods applied to an artificially-constructed mixture of cells of bacterial taxa frequently isolated from CSF shunt infections. (Table 1). This mock community then served as a useful control not only for comparing the results of different DNA extraction methods, but other potential sources of error and bias as noted below. To characterize the contribution of **reagent contamination** to the HTS-based microbiota profiles of a defined mock community, we generated Figure 2, a principal coordinates analysis (PCoA) displaying differences in overall sequencing-defined microbiota profiles for the mock community yielded by the three extraction methods, and for the two controls. Mock community profiles were qualitatively similar regardless of extraction technique but different from those of controls, which reflected reagent contaminants during DNA extraction—"no sample extraction control"—and PCR—"no-template PCR control"—respectively. In addition, the profiles of the two control types differed significantly from each other, as expected given these control types represent distinct sources of potential contamination. Therefore, extraction method was not a significant source of bias for this analysis, and the contribution of reagent contaminants to the 16S HTS-based microbiota profiles of a defined mock community was minimal.

To characterize the contribution of **DNA extraction efficiency and bias**, we extracted DNA from this mock community using two different commercially-available extraction kits, the AGOWA mag Mini DNA isolation kit (MAG) and BiOstic Bacteremia DNA isolation Kit either without (BNC) or with (BWC) carrier RNA, which can increase DNA extraction yield for some kits[11]. (Figure 3) We then compared extraction efficiency and the sequencing-based relative abundance of mock community members among these three processing methods, relative to the known composition of the mock community, within four different dilutions of the mock community to determine the effects of bacterial load (0 to 1000-fold), and including controls to identify the bacterial DNA contaminating extraction and PCR reagents

As shown in Figure 3, these three methods, followed by 16S HTS, generated overall similar microbiota profiles that reflected the mock community composition to varying degrees, but never perfectly, indicating the bias introduced by sequencing library construction. For example, the relative abundance of *Acinetobacter* in sequencing results averaged 23%, with an input relative abundance of 8%, while the results for *Pseudomonas* were 3% and 8%, respectively. Given the overall similarity in results from the three DNA extraction methods (Figure 3), the bias is likely to be attributable to that known to result from the primers used to amplify a portion of the 16S rRNA gene, V4, which is used commonly for this type of analysis).[22] Notably, this primer set and gene variable region are often used for this type of analysis because they are generally capable of genus-level taxonomic identification of bacteria, as was the case here for all mock community members other than *Klebsiella* (identified at the family level—"Enterobacteriaceae", and verified by higher-resolution sequencing of this species).

Based on the similarity of the mock community composition before and after processing, we turned our attention to the computational management of contaminant sequences, comparing post-processing using decontam R analysis (Figure 4, column b) and computational removal of all sequences identified in concurrent no-sample extraction controls (Figure 4, column c). We recommend the use of any of the three extraction methods and post-processing using decontam R to maximize DNA yield and minimize bias (Figure 4, columns a and b). Full removal led to exclusion of sequences that were known to be in the mock community, and decontam R qualitatively improved concordance with the mock community composition.

## Refined 16S HTS on CSF samples obtained from children diagnosed with bacterial meningitis (Objective 2)

To further investigate the utility of computational filtering of contaminant reads, we then applied all three DNA extraction methods and identical post-sequence analyses to the CSF samples from children diagnosed with meningitis. We analyzed CSF collected from 22 pediatric meningitis episodes were analyzed. Among the study population, 59% (n=13) were female, with a median age of 0.6 months [interquartile range (IQR) 0.2-0.9] and median gestational age of 39 weeks (IQR 35-40).

Before sequencing, we defined the number of bacterial genome equivalents (GE) recovered from all clinical CSF samples and controls after all three DNA extraction methods, using broad-range 16S rRNA gene quantitative PCR (qPCR, Figure 5). We found that the amount of bacterial DNA extracted from clinical samples generally did not exceed that from negative controls regardless of processing method. These results suggested that the bacterial load in the CSF samples was relatively low compared with many clinical and environmental samples,[5] raising the possibility that our approach (and perhaps any current sequencing-based approach) may not be able to discriminate true pathogens from reagent contaminants.

Sequencing resulted in a total of 4884 unique V4 sequences, of which 249 were identified as likely contaminants and removed. As shown in Figure 6, while this pipeline did often identify reads of the genera cultured from infection samples at the time of diagnosis, this approach did not reliably indicate the dominance, or even presence, of those cultured organisms. In only 3 out of 22 samples was the cultured genus identified as the dominant organism by 16S HTS (23I - *Staphylococcus*, 18T and 45X - *Streptococcus*). In 4 samples the cultured organism was not identified by 16S HTS at all. Therefore, we concluded that, even with computational filtering of likely or definite contaminant reads, refined 16S HTS was unable to consistently identify infecting pathogens in CSF samples from children with meningitis.

## Discussion

This study sought to optimize the performance of 16S HTS in CSF samples with low bacterial loads by defining and addressing potential sources of error (including the contributions of reagent contamination, variable DNA extraction efficiency and bias, and batch effect). We demonstrated optimal similarity of

sequencing-based mock community microbiota profiles and maximization of DNA yield, albeit with some previously-observed efficiency bias, with all three extraction techniques and post-processing using decontam R at the low bacterial loads observed in CSF samples. We then applied these methods to CSF samples from children with bacterial meningitis. We demonstrated that, even with a rigorously-controlled sample processing and 16S HTS sequencing approach, augmented by computational removal of contaminant sequencing signal, we were unable to reliably identify infecting organisms identified by culture in CSF samples with low bacterial loads from children with clinically-diagnosed meningitis. Due to the low abundance of organisms in these samples relative to that of DNA in most commercially-available processing reagents, the "signal" from infecting bacteria could not be distinguished from the "background noise" of contaminants. Therefore, without major methodologic advances, culture currently remains the best approach to identifying pathogens infecting in this clinical context. Our hypothesis that, unlike CSF shunt infection, there would be high levels of bacteria in bacterial meningitis samples was disproven.

16S HTS and other molecular microbiological techniques have proven to both complement and, in some cases, show promise for replacing culture in other clinical scenarios, usually where bacterial abundances are much higher than that of contaminants. For example, sequencing-based microbiome methods are in common use for studying the microbiota of the gastrointestinal tract, vagina, mouth, and other exposed mucosal surfaces. In addition, culture is frequently unable to identify pathogenic organisms in patients with meningitis, despite the clear purulence of the resulting samples.[23] Therefore, there is great interest in using culture-independent, DNA-based methods for pathogen identification in CSF samples.[23, 24] In this study, we found that the nearly consistent low bacterial abundances relative to the contamination levels of reagents greatly diminishes the utility of molecular approaches, even with the inclusion of rigorous controls and computational removal of contaminant sequencing reads, regardless of how the samples are processed or extracted.

A panel of experts previously performed a narrative review to describe the desirable characteristics for a molecular diagnostic assay for bacterial meningitis.[25] The authors concluded such an assay would detect multiple pathogens in a cost-effective, easy to use system that provides rapid and robust results. Unfortunately, our analysis indicates that, absent major methodologic advances, such an 16S HTS assay does not currently exist. However, current molecular methods are already in common use for characterizing higher-abundance samples, such as purulent respiratory infections,[26] gastrointestinal infections,[27] and even normally sterile fluids such as blood.[28] In contrast, infections with relatively low abundances of microbes in the resulting samples pose considerable challenges to these methods. As numerous authors have reviewed,[4, 6, 29–35] the uncertainty and error introduced by reagents, human error, and methodologic bias renders current methodology concerningly inaccurate, in some cases leading to controversial or variable study results. Indeed, investigators have reached conclusions similar to those here for studying nasopharyngeal samples and induced sputum, which can also exhibit low bacterial loads.[36]

Our results contrast with prior studies by others investigating the utility of 16S HTS for diagnosing or identifying pathogens in children and adults with meningitis.[31–35, 37, 38] While comprehensively

contrasting our approach with these others is beyond the scope of this manuscript, two of these studies defined detection of specific pathogens as exceeding a limit of 4 reads in the case of *Streptococcus pneumoniae*[33] and 19 reads for any pathogen,[34] respectively, raising questions regarding whether these limits exceeded the background signal from reagents. In general, these studies did not quantify the effects of sample extraction or reagent contaminants, precluding direct comparison with our approaches here.

We previously used 16S HTS to identify bacterial DNA reads in CSF from children with shunt infections, [12] where the bacterial loads in infection samples often exceeded  $10^5$  CFU/mL. While we have refined the technique of 16S HTS for use in CSF and demonstrated its limited utility in bacterial meningitis, we believe this approach may be more useful for CSF shunt infection, where persistence of bacteria on hardware may increase sensitivity of molecular diagnostics relative to that for meningitis. Together with the results of the current study, these findings highlight the possible role of restricting molecular microbiological analysis to cases where loads exceed the limits of detection, which must be defined precisely for the specific analytical pipeline being used. We also propose that a screening qPCR for total bacterial load, performed concurrently with culture, could be a useful step prior to considering the utility of HTS for clinical diagnosis.

Our study had a number of limitations. For example, our CSF sample set originated from a single geographic area and reflected a limited sample size (n=22), restricting generalizability. In addition, the volume of CSF available for our analyses was relatively low—100  $\mu$ L—which could conceivably have limited our sensitivity. Other DNA extraction methods could have improved sensitivity, an issue we attempted to address by comparing 3 different such methods. While we constructed a mock community to optimize conditions, with hindsight it should have been diluted further approach the low bacterial load of CSF. Despite these limitations, we believe we have comprehensively tested conditions for 16S HTS testing of low bacterial load clinical samples such as CSF.

In conclusion, we demonstrated optimal similarity in microbiota profiles yielded by sequencing of a mock community and maximization of DNA yield, albeit with some previously observed efficiency bias, after three extraction techniques and post-processing using decontam R under the low bacterial load conditions seen in CSF samples. We found that the limits of detection imposed by reagents and methodologic bias precluded the accurate detection of infecting bacteria in children with culture-confirmed meningitis using 16S HTS despite rigorous controls and sophisticated computational approaches. Although we did not find current DNA-based diagnostics to be useful for meningitis, the utility of these methods CSF shunt infection remains undefined. Future advances in sample processing methods to minimize or eliminate contamination will be required to improve the sensitivity and specificity of these methods for pediatric meningitis.

## Declarations

**Ethics approval and consent to participate:** The study received Institutional Review Board approval from the Seattle Children’s Research Institute, St. Louis Children’s Hospital and Children’s Hospital Los Angeles.

**Consent for publication:** Not applicable

**Availability of data and material:** The datasets generated and/or analyzed during the current study are available in the NCBI Sequence Read Archive repository, <https://www.ncbi.nlm.nih.gov/sra> as “PRJNA768849 : A refined, controlled 16S rRNA gene sequencing approach reveals limited detection of cerebrospinal fluid microbiota in children with bacterial meningitis”.

**Competing interests:** The authors declare that they have no competing interests

**Funding:** This study was supported by R01 NS095979. Dr. Limbrick receives research funding and equipment for unrelated projects through Microbot Medical, Inc. and Medtronic, Inc. Dr. Hoffman receives additional support from the National Institutes of Health via K24 HL141669.

**Authors' contributions:** CP, KW, PH, TS and LH conceived and designed the study. DL, PM, JH, and TS provided oversight over study enrollment and material collection. CP performed the experiments. CP and KW analyzed the data. CP, LH, and TS wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

**Acknowledgements:** Current membership of the CLIMB group includes: Alexander Cheong, Gabriel Haller, Julie McGalliard, Diego Morales, Amanda Morgan, Alexander Rangel-Humphrey, and Lisa Wick. Past membership includes: Dan Berger, Whitney Bond, Haley Botteron, Courtney Dethlefs, Jessica Foster, Mohammed Gabir, Robert Johnson, Deanna Mercer, and Linda Shih. We would like to thank the children and families who participated in this CLIMB study at St. Louis Children’s Hospital. We would also like to thank the children and families who have and are enrolling in CLIMB at participating centers, as well as the pediatric neurosurgeons, neurosurgical staff, and laboratory study staff who support enrollment of patients and ensure appropriate collection and storage of CSF specimens.

## References

1. Simon, T.D., et al., *Cerebrospinal Fluid Shunt Infection: Emerging Paradigms in Pathogenesis that Affect Prevention and Treatment*. J Pediatr, 2018.
2. Simon, T.D., et al., *Evaluation of microbial bacterial and fungal diversity in cerebrospinal fluid shunt infection*. PLoS One, 2014.
3. Simon, T.D., et al., *Use of quantitative 16S rRNA PCR to determine bacterial load does not augment conventional cerebrospinal fluid (CSF) cultures among children undergoing treatment for CSF shunt infection*. Diagn Microbiol Infect Dis, 2013.
4. Blaser, M.J., et al., *Lessons learned from the prenatal microbiome controversy*. Microbiome, 2021. **9**(1): p. 8.

5. Marsh, R.L., et al., *How low can we go? The implications of low bacterial load in respiratory microbiota studies*. *Pneumonia* (Nathan), 2018. **10**: p. 7.
6. Salter, S.J., et al., *Reagent and laboratory contamination can critically impact sequence-based microbiome analyses*. *BMC Biol*, 2014. **12**: p. 87.
7. Zhao, J., et al., *Impact of enhanced Staphylococcus DNA extraction on microbial community measures in cystic fibrosis sputum*. *PLoS One*, 2012. **7**(3): p. e33127.
8. Leek, J.T., et al., *Tackling the widespread and critical impact of batch effects in high-throughput data*. *Nat Rev Genet*, 2010. **11**(10): p. 733-9.
9. Randall, D.W., et al., *Batch effect exerts a bigger influence on the rat urinary metabolome and gut microbiota than uraemia: a cautionary tale*. *Microbiome*, 2019. **7**(1): p. 127.
10. Simon, T.D., et al., *Patient and Treatment Characteristics by Infecting Organism in Cerebrospinal Fluid Shunt Infection*. *J Pediatric Infect Dis Soc*, 2018.
11. Shaw, K.J., et al., *The use of carrier RNA to enhance DNA extraction from microfluidic-based silica monoliths*. *Anal Chim Acta*, 2009. **652**(1-2): p. 231-3.
12. Whitlock, K.B., et al., *Characterization of cerebrospinal fluid (CSF) microbiota from patients with CSF shunt infection and reinfection using high throughput sequencing of 16S ribosomal RNA genes*. *PLoS One*, 2021. **16**(1): p. e0244643.
13. Team, R.C., *R: A language and environment for statistical computing*. 2013, R Foundation for Statistical Computing, Vienna, Austria.
14. Kozich, J.J., et al., *Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform*. *Appl Environ Microbiol*, 2013. **79**(17): p. 5112-20.
15. Thompson, L.R., et al., *A communal catalogue reveals Earth's multiscale microbial diversity*. *Nature*, 2017. **551**(7681): p. 457-463.
16. Wang, Q., et al., *Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy*. *Appl Environ Microbiol*, 2007. **73**(16): p. 5261-7.
17. Nelson, M.T., et al., *Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles*. *Cell Rep*, 2019. **26**(8): p. 2227-2240 e5.
18. Pruesse, E., et al., *SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB*. *Nucleic Acids Res*, 2007. **35**(21): p. 7188-96.
19. Davis, N.M., et al., *Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data*. *Microbiome*, 2018. **6**(1): p. 226.
20. Falkow, S., *Molecular Koch's postulates applied to bacterial pathogenicity—a personal recollection 15 years later*. *Nat Rev Microbiol*, 2004. **2**(1): p. 67-72.
21. Rhoads, D.D., et al., *Comparison of culture and molecular identification of bacteria in chronic wounds*. *Int J Mol Sci*, 2012. **13**(3): p. 2535-50.

22. Soriano-Lerma, A., et al., *Influence of 16S rRNA target region on the outcome of microbiome studies in soil and saliva samples*. Sci Rep, 2020. **10**(1): p. 13637.
23. Men, X., et al., *Pathogen identification in culture-negative cerebrospinal fluid specimens of patients with purulent meningitis using next-generation sequencing technology*. Int J Clin Exp Pathol, 2020. **13**(9): p. 2427-2438.
24. Perdigao Neto, L.V., et al., *Polymerase chain reaction targeting 16S ribosomal RNA for the diagnosis of bacterial meningitis after neurosurgery*. Clinics (Sao Paulo), 2021. **76**: p. e2284.
25. Diallo, K., et al., *Molecular diagnostic assays for the detection of common bacterial meningitis pathogens: A narrative review*. EBioMedicine, 2021. **65**: p. 103274.
26. Rand, K.H., et al., *Relationship of Multiplex Molecular Pneumonia Panel Results With Hospital Outcomes and Clinical Variables*. Open Forum Infect Dis, 2021. **8**(8): p. ofab368.
27. Yoo, I.H., et al., *Quality Improvements in Management of Children with Acute Diarrhea Using a Multiplex-PCR-Based Gastrointestinal Pathogen Panel*. Diagnostics (Basel), 2021. **11**(7).
28. Opota, O., K. Jaton, and G. Greub, *Microbial diagnosis of bloodstream infection: towards molecular diagnosis directly from blood*. Clin Microbiol Infect, 2015. **21**(4): p. 323-31.
29. Glassing, A., et al., *Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples*. Gut Pathog, 2016. **8**: p. 24.
30. Lauder, A.P., et al., *Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota*. Microbiome, 2016. **4**(1): p. 29.
31. Liu, A., et al., *High-throughput sequencing of 16S rDNA amplicons characterizes bacterial composition in cerebrospinal fluid samples from patients with purulent meningitis*. Drug Des Devel Ther, 2015. **9**: p. 4417-29.
32. Moon, J., et al., *Rapid diagnosis of bacterial meningitis by nanopore 16S amplicon sequencing: A pilot study*. Int J Med Microbiol, 2019. **309**(6): p. 151338.
33. Zhang, X.X., et al., *The diagnostic value of metagenomic next-generation sequencing for identifying Streptococcus pneumoniae in paediatric bacterial meningitis*. BMC Infect Dis, 2019. **19**(1): p. 495.
34. Nakagawa, S., et al., *Rapid sequencing-based diagnosis of infectious bacterial species from meningitis patients in Zambia*. Clin Transl Immunology, 2019. **8**(11): p. e01087.
35. Ji, X.C., et al., *Reduction of Human DNA Contamination in Clinical Cerebrospinal Fluid Specimens Improves the Sensitivity of Metagenomic Next-Generation Sequencing*. J Mol Neurosci, 2020. **70**(5): p. 659-666.
36. Claassen-Weitz, S., et al., *Optimizing 16S rRNA gene profile analysis from low biomass nasopharyngeal and induced sputum specimens*. BMC Microbiol, 2020. **20**(1): p. 113.
37. Xing, X.W., et al., *Metagenomic Next-Generation Sequencing for Diagnosis of Infectious Encephalitis and Meningitis: A Large, Prospective Case Series of 213 Patients*. Front Cell Infect Microbiol, 2020. **10**: p. 88.

## Figures

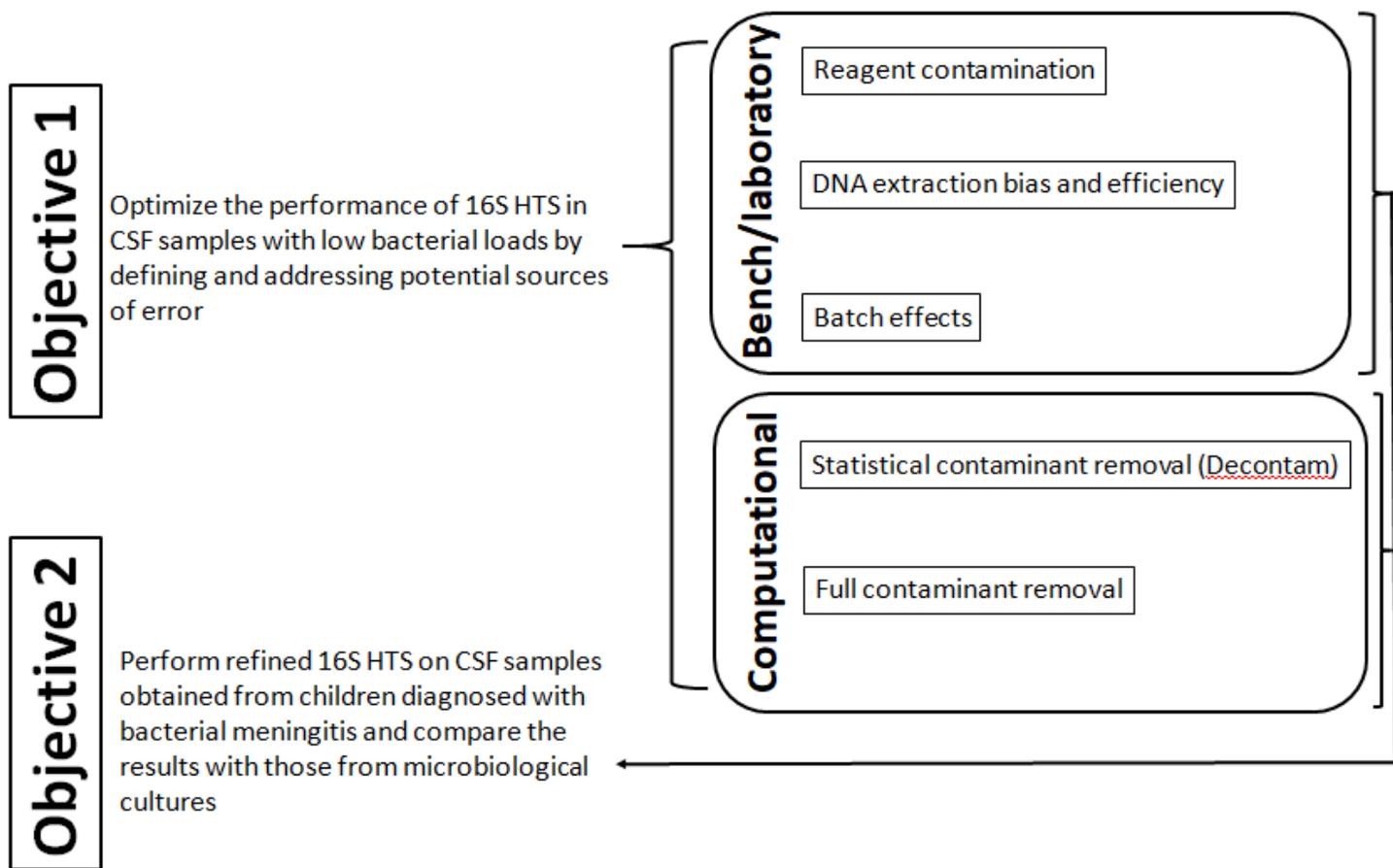
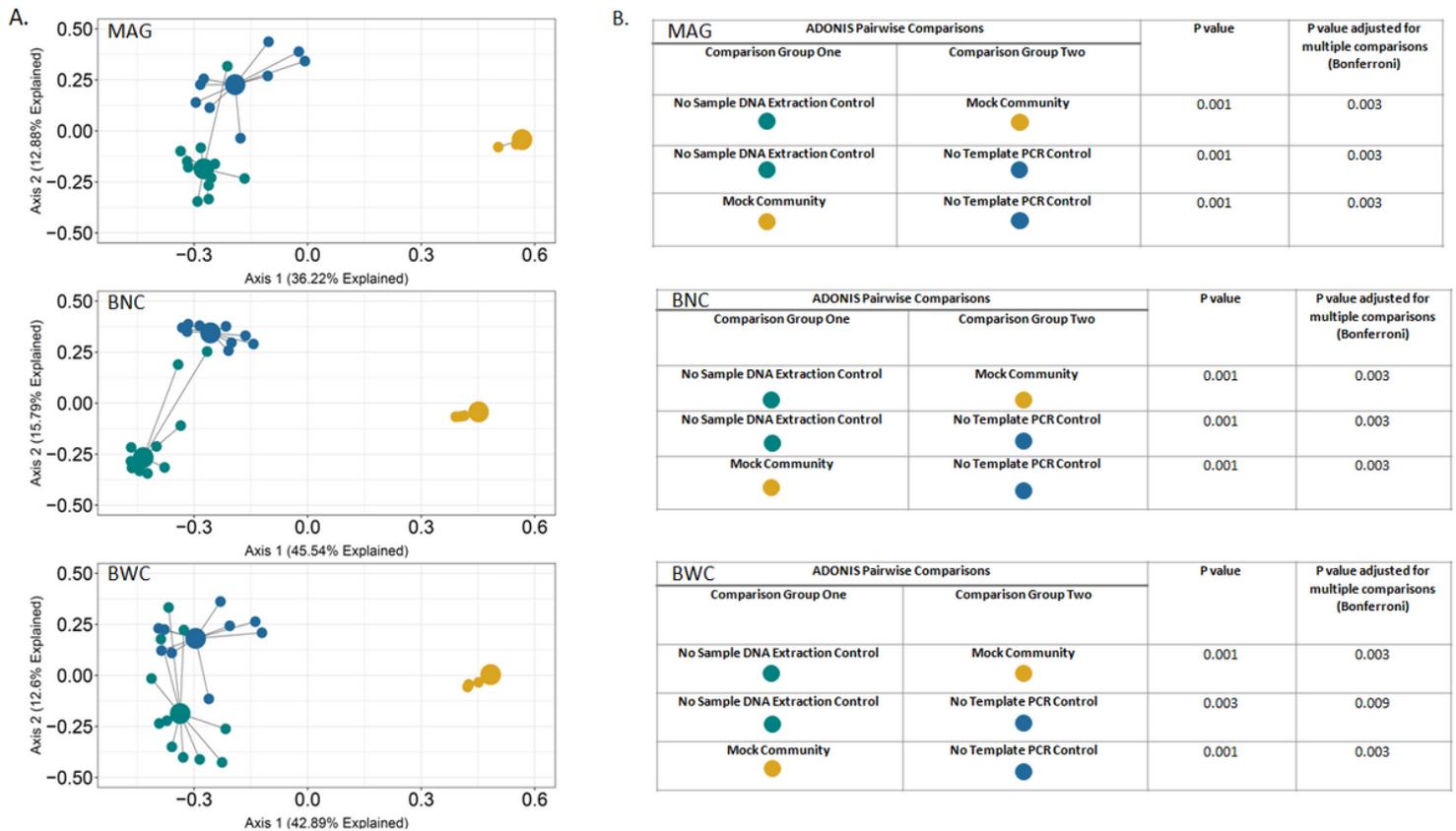


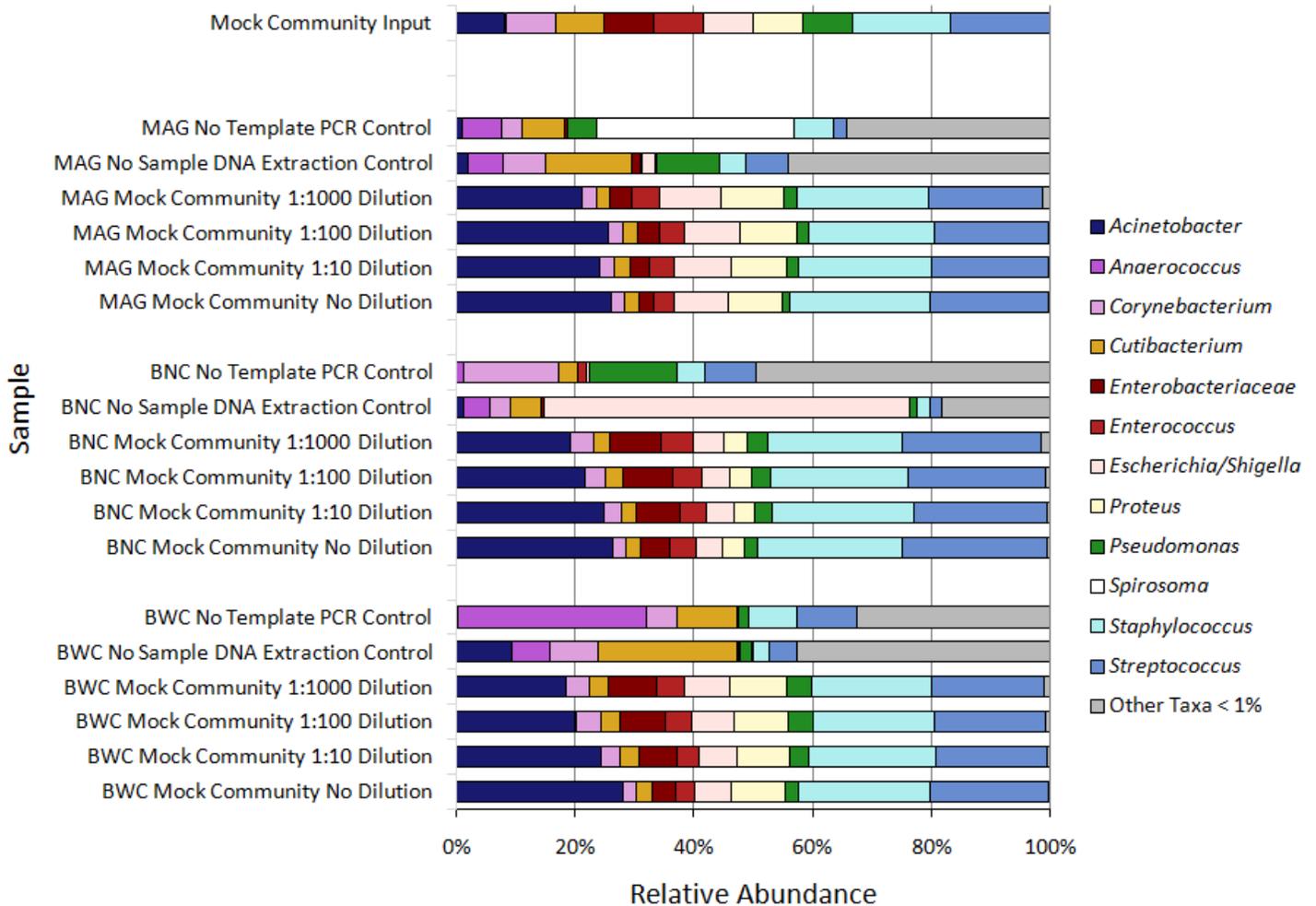
Figure 1

Schematic of study objectives. Bench and computational approaches were refined using mock communities and robust negative controls to define and address contributions of contaminants and bias to 16S rRNA HTS in Objective 1. These bench and computational pipelines were applied to CSF samples from children with meningitis in Objective 2.



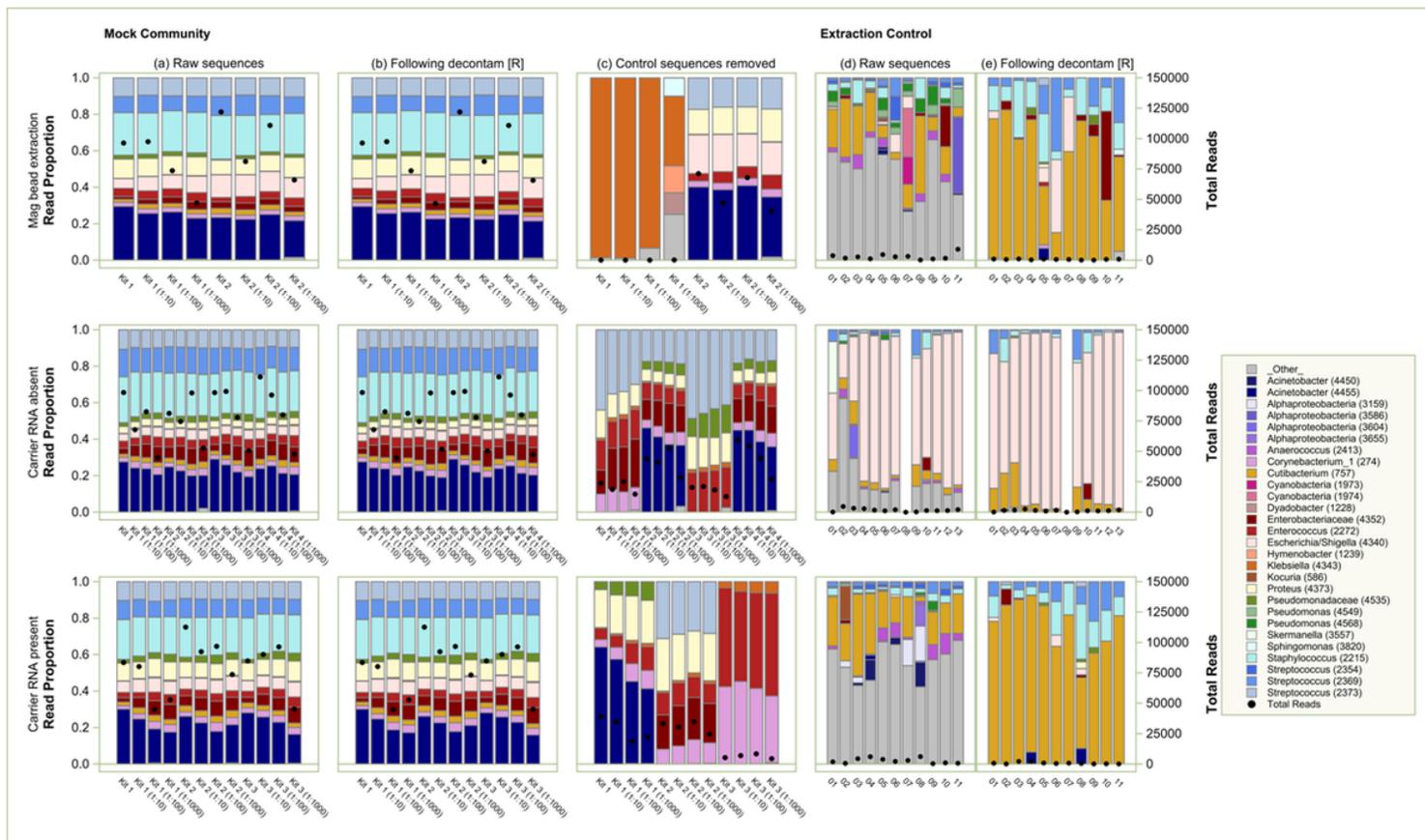
**Figure 2**

PCoA of mock community and negative control samples using 3 different DNA extraction methods A. PCoA plot of samples using Bray-Curtis distances based on sequence read composition. The PCoA describes relationships between taxonomic composition patterns of the individual samples and sample type groupings. Individual samples are represented by the small dots, colored by sample type. Large dots represent the centroid of each group, with lines connecting individual sample dots to their respective centroids. Centroids represent the averages of the taxonomic composition of each sample type. B. Data tables of ADONIS pairwise comparisons within DNA extraction method groupings. Comparisons were between mock communities and their no-sample extraction controls and no-template PCR controls.



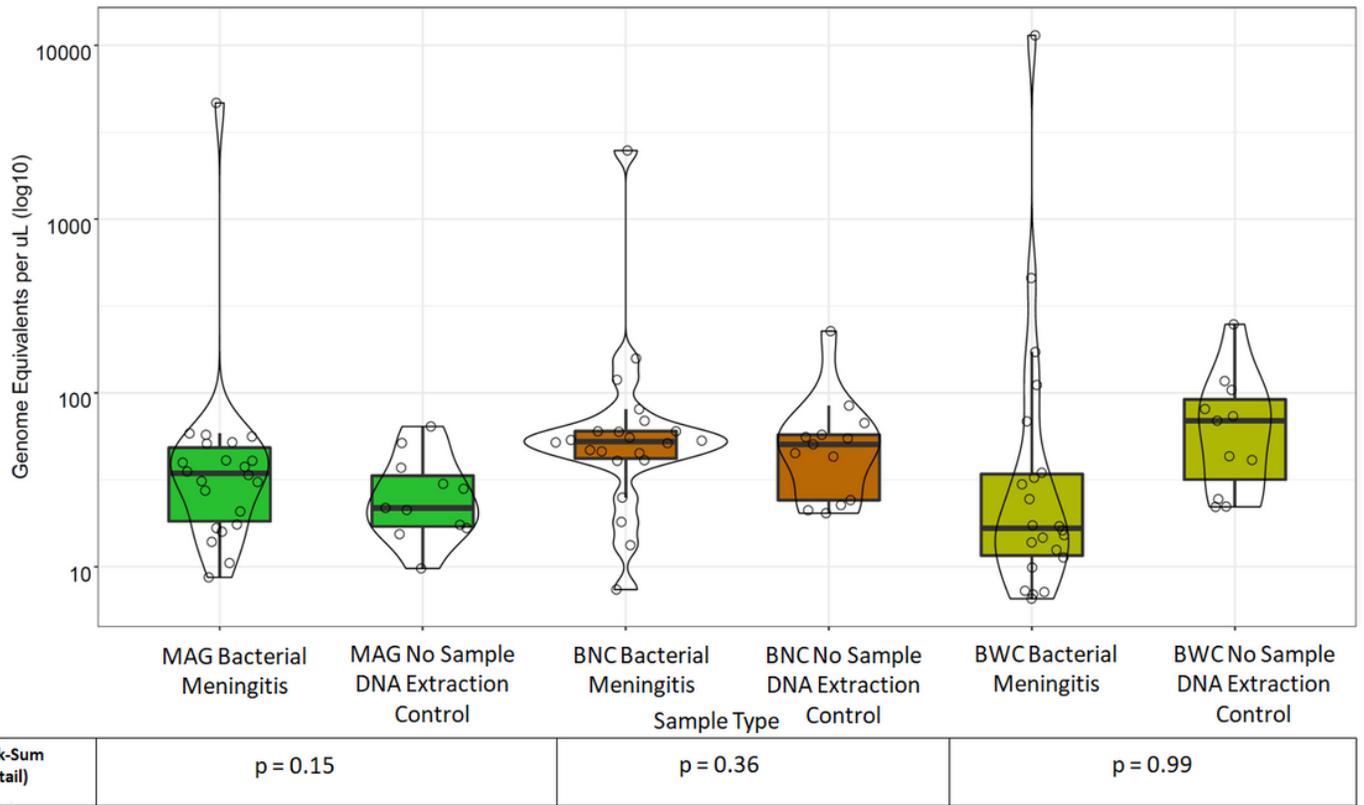
**Figure 3**

Mock community taxonomic profiles from V4 HTS are overall similar regardless of dilution or DNA extraction method. Shown are bar plots displaying the relative abundances of bacteria identified by sequencing of mock communities according to extraction method and dilution of the input mock community. “Mock community input”, the known composition of the original mock community. MAG, Agowa mag mini DNA isolation kit. BNC, BiOstic bacteremia DNA kit without carrier RNA. BWC, BNC with carrier RNA (which can increase DNA extraction yield). The DNA from mock community cells was extracted using MO BIO's BiOstic Bacteremia DNA Kit without carrier RNA (BNC), MO BIO's BiOstic Bacteremia DNA Kit with carrier RNA (BWC) and the AGOWA mag Mini DNA isolation kit (MAG). Sequence quality filter and analyses were carried out using the denoising program DADA2. Sequence taxonomy was classified through the SILVA 16S rRNA database.



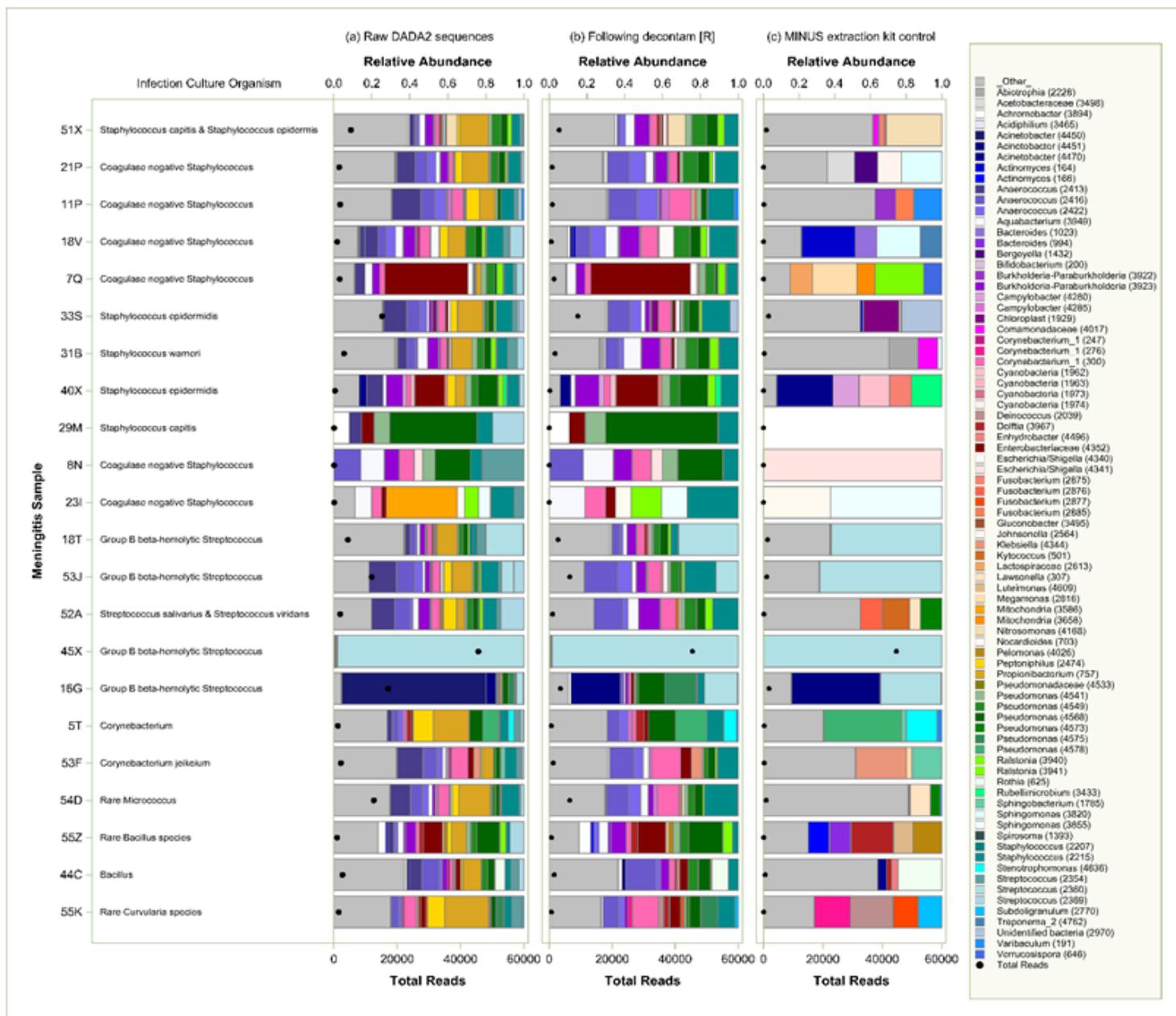
**Figure 4**

Effect of DNA extraction method and computational decontaminant removal on taxonomic composition of mock community and negative control samples. Stacked bar-chart results showing relative abundances of taxa after each extraction method and HTS, followed by contaminant sequence removal. Left, profiles shown were generated with (a) unfiltered sequence data; (b) those data after decontam R; (c) after computational removal of all sequences identified in concurrent no-sample extraction controls. Right, profiles from the controls themselves before (d) and after (e) decontam R. We recommend the application of the computational package decontam R to sequences derived from any of the three extraction techniques.



**Figure 5**

The bacterial DNA concentrations in CSF bacterial meningitis samples are comparable to bacterial DNA concentrations in no-template extraction controls. Box and violin plots of total bacterial abundance in CSF bacterial meningitis samples and their corresponding, concurrent extraction controls by 16S rRNA gene copies measured by qPCR, expressed as genome equivalents per ul (GE/ul).



**Figure 6**

16S HTS of CSF samples with from children diagnosed with meningitis, followed by computational filtering of contaminant reads, does not reliably identify culturable bacteria in those samples CSF samples from children with meningitis are identified by sample ID (left), where the organisms cultured from those samples at diagnosis are indicated. Relative abundances of taxa by genus in raw, unfiltered data using MAG (a) generally changed little with analysis by decontam R (b). While the further complete removal of reads in concurrent no-sample extraction (c) did occasionally enrich for the cultured genus (e.g., 53J), this approach more often did not reliably identify the cultured organism either at all (e.g., 51X, 21P, 11P, 18V, 7Q) or as a dominant constituent (e.g., 16G).