NORTHWESTERN UNIVERSITY

Advancing Understanding of Diversity and Antimicrobial Resistance in Indoor Microbiomes

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Civil and Environmental Engineering

By

Jiaxian Shen

EVANSTON, ILLINOIS

March 2023

Abstract

The importance of understanding indoor microbial exposure is increasingly recognized, particularly concerning the emergence of pathogens and antimicrobial resistance. With the advancements of sequencing technologies, our capability of exploring indoor microbial communities has dramatically increased. However, huge challenges remain to translate sequence-based knowledge to actionable interpretations to support human health and environmental sustainability. Specifically, we urgently need more effective, untargeted surveillance methods for environmental microbiomes in healthcare facilities. We need to conduct more and better monitoring of tolerance to antimicrobials that were long immune from concern (e.g., chlorhexidine) as it is emerging in hospital environments. We also need to expand the toolkit for evidence synthesis studies in the field of environmental microbiomes to fully unleash the potential of indoor microbiome research and to reveal more generalizable phenomena and mechanisms.

In this dissertation, I developed an improved metagenomics-based surveillance workflow with enhanced ability to handle low-biomass samples, distinguish viability, be quantitative, and estimate sequencing resources. The workflow not only can serve as a guidance for environmental surveillance practices but also enhances our understanding of surface microbiomes. I investigated chlorhexidine resistance in the hospital environment by combining controlled microcosm experiments and field surveys. My work uncovered the persistence patterns of chlorhexidine on indoor surfaces in response to disinfection and cleaning practices. I also found that chlorhexidine tolerance was widespread in a medical intensive care unit environment, and sinks were a dominant reservoir regarding chlorhexidine tolerance. This knowledge will help

improve clinical application of chlorhexidine. Additionally, I developed an R package, RefDeduR, that performs accurate and high-throughput reference deduplication for large datasets to promote broader applications of evidence synthesis research (e.g., in indoor microbiomes).

Taken together, this dissertation enhances our knowledge of diversity and antimicrobial resistance in indoor microbiomes as well as improves our capability for future exploration. The outcomes of this dissertation will not only facilitate future research in the field, but also inform clinical and epidemiological surveillance, disinfection and cleaning practices, and building design. Ultimately these will lead to better infection prevention and health promotion.

Acknowledgment

I would like to express my deepest appreciation to my Ph.D. adviser, Dr. Erica M. Hartmann, for her tremendous support over the past years. I could not have been the researcher and the person I am without her creative advice and considerate mentoring.

I would also like to extend my sincere thanks to my committee members and collaborators. Dr. Luisa Marcelino, Dr. Patrick C. Seed, Dr. George Wells, Dr. Mary K. Hayden, Dr. Vincent B. Young, Dr. Fangqiong Ling, and Dr. Curtis Huttenhower contributed significantly to my research and career development. I had the great pleasure of working with all previous and current Hartmann lab members, and peers in the department of Civil and Environmental Engineering, at Northwestern, and in our collaborators' teams. In particular, I would like to thank Dr. Alexander G. McFarland, Dr. Teppei Shimasaki, and Khaled Aboushaala for their help in collecting hospital environmental samples.

I must also thank all funding agencies for supporting my research, conference travel, and publishing. Particularly, the Centers for Disease Control and Prevention not only funded a significant part of my work, but also provided invaluable feedback on my research.

I'm extremely grateful to my family and friends for their continuous support. I cannot leave Northwestern without acknowledging my father, mother, and sister for encouraging me to pursue my dream throughout my life. My friends sparked research ideas, gave practical suggestions, and provided mental support during our conversations. I am especially grateful to Ruochen Jiao for always keeping me company. Without their support, I would not have been able to bravely face uncertainty and overcome challenges.

List of Abbreviations

Abbreviation	Explanation
MICU	Medical intensive care unit
PMA	Propidium monoazide
LoD	Limit of detection
PMA-Seq	PMA treatment coupled with 16s rRNA gene amplicon sequencing
qPCR	Quantitative PCR
MetaSeq	Metagenomic sequencing
PMA-MetaSeq	PMA treatment coupled with metagenomic sequencing
RUMC	Rush University Medical Center
PBST	Phosphate buffered saline with 0.02% tween 80
PBS	Phosphate buffered saline
TSA	Tryptic soy agar
TSAI	Tryptic soy agar supplemented with 4 mg/l itraconazole
R2A	Reasoner's 2A agar
BA	Blood agar
MALDI-TOF MS	Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
PCoA	Principal coordinate analysis
Nd	Nonpareil diversity
PTR	Peak-to-trough ratio
CHG	Chlorhexidine digluconate
MICs	Minimum inhibitory concentrations
MBCs	Minimum bactericidal concentrations
MHA	Mueller-Hinton Agar
QC	Quality control
BAM	Binary alignment map
SAM	Sequence alignment map
ORFs	Complete open reading frames
CARD	Comprehensive antibiotic resistance database
RGI	Resistance gene identifier
ARGs	Antibiotic resistance genes
RND	Resistance-nodulation-cell division
SMR	Small multidrug resistance
MFS	Major facilitator superfamily
ABC	ATP-binding cassette

Glossary of Terms

Antimicrobial: An antimicrobial is an agent that kills microorganisms or stops their growth. According to the microorganisms they act primarily against, antimicrobials can be classified as antibiotics which are used against bacteria, and antifungals which are used against fungi [1].

Antiseptic: Antiseptic is a germicide that is used on skin or living tissue for the purpose of inhibiting or destroying microorganisms. Examples include alcohols, chlorhexidine, chlorine, hexachlorophene, iodine, chloroxylenol (PCMX), quaternary ammonium compounds, triclosan, and benzalkonium chloride [2, 3].

Disinfectant: Disinfectant is a chemical agent used on inanimate objects (i.e., nonliving) (e.g., floors, walls, sinks) to destroy virtually all recognized pathogenic microorganisms, but not necessarily all microbial forms (e.g., bacterial endospores). The EPA groups disinfectants on whether the product label claims “limited,” “general” or “hospital” disinfectant [2]. Examples include triclosan and benzalkonium chloride.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Under the One Health framework, the importance of understanding indoor microbial exposure is increasingly recognized [4, 5]. This importance is due in part to the fact that people in modern society spend the majority of their time indoors [6, 7], along with the increase in the number of observed associations between diseases and exposure to indoor pathogens [8-11]. In addition to direct exposure to pathogens, antibiotic resistance is another critical issue associated with the indoor microbiome. Researchers have shown that antibiotic resistance is concentrated in the built environment relative to outdoors [12], and that built environment factors in the home can favor the recurrence of antibiotic-resistant infections [13]. To promote a healthy indoor microbiome, it is imperative to understand indoor microbial diversity, antimicrobial resistance, and factors determining these phenomena.

With next-generation sequencing technologies, numerous efforts have been made to characterize the indoor microbial communities, explore their ecological features, infer their public health impacts, and so forth (e.g., [14-18]). While these endeavors increase our understanding of indoor microbiomes, huge challenges remain to bridge the gap between deploying sequence-based knowledge and generating actionable interpretations to support human health and environmental sustainability.

1.1 More effective environmental surveillance methods are needed

Healthcare facilities are a high-priority indoor environment because their occupants are particularly vulnerable to infection. More effective surveillance of environmental microbiomes is urgently demanded in healthcare facilities, specifically concerning pathogens and the

development of antimicrobial resistance. Emergent evidence indicates synergy among pathogens and dissemination of antibiotic resistance in built environment [19, 20]. Metagenomics thus represents a promising surveillance technique because it is untargeted. Compared with targeted surveillance, untargeted methods may also identify novel or rapidly emerging pathogens and antibiotic resistance genes [21]. However, metagenomics-based techniques are still challenged by several limitations: 1) they are not powerful enough to extract valid signals out of the background noise for low biomass samples; 2) they do not distinguish between viable and non-viable organisms; and 3) they do not reveal the microbial load quantitatively. An additional practical challenge towards a robust pipeline is the inability to efficiently allocate sequencing resources a priori. Assessment of sequencing depth is generally practiced post hoc, if at all, for most microbiome studies, regardless of the sample type. This practice is inefficient at best, and at worst poor sequencing depth jeopardizes the interpretation of study results. These challenges are both conceptual and technical in nature. They arise from the diversity of metagenomics research objectives and are often exacerbated by intrinsic features of low-biomass environments that need to be monitored.

To address these challenges, I first discussed conceptual and technical perspectives toward accurate and robust environmental surveillance (**Chapter 2**). Viability assessment, taxonomic resolution, and quantitative metagenomics were covered regarding their current advancements, necessary precautions and directions for further development. Overall, I highlight the importance of building solid conceptual frameworks and identifying rational limits to facilitate the application of techniques. This perspective was published in *Frontiers in Genetics* under the title "Toward accurate and robust environmental surveillance using metagenomics" where I am the

first author. Under the guidance of this perspective, I then developed an improved workflow for metagenomics-based environmental surveillance (**Chapter 3**). Techniques were evaluated, optimized, and developed based on a representative microbiome sample created by aggregating 120 surface swabs collected from a medical intensive care unit. The developed workflow is appropriate for low-biomass samples, distinguishes viability, is quantitative, and estimates sequencing resources. The research in this chapter was published in *Microbiome* under the title "An improved workflow for accurate and robust healthcare environmental surveillance using metagenomics" where I am the first author.

1.2 Emerging tolerance to chlorhexidine in environmental bacteria merits more attention

Chlorhexidine has been used as an antiseptic for decades with little concern for resistance. However, more and more cases of reduced susceptibility to chlorhexidine have been reported recently, raising concerns of increased resistance. The concern is magnified considering the dissemination of antimicrobial resistance genes in microbial communities. The hospital environment is a known reservoir of nosocomial pathogens and antibiotic-resistant bacteria. It also has a high probability of containing sublethal concentrations of antimicrobials, and this may promote the development or enrichment of antimicrobial tolerance [22]. Moreover, researchers have traced hospital outbreaks to chlorhexidine-resistant, environment-sourced strains [23]. Despite that, previous monitoring efforts have been largely restricted to clinical isolates, and scrutiny on chlorhexidine resistance in the environment is very limited.

In **Chapter 4**, I investigated chlorhexidine resistance in the hospital environment by combining microcosm experiments and field surveys. Employing microcosms under controlled conditions, persistence patterns of chlorhexidine in response to different cleaning and disinfection practices on 3 surface materials were investigated. Survival of 4 clinically relevant bacterial strains was examined under 24 h chlorhexidine exposure on surfaces in the concentration range that chlorhexidine persisted. To determine how widespread chlorhexidine tolerance is *in situ*, I further conducted a field survey in a medical intensive care unit (MICU), revealing both phenotypic and genomic indicators of chlorhexidine tolerance. A manuscript describing this work, on which I will be first author, is in preparation.

1.3 Toolkit of evidence synthesis research needs to be expanded

The copious and diverse research efforts pertaining to indoor microbial communities advance our overall perception of the indoor microbiome and provide a large dataset that can be used to develop evidence synthesis studies (e.g., systematic reviews, meta-analyses). Evidence synthesis studies reveal more generalizable phenomena and underlying mechanisms compared to individual studies. In general, they support statistically stronger exploration than individual analyses because they comprise increased numbers and greater diversity of subjects [24]. In addition, conflicting results and biases from individual studies could be resolved or reduced [25]. For instance, Pollet et al. pointed out that patterns apparent at one scale regarding the identification and distribution of microbial communities in ticks will probably collapse into noise when viewed from other scales [26]. However, unlike social and biomedical sciences, evidence synthesis research based on systematic literature reviews have not kept pace in the field of indoor environmental microbiology, due in part to unique challenges in the field. Because the field lacks

standardized terminology encoded in an ontology and encompasses diverse research topics, researchers usually implement broad search keywords and search multiple platforms (e.g., PubMed, Web of Science) to ensure the completeness of retrieved records. This results in undesirable duplicated records, dramatically increasing manual workload in screening without effective deduplication. Existing deduplication modules are often labor-intensive by themselves, insufficient in accuracy, and limited to certain fields or databases (primarily clinical). While still useful for small datasets, due to these limitations, many approaches quickly become impractical in the fast-growing era of big data. Already, systematic reviews take tens of weeks. It has been shown that a systematic review takes an average of 164 full-time equivalent days in environmental science [27] and 67.3 weeks for a five-person team in medical fields [28]. Any processes that can accelerate computation and decrease manual labor are thus valuable improvements.

To remove hurdles toward broader application of evidence synthesis research (e.g., in indoor microbiomes), I developed an R package, RefDeduR, that performs accurate and high-throughput reference deduplication for large datasets (**Chapter 5**). The pipeline is modularized into text normalization, three-step exact matching, and two-step fuzzy matching, making it highly customizable. With finely tuned text cleaning and normalization, RefDeduR's high-confidence exact matching process outperforms many tools, even with their fuzzy matching procedure included. We semi-automate the time-consuming and error-prone manual review process in fuzzy matching by introducing a decision-tree algorithm, making it high-throughput while still maintaining accuracy. Additionally, we propose to use the inflection point of the similarity distribution curve as the cutoff threshold, making the pipeline more practical. The tool

also takes into account preprints and conference proceedings, discarding them when a peer-reviewed version is present. This is and will be increasingly important with the rise of preprint servers. A manuscript describing this work, on which I am the first author, is under review in *Research Synthesis Method*s with a preprint available in bioRxiv (https://www.biorxiv.org/content/10.1101/2022.09.29.510210v1). Additionally, users can access RefDeduR on GitHub (https://github.com/jxshen311/RefDeduR) and view further documentation and examples on the website (https://jxshen311.github.io/RefDeduR/).

1.4 Conclusions and future directions

Finally, based on my experience and the results of my work, I offer my perspective on the future of the field. Overall, I see enormous potential in the use of metagenomics as an environmental surveillance tool. I further offer my suggestions for how to capitalize on this potential and improve it for future implementation. I highlight some inconsistencies in the nascent field of indoor microbiome research and outline some practices that would improve these issues. With these and other continued advances, I am confident we will develop a deep understanding of the microbiology of indoor environments and successfully use this understanding to promote human and environmental health.

Chapter 2 Toward Accurate and Robust Environmental Surveillance Using Metagenomics

2.1 Abstract

Environmental surveillance is a critical tool for combatting public health threats represented by the global covid-19 pandemic and the continuous increase of antibiotic resistance in pathogens. With its power to detect entire microbial communities, metagenomics-based methods stand out in addressing the need. However, several hurdles remain to be overcome in order to generate actionable interpretations from metagenomic sequencing data for infection prevention. Conceptually and technically, we focus on viability assessment, taxonomic resolution, and quantitative metagenomics, and discuss their current advancements, necessary precautions and directions to further development. We highlight the importance of building solid conceptual frameworks and identifying rational limits to facilitate the application of techniques. We also propose the usage of internal standards as a promising approach to overcome analytical bottlenecks introduced by low biomass samples and the inherent lack of quantitation in metagenomics. Taken together, we hope this perspective will contribute to bringing accurate and consistent metagenomics-based environmental surveillance to the ground.

2.2 Introduction

Approximately 56% of the world's population lives in urban areas (UN, 2018) and people in developed nations spend at least 90% of the time indoors [6, 7, 29], making built environments hotspots with which humans frequently interact. Understanding and monitoring fomite transmission is critical in infection prevention [30]. The need for environmental surveillance

particularly stands out given emerging issues like the covid-19 pandemic and the continuous increase of antibiotic resistance in pathogens. Metagenomics-based methods have shown promising potential to meet this need, as they can detect entire microbial communities, as opposed to targeted identification.

However, there are several obstacles that we must overcome to bridge the gap between deploying metagenomics and generating actionable interpretations to guide infection prevention. Cultivation provides direct evaluation of microbial removal by revealing observable colonies, but this approach lacks precision [31]. Whole genome sequencing leads transmission prevention actions by monitoring strain-level dynamics of the targeted pathogen but is difficult to apply for multiple organisms simultaneously [32]. Although metagenomics dramatically expands the scope of detectable organisms compared with the aforementioned methods, it urgently needs the ability to differentiate viability, which may otherwise cause an overestimation of infection risk, and reveal the actual load of pathogens (i.e., be quantitative), for direct correlation with infection risk [33, 34]. Finally, the taxonomic resolution needs to be high enough to discriminate pathogens from closely related non-pathogens.

These challenges are both conceptual and technical in nature. They arise from the diversity of metagenomics research objectives, and are often exacerbated by intrinsic features of low-biomass environments that need to be monitored. Low-biomass samples are typical of built environment surface swabs, air, water, and rocks. Such samples are dilute, containing approximately 10^2-10^4 cells/mL for liquid samples [21, 35]. Moreover, these samples are usually limited in total cells, making it harder to obtain enough biomass. For example, when swabbing door handles, the biomass cannot be increased by enlarging the sampling area, which is finite.

Thus, special precautions are often necessary due to the low success rate of sample preparation and high possibility of contamination [36].

In this context, we focus on three critical conceptual and technical advances that need to be incorporated throughout the metagenomic environmental surveillance process: viability assessment, taxonomic resolution, and quantitation (Fig. 2.1A).

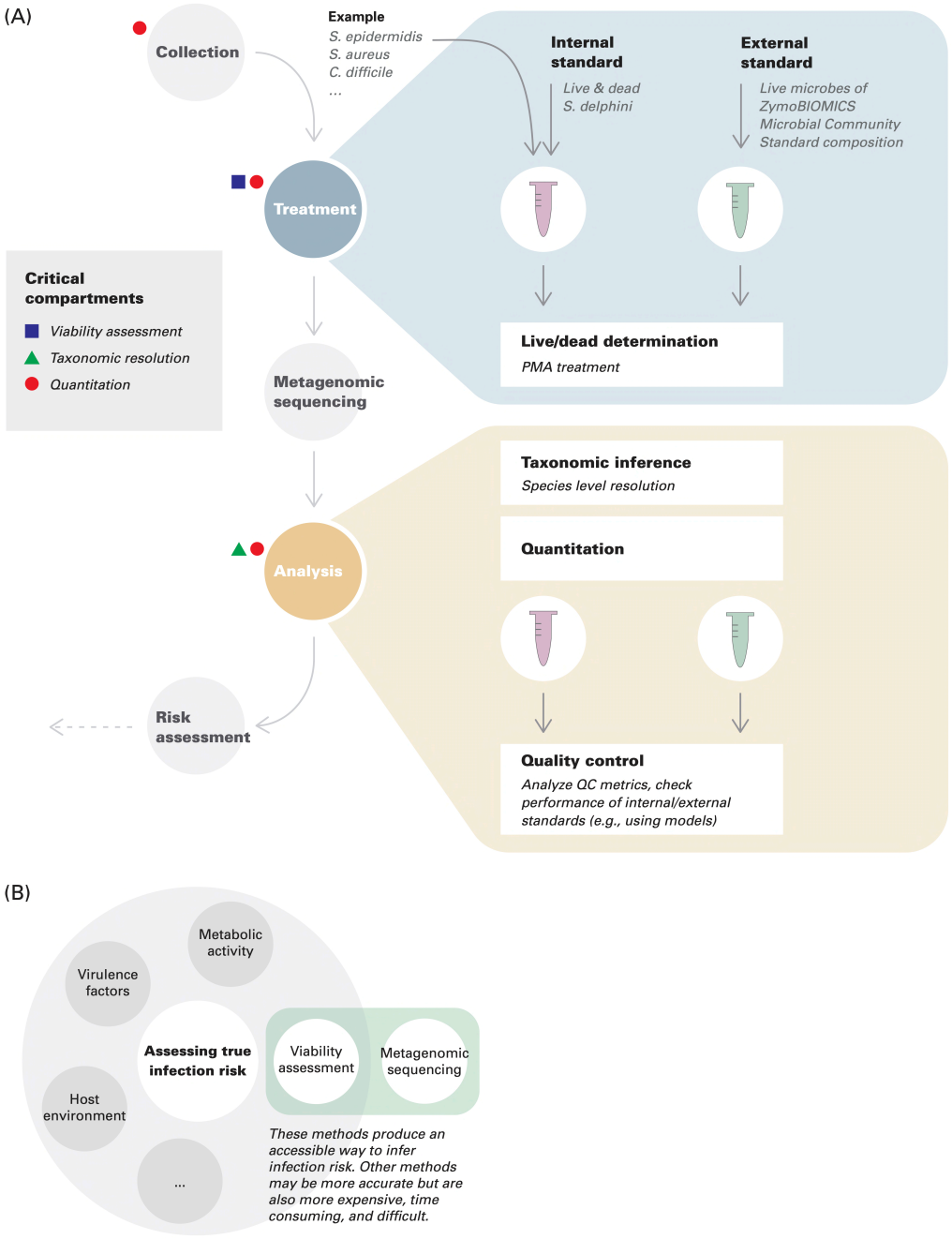


Fig. 2.1 (A) Best practices in environmental surveillance using metagenomics (with examples). Internal standards are added to collected samples, while external standards are run in parallel with samples throughout the pipeline to assure its performance. An example is provided for demonstration purposes. Note that the standards given in this example only have theoretical

potentials; more investigations are needed for benchmark and optimization. In this example, species level resolution is needed to distinguish *S. epidermidis, S. aureus*, and *S. delphini*. Assuming *S. delphini* is a good internal standard for *Staphylococcus* but not for *Clostridium*, in this case, quantitative risk assessment can only be achieved for *S. epidermidis* and *S. aureus*, but qualitative lesson can still be gained for *C. difficile*. **(B) Viability assessment coupled with metagenomic sequencing represents an accessible way to infer infection risk.**

2.3 Viability assessment using propidium monoazide (PMA)

Locations of environmental surveillance (e.g., built environment surfaces) harbor a significant proportion of dead microbes, which are captured by traditional DNA-based methods, including metagenomic sequencing [37]. Failure to assess viability could cause overestimation of infection risk. Approaches have been proposed to address this issue, with PMA treatment as a representative.

PMA treatment directly assesses membrane integrity. However, viability more broadly includes multiple underlying features, such as replication, metabolic activity [38], and virulence. These phenomena are not always interchangeable. In environmental surveillance, connecting viability to infection risk is the most informative criterion. This also highlights the importance of clarifying which criteria are being evaluated in the assessment of methods (Fig. 2.1B).

Technical challenges and optimization efforts have accompanied PMA treatment throughout its development [38-41]. The outcome is related to multiple factors, including experimental conditions (e.g., dye concentration, incubation time, light exposure time), the diversity of microbes (e.g., target gene length, differences in cell membranes, formation of spores), and the complexity of the matrix (e.g., turbidity, pH, dead cell density). Variations in these factors make PMA treatment seemingly unreliable. It may nevertheless be valuable for environmental surveillance when certain conditions are satisfied.

Application of PMA treatment to environmental samples has been hindered partly because these samples contain a diverse microbial community in a complex matrix. As pointed out by Wang et al., PMA-seq with a universal protocol is not feasible to quantify viability of realistic communities, even with *E. coli* controls spiked in [42]. To facilitate its application in surveillance, instead of insisting on differentiation of the viability for every community member, we should start by identifying sets of similar pathogens, as these groups will have the highest potential to fit in one protocol while maintaining relatively good efficiencies. Comparing this to the concept of pinpointing dynamic range in quantification, by sacrificing part of metagenomics' randomness, viability quantification may be achieved. For example, Yang et al have tried to simultaneously detect three viable *Salmonella enterica* serovars using multiplexed PCR coupled with PMA treatment [43]. Analogous principles have also been applied to the development of reagent enhancers by focusing on Gram-negative bacteria [38].

Internal standards help address biases introduced by complex matrices. To that end, peroxide-killed *Campylobacter sputorum* cells were spiked into chicken rinses in the quantification of viable *Campylobacter* [44] with encouraging results. Nevertheless, further progress should be made for widespread adoption of internal standards in risk assessment, particularly regarding the diversity and viable proportion of microbes forming the standard. Internal standards containing more than one organism are necessary to cover the diversity of the targeted microbial group, with different viable proportions to account for variations in PMA efficiency at different live/dead ratios. In routine application, once a stable relative response factor is determined for a microbe (or microbial group)-standard combination, the number of internal standards might be reduced. Creating quality control metrics, analogous to sequencing coverage and depth in metagenomics,

or adopting calculation schematics exemplified in [45] may also be noteworthy directions for future research.

Briefly, building a well-defined and continuously polished framework that limits its usage to a feasible scope but also maximizes the supporting functionality paves the way towards implementation of PMA treatment in environmental surveillance.

2.4 Inferring taxonomy in low-biomass metagenomes

Short-read shotgun metagenomic analysis reveals taxonomy without the limitations of amplicon sequencing or culture-based methods [46]. However, low-biomass samples can be more susceptible to technical factors including library size, community complexity, host DNA, and contamination. Therefore, mitigating strategies should be carefully considered. Afterwards, choosing a suitable taxonomic identification method is crucial for reliable metagenome analysis and interpretation, particularly for preventing false over-estimation of pathogens based on detection of non-pathogenic relatives or under-estimation of risk from pathogens with very low infective doses due to limitations in detecting rare taxa.

Differences in coverage and depth can result in differing estimates of taxonomic richness and diversity in identical samples, primarily at low level ranks, such as genus or species (Table 2.1) [47, 48]. Smaller read libraries are particularly challenged by a diminished capacity to detect rare taxa and accurately estimate overall taxonomic abundance of samples [49] because more abundant members in the metagenome have a higher likelihood of detection [50]. Additionally, the lowered overall information content of low coverage and depth read libraries impact the ability to identify low level taxonomic ranks.

Table 2.1 Definitions and calculations of common sequencing terms.

Term	Definition	Approach(es)	Reference(s)
Read library	The number of reads generated from a single metagenomics sample	-	-
Coverage	The fraction of a metagenome represented by the read library	Taxonomy-based rarefaction curve Read redundancy-based rarefaction curve	[51, 52]
Read depth	The number of times a particular base is captured by a read	$base\ x\ reads\ that\ map\ to\ base$	[53, 54]
Mean read depth	The read depth averaged across the metagenome assembly	$\frac{reads\ mapped\ to\ assembly\ x\ average\ read\ length}{assembly\ size}$	[55]

Low-biomass samples are especially sensitive to the presence of contaminants, as the true signal can be easily overwhelmed [36]. Multiple avenues of contamination exist, including sample preparation and DNA extraction, from either the reagents or the researchers themselves, and carryover between sequencing runs. Methods to reduce contamination include UV radiation and DNase treatment of kit reagents to specialized library preparation workflows [56-59]. Metagenomic samples should be accompanied with kit extraction negative controls and DNA-spiked positive/internal controls during sequencing runs to identify sources of contamination [36, 56, 60].

Given the susceptibility of low biomass samples to contamination, special care should be taken in preparing appropriate controls to avoid misidentifying contaminants as true signals, as cross-contamination can confound epidemiological or strain-tracking efforts. [61, 62]. The simplest approach is to remove sample reads that align to taxa found in the negative controls [63]. This can result in removing reads belonging to the true taxonomic composition of a sample, and especially problematic in instances where negative controls are contaminated by sample DNA or belong to a pathogen under surveillance. Other approaches include filtering sequences that fall below a designated relative abundance threshold or map to taxa in a contaminant database [64]. Approaches that remove low frequency sequences are not recommended for low biomass samples. Employing blank negative controls and study-specific kit negative controls could help in identifying genuine instances of contamination in low biomass samples and detecting kit-based contamination. Bioinformatics pipelines that incorporate either one or a combination of the above approaches have been developed to streamline identification of contaminants and/or cross-contamination [64-66].

A variety of tools are available to characterize the taxonomic composition of a metagenomic sample and broadly follow two approaches: using reads as inputs or assembling reads and then using the genes/contigs as input [63, 67]. Both approaches have tools that use k-mer, alignment, and marker gene matching implementations. A meta-analysis of both approaches demonstrated that at artificially lowered read library sizes, read-based classification methods maintained their accuracy compared to assembly-based methods [68] because assembly-based methods rely on having sufficient overlapping read depths. Metagenomic samples from low-biomass environments with insufficient coverage (<20X read depth over the whole metagenome) may require read-based

taxonomic classification [46, 68]. Similarly, inherently low read depths may limit the level of taxonomic resolution, as strain-level analysis requires high read depth to distinguish between SNP variants or marker gene variants (e.g. characterizing the relatedness of strains during an outbreak using SNPs) [49, 69-71]. If strain-level variants are desired, merging paired-end reads or using sequencing technologies that generate longer reads may be necessary [69].

Choosing an appropriate taxonomic reference database can greatly impact the breadth of taxa identified [72, 73]. For example, a reference database built from gut bacteria may not identify environmental taxa but may be suitable for identifying gut pathogens in the environment. Many tools offer the option of using either precompiled or custom reference databases [63]. CAMISIM, a tool that simulates microbial metagenomic datasets, can be used by researchers to test different approaches [74].

2.5 Quantitative metagenomics in environmental surveillance

Conceptually, quantitative metagenomics has appeared in many ways in microbiome research, ranging from performing basic calculations of abundance, to normalizating metrics to these calculations, to the ultimate goal of absolute quantification as in qPCR [75].

At any level, parameters or metrics for profiling a microbial community are the basis of analysis. As such, selecting meaningful parameters is the first step toward quantitative metagenomics. Of the five parameters summarized in their review [50] (Table 2.2), Nayfach and Pollard suggested that cellular relative abundance and average genomic copy number are the more biologically meaningful and quantitative strategies. However in reality, relative abundances are more frequently used. For instance, quantitative metagenomics is applied in gut microbiome studies to

identify unique biomarkers [76, 77], to compare disease and health states [78], and to predict resistome [79], all of which use cellular/gene relative abundance normalized by genome/gene length through mapping reads to reference genomes/genes (e.g., the MetaHIT gene catalogue).

Table 2.2 Parameters to profile microbial communities summarized by Nayfach and Pollard. Theoretical calculation equations are created based on our interpretation of the authors' descriptions, which are also adapted to the context of environmental surveillance.

Parameter	Theoretical calculation equation
Cellular relative abundance	$CRA(x) = \frac{number\ of\ cells\ of\ taxon\ x}{number\ of\ cells\ in\ the\ community}$
Gene relative abundance	$GRA(x) = \frac{number\ of\ gene\ x}{number\ of\ genes\ in\ the\ community}$
Average genomic copy number	$AGCN(x) = \frac{number\ of\ gene\ x}{number\ of\ cells\ in\ the\ community}$ $AGCN(x) = GRA(x)\ \times\ average\ genome\ size\ (in\ number\ of\ genes)$
Cellular absolute abundance	$CAA(x) = \frac{number\ of\ cells\ of\ taxon\ x}{volume/weight/area}$
Gene absolute abundance	$GAA(x) = \frac{number\ of\ gene\ x}{volume/weight/area}$

However, cellular and gene absolute abundances are the most promising parameters in environmental surveillance, which is predicated on the actual load of pathogens or pathogenic genes. Moreover, absolute abundances allow better comparisons across samples [80] and across taxa/genes [50, 81].

Technically, several challenges remain to be overcome towards accurate and unbiased estimation of absolute abundances. It requires careful re-design of the entire study in a stringently quantitative framework, beginning with sample collection. Samples should be collected in an absolute framework (per unit volume, weight, area, etc.), and this framework should be

maintained throughout sample preparation. Taking surface microbiome studies as an example, extra considerations include the measurement and documentation of the swabbing area as well as volume of sampling buffer and other solutions used in the entire workflow, and the examination of the recovery rate where sample loss is non-negligible. Furthermore, normalization by genome/gene length is necessary to account for the varying representativity in sequencing a read from genomes/genes of different lengths.

This is not always easy in reality. For example, the sampling area of sink biofilms is difficult to assess when destructive sampling is not permitted. Even if the samples are collected in a strictly quantitative way, other steps in the sample treatment process still need to be conducted quantitatively. Taking DNA extraction as an example, instructions like "transfer up to 600 µL of supernatant to a clean tube" destroy the quantitative chain and prevent us from calculating dilution factors. Accurate normalization by genome/gene length requires continuous effort in expanding genome/gene databases and in incorporating genome/gene normalization into bioinformatic pipelines [82]. In the interim, mapping reads to a set of well-studied while also universal (within the study scope) marker genes (e.g., 16S rRNA genes in bacteria) could serve as a workaround [50] but unfortunately introduces its own biases.

Besides incorporating qPCR or flow cytometry, introducing standards has great potential to enhance the quantitation ability of metagenomics. In this context, internal standards outcompete external standards, partly because variations in sample treatment seem inevitable (e.g., shotgun libraries undergo equimolar normalization) and because the relationship between the amount of input material and the number of output reads remains obscure. Internal standards also

compensate for errors resulting from any non-quantitative processing steps following their addition. Some efforts have been made to incorporate internal standards into the metagenomic pipeline, such as spiking mock-community cells into the collected samples [60], adding genomic DNA just prior to cell lysis in DNA isolation [80, 83], and including a set of synthetic DNA before library preparation [84]. Despite these advances, systematic investigations are needed to benchmark methods, identify limitations, and validate use in various contexts. Clearly, a set of standards are needed to account for the complexity of samples and the diversity of targets. But which performs the best among mock communities, genomic DNA isolated from cultured microorganisms, and synthetic DNA remains unclear. Nor is it known at which step the standards should be added and at what dose. Moreover, when exogenous materials are hard to find, the standard addition method may be worth exploring [85]. Finally, the standards should be evaluated holistically at the pipeline level for their compatibility and functionality across multiple steps (e.g., PMA treatment, metagenomic sequencing). Ideally, an optimal pipeline should also be equipped with quality control compartments such as external standards and mathematical models which assess and calibrate biases [60].

2.6 Limits of detection

Incorporating viability assessment, adequate taxonomic resolution, and quantitation into metagenomics will yield invaluable insights into environmental surveillance. But perhaps more critical than interpreting observed data is interpreting non-detects. Ultimately, a viable signal must be linked to infection risk by determining the threshold load of pathogens to cause an infection when they are contracted from a fomite. This threshold is pathogen-specific. Thus,

reference values, like clinical standards for antimicrobial susceptibility testing (M100, CLSI) [86], as technological standards are necessary.

Moreover, sensitivity of every step in the pipeline must be accounted for in data interpretation, as the overall sensitivity is determined by the lowest step. As mentioned above, features of the community being sampled can affect the expected breadth of coverage. Complex samples with many taxa, host DNA, or stochastic eukaryotic DNA (for example in surface swabs) may require high read library sizes to ensure sufficient breadth of coverage [73, 87]. This is especially important when non-target DNA can represent the majority of the reads generated, resulting in decreased capacity to detect rare taxa and at fine-grained resolutions. Pilot studies that assess coverage using taxonomy or read redundancy-based rarefaction analysis can help determine an appropriate library size [88]. When pathogens are rare compared to other organisms, limit of detection (LoD) is a crucial parameter as it determines the maximum possible load of pathogens when they are not detected. Given the inherent nature of metagenomic shotgun sequencing that a fixed total number of reads are distributed based on the relative proportion of genetic materials present in a batch, LoD must be approximated with the microbial community to be sequenced at a batch-based pace. Because of this matrix-dependent characteristic, it is impractical to get a universal LoD for the technique “metagenomic sequencing”. Empirically, LoD can be estimated relative to the least abundant but detected members in the internal standards or the sample itself.

2.7 Discussion

In summary, metagenomics has enormous potential in environmental surveillance of pathogens as it simultaneously detects multiple organisms and functional genes of interest, e.g., antibiotic

resistance. However, the following steps need to be taken to ensure that metagenomic data can practically be applied to risk assessment:

1. Rationally address inherent conceptual limitations regarding viability. For example, PMA treatment assesses membrane integrity, not infectivity; but relationships can be deduced between intact cells and infectious organisms.
2. Rationally address inherent limitations regarding taxonomy. For example, while almost all *Salmonella* are pathogenic, higher taxonomic resolution is needed to distinguish pathogenic *Pseudomonas*.
3. Incorporate internal standards. Doing so will compensate for biases introduced by complex environmental matrices, yield quantitative results, and correct both random and systematic errors.
4. Holistically integrate multiple steps in pipeline optimization. Specifically, internal standards can be incorporated for multiple operations including PMA treatment, taxonomic inference, and quantification.

Metagenomics-based environmental surveillance has potential for developing rich datasets that aid surveillance. Metagenomic data can aid in linking taxa with virulence factors and antibiotic resistance genes. Strain-level data can further track transport of pathogens in the environment or reveal microbial networks of interactions among patients, employees, medical devices or wastewater. Promoting crucial standardizations ranging from sampling protocols to data analysis, curation and presentation, can not only help produce internally consistent results but also increase external compatibility with data generated in different studies [50] or with different protocols [89].

Chapter 3 An improved workflow for accurate and robust healthcare environmental surveillance using metagenomics

3.1 Abstract

Background: Effective surveillance of microbial communities in the healthcare environment is increasingly important in infection prevention. Metagenomics-based techniques are promising due to their untargeted nature but are currently challenged by several limitations: 1) they are not powerful enough to extract valid signals out of the background noise for low biomass samples; 2) they do not distinguish between viable and non-viable organisms; and 3) they do not reveal the microbial load quantitatively. An additional practical challenge towards a robust pipeline is the inability to efficiently allocate sequencing resources a priori. Assessment of sequencing depth is generally practiced post hoc, if at all, for most microbiome studies, regardless of the sample type. This practice is inefficient at best, and at worst poor sequencing depth jeopardizes the interpretation of study results. To address these challenges, we present a workflow for metagenomics-based environmental surveillance that is appropriate for low-biomass samples, distinguishes viability, is quantitative, and estimates sequencing resources.

Results: The workflow was developed using a representative microbiome sample, which was created by aggregating 120 surface swabs collected from a medical intensive care unit. Upon evaluating and optimizing techniques as well as developing new modules, we recommend best practices and introduce a well-structured workflow. We recommend adopting liquid-liquid extraction to improve DNA yield and only incorporating whole-cell filtration when the non-bacterial proportion is large. We suggest including propidium monoazide treatment coupled with

internal standards and absolute abundance profiling for viability assessment, and involving cultivation when demanding comprehensive profiling. We further recommend integrating internal standards for quantification, and additionally qPCR when we expect poor taxonomic classification. We also introduce a machine learning-based model to predict required sequencing effort from accessible sample features. The model helps make full use of sequencing resources and achieve desired outcomes.

Conclusions: This workflow will contribute to more accurate and robust environmental surveillance and infection prevention. Lessons gained from this study will also benefit the continuing development of methods in relevant fields.

3.2 Introduction

Effective microbial surveillance in the built environment is increasingly important in infection prevention, given the persistence of pathogens in environmental reservoirs and their potential transmission to patients [90-96] (e.g., carbapenem-resistant *Klebsiella pneumoniae* in sink drains [97, 98]). Furthermore, with the emergent studies showing synergistic relationships among pathogens on hospital surfaces [19] and the possibility for pathogenic bacteria to acquire antibiotic resistance genes from non-pathogenic neighbors [20], it is necessary to expand from targeted surveillance to untargeted methods. Untargeted methods are advantageous in identifying novel or rapidly emerging pathogens [21]. Metagenomics-based techniques are the most promising option to achieve these goals but are currently challenged by several limitations: 1) they are not powerful enough to extract valid signals out of the background noise for low biomass samples; 2) they do not distinguish between viable and non-viable organisms; and 3) they do not reveal the microbial load quantitatively [50, 99].

For challenge 1), adoption of appropriate negative controls has been emphasized [36, 60], along with various bioinformatic filtering tools to remove putative contaminants [64, 65]. While current efforts have largely focused on contamination prevention, increasing the biomass itself remains understudied [21]. Having adequate biomass is essential, as previous work has indicated that a small amount of starting material (i.e., DNA) has adverse impacts on the outcome regardless of sample processing methods [100]. In practice, methods have been adopted as temporary fixes, such as pooling samples from different sites or dates [101], and using wipes instead of swabs as sample collectors [90]. However, these workarounds are not always available [56]. Moreover, systematic evaluation and benchmarking of optimization strategies for metagenomic sample preparation remain largely unexplored.

For challenge 2), propidium monoazide (PMA) is the most widely used viability indicator compatible with molecular techniques. Though intensively optimized [102], its efficacy and applicability in combination with metagenomics are controversial. A semi-quantitative systematic evaluation concluded that PMA treatment coupled with 16S rRNA gene amplicon sequencing (PMA-Seq) is reliable when the microbial community is not very complex, while uncertainties increase dramatically with complexity [103]. The uncertainties come from both heterogeneity of microorganisms (e.g., cell envelope structure differences, spore formation), and complexity of the background matrix (e.g., turbidity, salt concentration, dead cell density) [39, 40, 99]. While the microbial communities to be surveilled have their inherent advantage of being low complexity, little is known about the effectiveness of incorporating PMA with multi-species internal standards. To be appropriately rigorous, comparisons are needed relative to standard

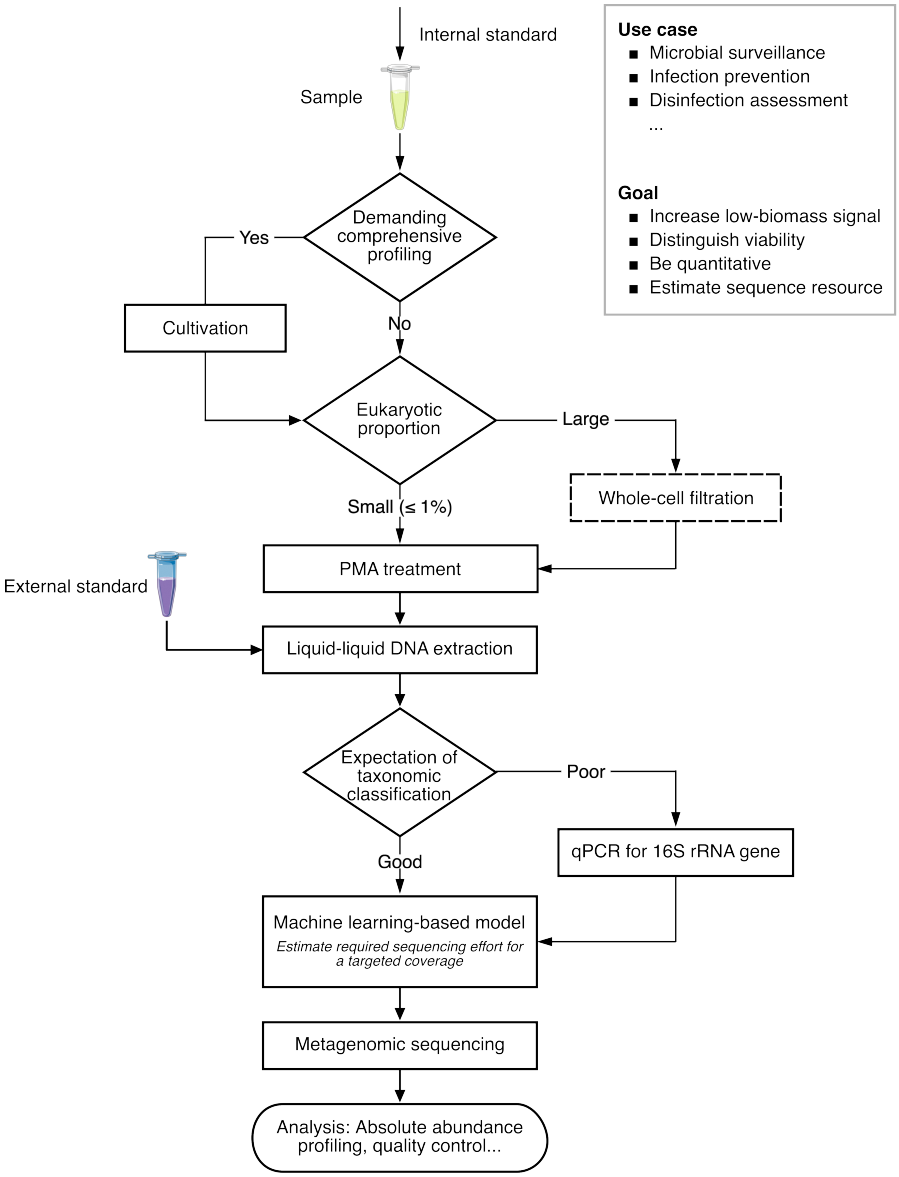
surveillance that does not consider viability (i.e., no PMA), as well as traditional methods (i.e., cultivation).

For challenge 3), pitfalls of using relative abundances in microbial profiling have been widely indicated. Such pitfalls include but are not limited to lack of unique connections between biological interpretations and experimental observations and unreliable comparisons across samples [50, 104, 105]. Strikingly, mis-selection of analytical tools for relative abundance data could lead to as high as 100% false discovery rates [104]. Besides flow cytometry [106, 107], combing sequencing with quantitative PCR (qPCR) and including internal standards [80] are two major means of making quantitative estimations out of next-generation sequences. In practice, previously reported applications for qPCR include air and dust samples in classrooms [108]; for internal standards, applications include Amazon River plume [83], soil [109], and stool samples [110]. However, comparisons are not yet available between metagenomics coupled with qPCR and with internal standards using low biomass environmental samples in the immediate vicinity of humans (e.g., healthcare settings).

An additional practical challenge in developing a robust pipeline with metagenomics is how deep one should sequence. While useful in whole genome sequencing, recommendations of coverages expressed by folds of genome sizes (e.g., 15X to 60X) are not readily transferable to metagenomic sequencing (MetaSeq), as reads do not equally distribute across members with substantially different abundances. Nonpareil, a redundancy-based tool, estimates and projects abundance-weighted average coverage for metagenomics (expressed in percentage) [52, 88, 111]. This helps reduce erroneous interpretations out of metagenomic results. Yet expected coverage is still largely unpredictable before sequencing is run. Researchers usually rely on

previous experience of similar samples and the available budget to determine the sequencing effort (read size, unit: bp), which may lead to either a coverage too low, thus limiting the extractable information [20], or a waste of resources [90].

To address these challenges, we present a workflow for metagenomics-based environmental surveillance that is appropriate for low-biomass samples, distinguishes viability, is quantitative, and estimates sequencing resources (Fig. 3.1). Liquid-liquid extraction, PMA treatment equipped with internal standards and absolute abundance profiling, qPCR, and a machine learning-based model are the recommended components for the comprehensive workflow, with whole-cell filtration and cultivation as optional accessories.



Treatment in dashed box needs further investigation.

Fig. 3.1 A workflow for metagenomics-based environmental surveillance that is appropriate for low-biomass samples, distinguishes viability, is quantitative, and estimates sequence resources.

3.3 Methods

3.3.1 Sample collection, aggregation, and cultivation

We collected 120 surface swabs from the 28-bed medical intensive care unit at Rush University Medical Center (RUMC) in October 2018. RUMC is a 720-bed tertiary care teaching hospital in Chicago, IL. Samples were collected from door sills, computer keyboards, light switches, nurse calls, and bed rails in 13 single-bed patient rooms, as well as door sills in 4 medication rooms, 2 public restrooms, 1 staff-only restroom, and the communicating space of MICU (Table S3.2). Weighted mean area of sampled surfaces was 216 cm^2. Patient rooms were selected to keep a relatively balanced number for both contact isolation and non-contact isolation rooms. Healthcare providers and visitors entering contact isolation rooms are required to wear gowns and gloves, which may reduce transmissions via contaminated healthcare providers. Room temperature and relative humidity were documented during the collection, which varied slightly across rooms, with the average being 23.8°C and 45%, respectively. Each sample was collected by 3 COPAN Nylon Flocked Swabs (Copan Diagnostics, Murrieta, CA, USA) and 1.5 mL Phosphate Buffered Saline with 0.02% Tween 80 (PBST), and stored at 4°C for up to 24 h prior to extraction, aggregation, and cultivation [112, 113]. Swabs were extracted and aggregated to create a representative microbiome sample [37, 113, 114]. Aliquots of this aggregation sample were then subjected to different processing methods (i.e., several DNA extraction methods, microbial community standard spike-in, PMA treatment and whole-cell filtration) to find best practices of the workflow (Fig. S3.1).

To capture a large fraction of the indoor microbiome diversity, we cultured the samples with 4 different media: tryptic soy agar (TSA), Reasoner's 2A agar (R2A), 0.1 strength R2A at 25°C,

and blood agar (BA) at 37°C, all supplemented with 4 mg/L itraconazole [113]. This resulted in 233 cultivable isolates. All colonies that could be individually picked or purified were subject to taxonomic identification by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) using the VITEK® MS Mass spectrometry microbial identification system (BioMerieux, Marcy-l'Étoile, France) and the VITEK MS V3.2 FDA 510(k) cleared database. Among the 233 isolates, 201 were identified. It is important to note that, because multiple media types were used, the number of isolates for each species identified does not represent the relative abundance of this species in the sample, as some species may have grown on multiple media.

3.3.2 Standard addition, PMA treatment and whole-cell filtration

All treatments were done in triplicate, including cultivation.

Standard addition

Aliquots were snap frozen and stored at -80 °C until further processing to maximize the integrity of samples and avoid degradation resulting from long-term storage at 4°C [37, 115]. Samples were thawed at 4°C prior to treatments. ZymoBIOMICS Microbial Community Standard (Zymo Research, Irvine, CA, USA) was used as both the internal standard and the external standard. As the internal standard, 6.50 µL Zymo community was spiked into 1 mL aggregate sample, following the criterion that DNA of the species with the highest abundance in the Zymo community approximates 1% of the total DNA in the aggregate sample [83, 116]. As the external standard, aliquots of the Zymo community were run in parallel with aggregate samples throughout the workflow to assure its performance.

PMA treatment

Following standard addition, PMA treatment (Biotium, Fremont, CA, USA) with an optimized protocol was applied to half of the samples within each group [39, 40, 102, 117, 118]. The protocol was first validated by reproducing the work of Nocker et al. (2006) using *Escherichia coli* (ATCC 8739) as model strain (Fig. S3.2). *E. coli* was grown to the exponential phase. The culture was then split into two aliquots, one of which was killed by heat inactivation at 95°C for 7 min in Eppendorf ThermoMixer shaking at 400 rpm for homogenized heating. After cooling to room temperature, live and heat-killed aliquots were mixed following the same ratios as in Nocker et al. (2006), yielding samples of 6 different expected live cell ratios. Viability of both live and heat-killed cultures was confirmed by spread plating onto TSA and incubating at 37°C overnight. Half of each constructed sample underwent PMA treatment. The results were evaluated by both DNA concentration ratio quantified by Quant-iT™ PicoGreen™ dsDNA Assay (ThermoFisher, Waltham, MA, USA) and copy number ratio by qPCR with 16S universal primers (341F and 534R) (Fig. S3.2). Briefly, a final concentration of 25 μM PMA was used, and several steps were conducted to ensure the consistency across samples and minimize nonspecific reactions between PMA and random sample components, including 1) adding PMA to tube caps and inverting all tubes simultaneously, 2) working under red light, and 3) protecting samples from light as much as possible before the light activation step in the PMA-Lite™ device (Biotium, Fremont, CA, USA). An aliquot of samples for each replicate was preserved at -80°C until DNA extraction, with the rest stored at 4°C for downstream filtration.

Whole-cell filtration

Whole-cell filtration was conducted using EMD Millipore 25 mm Glass Vacuum Filter kit (MilliporeSigma, Burlington, MA, USA), 125 mL filter flask, and Gemini vacuum pump in a biosafety cabinet following aseptic techniques (Fig. S3.3). Notably, autoclaved tweezers were used to avoid possible contaminations from touching sensitive parts of the set-up. Samples were filtered by 100 μm nylon membrane, followed by 80 μm and 41 μm nylon membranes and 5 μm PVDF membrane (MilliporeSigma, Burlington, MA, USA). 1 mL PBS was added to the falcon tube and flask at each step to increase the sample recovery by rinsing the inner wall. The filtered samples were then subjected to 3-fold (relative to the volume before filtration) vacuum concentration with an Eppendorf Vacufuge plus. Filtered liquid samples and filter papers were preserved at -80°C until DNA extraction. To increase the extraction efficiency from filter papers, we compared 1) cutting them with scissors into 9 pieces, 2) grinding them with metal spatula after snap freezing in liquid nitrogen, and 3) directly putting the whole filter paper into the preservation tube. We finally selected the third option as this was the most operationally feasible way without high risk of contamination.

3.3.3 Negative controls

To combat the susceptibility of low-biomass samples to contamination, we included 4 types of negative controls along the workflow, namely, 6 negative field controls, 6 negative media controls, 12 negative filter controls, and 7 negative kit controls [36, 60, 99, 119]. The negative controls were processed in parallel with the surface samples, including metagenomic sequencing and bioinformatic analysis.

Negative field controls were collected exactly the same as surface samples, except that the swabs were exposed to the air without contacting targeted surfaces. One negative field control was collected at the beginning and the end of each sampling session. Two negative media controls were included in each sampling session, which were unopened media with swabs from the same lot. Each collected control was split into triplicate and processed along with samples [112, 120]. Negative filter controls were included in triplicate for each pore size by letting sterile PBST flow through the vacuum filtration system attached with blank filter papers. Additionally, 7 negative kit controls were processed across batches of DNA extractions.

3.3.4 DNA extraction, qPCR and metagenomic sequencing

To ensure enough DNA recovery, we performed an initial optimization on a separate set of surface swab samples collected from the same MICU prior to working with the aggregate sample (Fig. S3.4). Column-based methods were first tried due to its widespread usage in the field. We examined Qiagen DNeasy PowerSoil Kit with standard protocol and a modified version by 1) changing from vortex lysis to bead-beating lysis, 2) introducing heat incubation after bead-beating, and 3) adding 50 µL water each time for twice in total at the elution step. DNA yields of both were below the limit of detection. Liquid-liquid extractions were performed afterwards for their high-yielding potentials. Phenol-chloroform extraction resulted in the highest yield (186.27±55.51 ng/µL by NanoDrop), but the purity indicated by 260/280 was not acceptable (1.36±0.03). Lucigen MasterPure™ Complete DNA and RNA Purification Kit also resulted in high yields when coupled with bead-beating and heat lysis and better purity than phenol-chloroform extraction (260/280 1.62±0.02). We attempted to improve the purity using the Agencourt AMPure XP PCR Purification kit. However, we did not see a purity increase (260/280

1.62±0.03) and incurred a 64.70% DNA yield drop. Based on the above tests, we noticed that methods involving columns (Qiagen PowerSoil) or magnetic beads (Agencourt AMPure) greatly decreased the DNA yield. Because the primary concern for surface samples is low biomass, increasing DNA yield is considered more critical than bringing 260/280 to the desired range of 1.8-2.0. Therefore, the Lucigen MasterPure™ Complete DNA and RNA Purification Kit with the adapted protocol was chosen for all subsequent analyses [121]. Samples were thawed at 4°C prior to DNA extraction and DNA concentrations were quantified by Quant-iT™ PicoGreen™ dsDNA Assay [122].

V3 region of the 16S rRNA gene was amplified in qPCR using universal primers (341F: 5'-CCT ACG GGA GGC AGC AG-3', 543R: 5'-ATT ACC GCG GCT GCT GGC A-3') [123]. The 20 µL reaction mixture consisted of 10 µL PowerUp™ SYBR™ Green Master Mix (Applied Biosystems), 0.6 µL forward primer (10 µM), 0.6 µL reverse primer (10 µM), 5.0 µL DNA templates (pre-diluted if necessary), and 3.8 µL nuclease-free water. The reaction was run in technical triplicate on a QuantStudio 3 Real-Time PCR System (Applied Biosystems) with an initial denaturation step at 95°C for 2 min, followed by 40 amplification cycles (95°C, 15 s; 56°C, 15 s; 72°C, 1 min) and a melting curve stage (95°C, 15 s; 60°C, 1 min; 95°C, 15 s). No-template control and 5~8 standards were included in each batch to generate the standard curve (efficiency > 90%; $R^2 > 0.99$). Plasmid DNA constructed by TOPO™ TA Cloning™ Kit (Invitrogen, Waltham, MA, USA) was used as standards.

Extracted DNA was shipped on dry ice to the UMICH Microbiome Core (Ann Arbor, MI, USA) for library preparation using Nextera™ DNA Flex Library Prep Kit and paired-end 250-bp shotgun metagenomic sequencing on an Illumina MiSeq platform (MiSeq Reagent Kit v2).

Libraries were normalized at equal molarity in a 4 nM final concentration pool before sequencing, and for samples without enough DNA (e.g., negative controls), all available materials were used.

3.3.5 Data analysis

Sequence data processing

KneadData (v0.6.1) was first used to clean the shotgun sequences with default parameters. Reads present in the human reference database (hg37_and_human_contamination) and negative controls were filtered out. Metaxa2 (v2.2) [124-126] coupled with SILVA 132 SSU database [127, 128] was chosen to generate taxonomic profiles after comparing it with MetaPhlAn2 (v2.6.0) [129] and MetaPhlAn3 (v3.0.7) [130, 131]. The evaluation was conducted based on their performance on external standards and cross validation with cultivation results for untreated aggregate samples. Default parameters were used for all three tools. MetaPhlAn3 was ruled out mainly because it only generates marker genes at the species level and an average of 80.32% metagenome was deemed unknown for our samples. For external standards, both Metaxa2 and MetaPhlAn2 recognized all 8 bacteria species demonstrated in the theoretical composition, but MetaPhlAn2 failed to classify the 2 eukaryotic species. Moreover, it did not classify *Pseudomonas fluorescens* and barely classified *Pseudomonas stutzeri* from aggregate samples, while Metaxa2 recognized both. Though Metaxa2 included a few spurious taxa, all can be eliminated by removing singletons. Since the primary goal of this study was to compare techniques and recommend best practices, sensitivity outweighed specificity. Therefore, Metaxa2 was selected and singletons were removed for downstream analyses. Taxa were labeled to the lowest classifiable level, with species level as the ultimate target [103]. Metagenomic sequencing

coverage for all samples was estimated by Nonpareil (v3.303) under kmer mode using default settings [52, 111].

Comparison between gene- and genome- centric approaches

An evaluation of genome-centric analysis was based on 9 untreated aggregate samples (Table S3.3). We first compared Megahit (v1.0.6.1) [132] and metaSPAdes (v3.14.1) [133], and selected metaSPAdes as the assembly tool because it produced better-quality contigs (i.e., longer and fewer contigs, higher N50) as assessed using Quast (v4.4) [134]. MetaBAT2 (v1.7) [135] was then used to bin the contigs and reconstruct genomes, with bin quality checked by CheckM (v1.0.18) [136]. Each sample yielded a single high-quality bin (completeness > 92%, contamination < 2%). Taxonomy of the bins was subsequently assigned using GTDB-Tk (v1.7.0) [137]. All the 9 metagenome-assembled genomes were classified as *Pseudomonas fragi*. The genome-centric approach thus represents the least sensitive method compared with cultivation and gene-centric metagenomics, especially for less abundant taxa. Because the genome-centric method only detected 1 taxon, while cultivation and gene-centric metagenomics detected 12 and 11 taxa, respectively (Fig. 3.4a), this approach was excluded from further analysis.

Statistical analysis

Statistical analyses and data visualization were conducted in R (v4.0.4) [138] with packages such as Nonpareil, vegan, ape, ggplot2 [139], and dplyr. Principal coordinate analysis (PCoA) based on Jaccard metric was performed to demonstrate beta diversity [119]. Differences between groups were determined by Student's T-test or ANOVA coupled with Tukey's post-hoc test, depending on the number of groups under comparison. $P \leq 0.05$ was defined as statistically significant. Significance codes are: $p > 0.05$ (ns), $0.01 < p \leq 0.05$ (*), $0.001 < p \leq 0.01$ (**), $p \leq$

0.001 (***). Package “fitdistrplus” was implemented to find the best-fit distribution for Nonpareil diversity. Machine learning models were trained using the package “caret”.

3.3.6 Availability of data and materials

The raw shotgun metagenomic sequencing data are available in the NCBI SRA repository under Bioproject number PRJNA765404. Source code and supplementary data for reproducing analyses are available under MIT license at https://github.com/jxshen311/workflow_metagenomic_environmental_surveillance. Protocols are available at Protocol Exchange with DOIs: 10.21203/rs.3.pex-1656/v1 (sample collection, extraction, and cultivation), 10.21203/rs.3.pex-1657/v1 (snap freezing), 10.21203/rs.3.pex-1659/v1 (PMA treatment), and 10.21203/rs.3.pex-1658/v1 (DNA extraction).

3.4 Results

3.4.1 Liquid-liquid extraction improves the power of handling low-biomass samples

To improve DNA yield of low biomass samples, we first compared 3 categories of extraction methods. Bead-beating and heat lysis followed by liquid-liquid extraction was the optimal method, as opposed to widely used column- and magnetic bead-based methods (“Methods”, Fig. S3.4) [140, 141]. Notably, no detectable DNA was recovered using Qiagen DNeasy PowerSoil Kit. Supplementary to the recommendation that DNA input ≥ 1 ng for Nextera Flex Library Prep kit, we correlated it to the practical outcome and found that DNA > 11.2 ng corresponded to raw reads > 1e+05 (Fig. S3.4).

In addition to being low biomass, environmental samples of interest are usually in the immediate vicinity of humans, and thus often contain eukaryotic cells. These cells may compete with

bacteria for sequencing depth, lowering detectable resolution on bacteria, especially for low-abundance members. Collection methods such as swabs and wipes can further recover abiotic debris along with biological materials, to which chemicals potentially interfering with downstream experiments may adsorb [123]. To address these issues, we evaluated the implementation of whole-cell filtration in the workflow. Four filtration steps (100, 80, 41, 5 μm) were conducted in descending order of pore sizes. For our samples, filtration did not exert a significant effect on detected proportions of bacteria (Fig. 3.2a) or eukaryotic reads (Fig. S3.5), according to paired t tests ($p \geq 0.05$). Considering that the non-bacterial proportion of our samples was relatively small (~1%), filtration appears ineffective (or unnecessary) in increasing the bacterial proportion by excluding eukaryotic cells for samples with similar characteristics. Instead, most of the eukaryotic reads were human-associated and thus able to be removed *in silico* (Fig. S3.5). Moreover, we did not observe an increase in the number of rare taxa post filtration. Nevertheless, filtration did not negatively affect the number of recoverable taxa (Fig. S3.6) [142, 143].

As expected, filtration introduced biomass loss of ~13-44%, according to 16S rRNA gene copy number (Fig. 3.2b). The biomass loss may be compensated by a two-fold concentration, material permitting. Alternatively, the total biomass loss can be reduced in practical applications where one-step filtration is streamlined. Filtration did not impact the overall bacteria composition (Fig. 3.2c), nor did it change the relative abundances of top abundant taxa (average abundance > 1%) (Fig. S3.7). This evidence supports the validity of using filtration to concentrate bacterial samples in sequencing-based experiments for profiling relative abundances. However, the absolute abundances would be affected disproportionately, as the extent of biomass loss varied

across samples (Fig. 3.2b). Therefore, when an absolute metric is of interest, the recovery rate needs to be rigorously measured. Bacteria retention profiles on 5 μm filters were similar to those of the liquid samples. However, bacterial members were not proportionally retained by filters of a larger pore size (100, 80, 41 μm) (Fig. 3.2d). Hence treating microbial samples with large pore-size filters may introduce biases, even when relative abundances are used.

Taken together, for samples whose non-bacterial proportion is small (e.g., ~1%), it is unnecessary to incorporate filtration to increase the bacterial fraction. Filtration is valid in concentrating samples. However, for low biomass samples which are low in both cell density and quantity, biomass loss outweighs the slight increase of bacterial signal. Instead, switching to a high-yield DNA extraction method, such as liquid-liquid extraction, can achieve higher folds of signal improvement (DNA concentration from undetectable to 18.62±1.16 ng/μL).

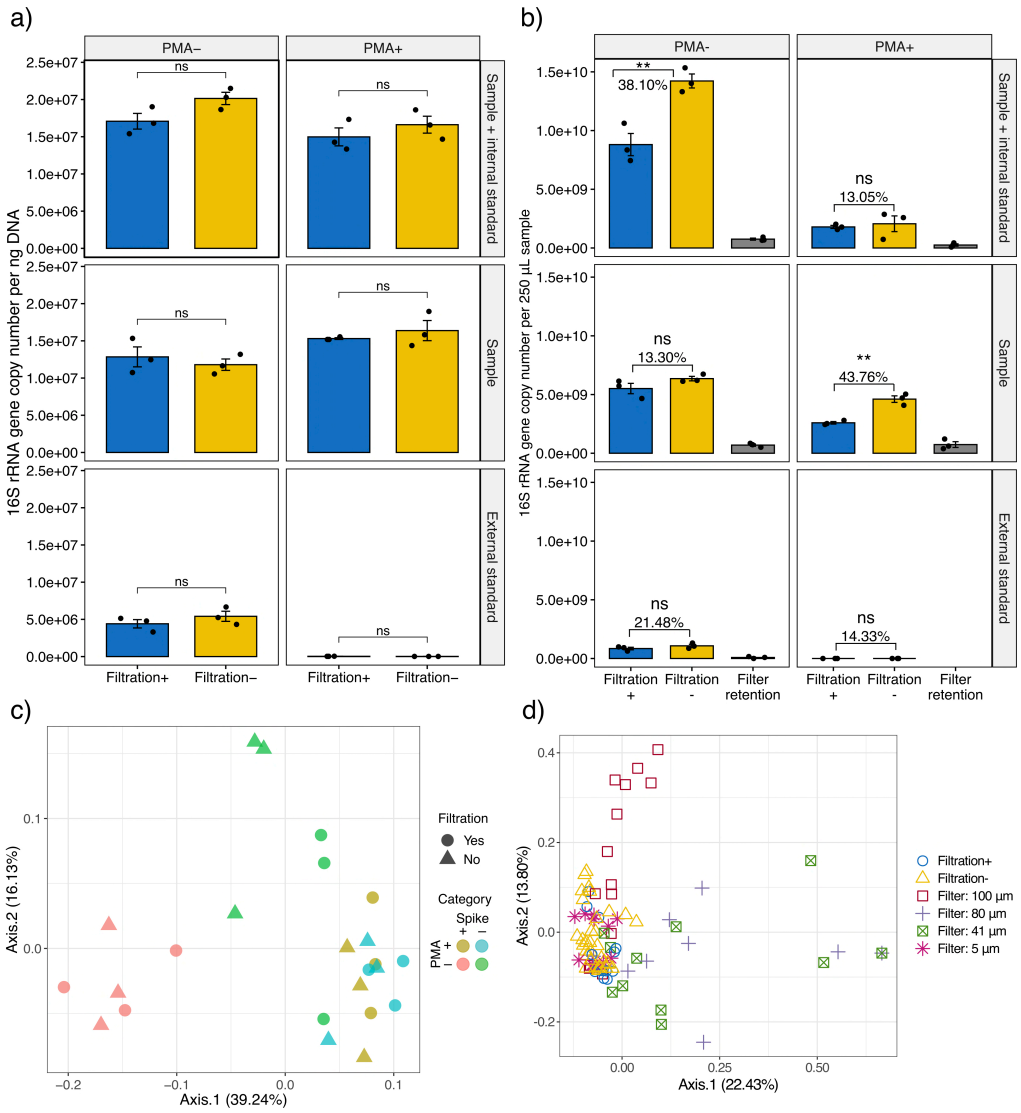


Fig. 3.2 Effects of sequential filtration on hospital-associated surface samples. a) Bacteria proportion was not significantly increased after filtration according to paired t tests. b) Biomass of samples with and without filtration as well as retained by filters according to 16S rRNA gene copy number. In a) and b), error bars represent the mean standard error of triplicates. Filter retention includes all biomass captured by 100, 80, 41, and 5 µm filters. Ns and ** are significance codes, representing $p > 0.05$ and $0.001 < p \leq 0.01$, respectively. A linear scale was used for both a) and b) because for a), a linear scale is more conservative than a log scale when no significant difference was concluded; for b), linear-scale biomass loss is more informative for metagenomic sequencing. c) Principal coordinate analysis using Jaccard distance metric among samples with and without filtration. d) Principal coordinate analysis based on Jaccard distance metric revealed that bacterial profiles retained on 5 µm filters clustered together with liquid samples, while those on 100, 80, and 41 µm filters were away from the major group.

3.4.2 PMA and cultivation improve the ability to determine viability

We examined the efficiency of PMA treatment coupled with metagenomic sequencing (PMA-MetaSeq) on hospital-associated surface samples with the ZymoBIOMICS Microbial Community spike-in as the internal standard. The Zymo community consists of 8 bacterial species and 2 yeasts, which presumably will function more comprehensively and accurately regarding bias correction and quality control than a single-species standard [40, 99, 103, 144]. Sequencing outcomes were compared with cultivation results for benchmarking, as cultivation is the gold standard for determining microbial viability.

Absolute abundance of samples decreased after PMA treatment, indicating the depletion of non-viable signals (Fig. 3.3c). This was further supported by the observation that **α** diversity was lower in PMA-treated samples (Fig. 3.3a), and that inter distances between paired samples were larger than intra distances within each sample group (Jaccard Distance; Fig. 3.3b). We note that absolute abundance should be used when analyzing sequence data involving viability assessment, as relative abundance profile is likely distorted (Fig. 3.3c-d) [109]. Although relative abundance is informative in demonstrating presence/absence, it neglects the amount of overall biomass and thus may inflate the apparent abundance of even low-abundance organisms. While absolute abundance is more reflective of reality, field trials are necessary to determine whether absolute or relative abundance, or either, can be linked to infection or other clinical outcomes.

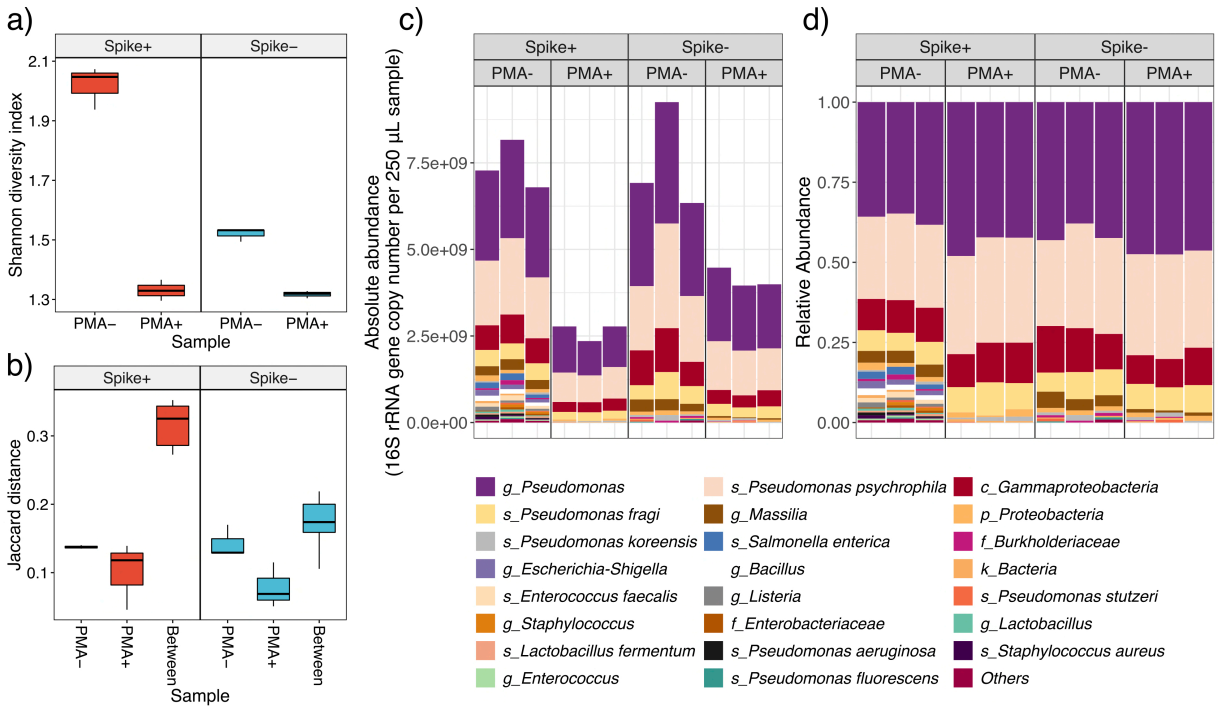


Fig. 3.3 Effects of PMA treatment on hospital-associated surface samples. a) PMA treated samples had lower α diversity based on Shannon index. b) Inter distances between paired samples with and without PMA treatment were larger than intra distances within each sample group (based on Jaccard metric). Comparisons of profiling the bacterial composition by c) absolute abundance and c) relative abundance.

We calculated the efficacy of 8 spike-in bacteria [103]. The efficacy should be 1 under ideal conditions, given that the percentage of viable microbes in the Zymo community is negligible (Fig. S3.8). The efficacy equaled 1 for all taxa, suggesting that PMA treatment is effective in low biomass samples regardless of taxonomy. This conclusion is partly consistent with our previous evaluation of PMA-Seq where *E. coli* was spiked in [103].

Focusing on specific taxa (Fig. 3.4b), we observed occasions of a complete depletion for high-absolute-abundance taxa and retention for low-absolute-abundance taxa, suggesting an effective viability distinction. Relative abundance for some taxa increased after PMA treatment (*g_Pseudomonas*, *s_Pseudomonas psychrophila*, *c_Gammaproteobacteria*, *s_Pseudomonas*

fragi, *s_Pseudomonas koreensis*, *k_Bacteria*), while all taxa showed a decrease in absolute abundance. This indicates that PMA treatment may increase the ability to detect taxa with majoritarily viable populations. We did not detect new taxa that were previously undetectable in PMA-treated samples. However, if nonviable microbes are not of interest, treating samples with PMA will improve the detection power for the overall community with comparable sequencing resources (Fig. 3.6c), as it reduced the overall **α** diversity (Fig. 3.3a).

Cross referencing between cultivation and PMA-MetaSeq was greatly impeded by their inherent limitations (e.g., detection limit for MetaSeq, biases with bioinformatics; viable but non-culturable cells for cultivation). Even for PMA-untreated samples, cultivation and MetaSeq only agreed with each other on a small number of taxa (Fig. 3.4a). Among the 3 viable taxa confirmed by cultivation, viability of *s_Pseudomonas fragi* and *s_Pseudomonas stutzeri* was reflected by PMA-MetaSeq, while *s_Pseudomonas fluorescens* became undetectable after PMA treatment. This might imply over-depletion, but could also be because its abundance went below the detection limit of MetaSeq. As indicated by Barbau-Piednoir et al., less-abundant taxa were more likely to be eliminated (to undetectable) by PMA treatment [145]. This is consistent with our observation, as the abundance of *s_Pseudomonas fluorescens* was the smallest among the cultivation-confirmed taxa. Thus for low-abundance taxa, cultivation could serve as a supplement to sequence-based viability assessment techniques, as a small unintentional removal of viable cells may lead to a large presence/absence difference. Moreover, incorporating cultivation can expand the detection spectrum in general, and particularly for low-abundance taxa, due to MetaSeq's restrictions such as detection limit and failure to distinguish closely related taxa.

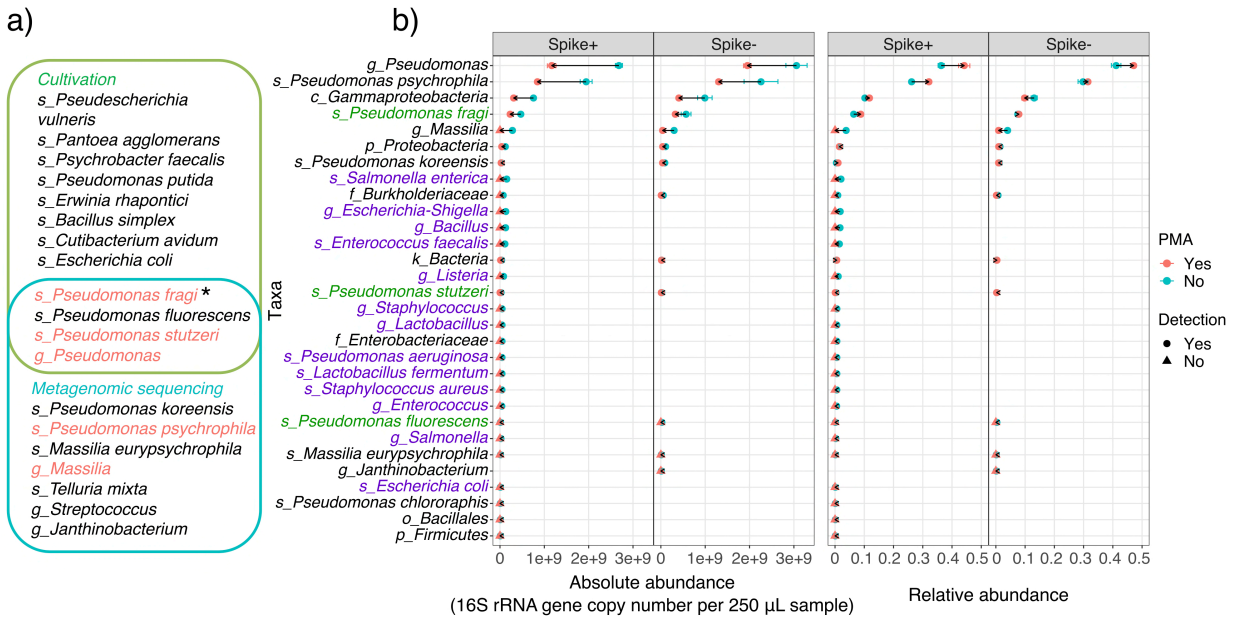


Fig. 3.4 Performance of cultivation and PMA-MetaSeq in viability distinction of hospital-associated surface samples. a) Venn diagram showing the detected taxa by cultivation and MetaSeq. Taxa detected by PMA-MetaSeq are color coded in red. *Pseudomonas fragi* was also detected in the genome-centric approach and is marked with an asterisk. b) The abundance change of all taxa detected under the framework of absolute abundance and relative abundance. Taxa in the theoretical composition of the internal standard and recovered in cultivation are color coded in purple and green, respectively. The Y axis follows a descending order of the average abundance across samples. Error bars represent the mean standard error of triplicates.

Collectively, we emphasize the importance of using absolute abundance and demonstrate a successful application of multi-species internal standards in PMA-MetaSeq. PMA is effective in low biomass samples and can improve the detection power by eliminating irrelevant signals. Cultivation remains a valuable supplement to sequence-based techniques for capturing a comprehensive picture.

3.4.3 Poor taxonomic classification is a major hurdle for internal standards in quantitative metagenomics

Quantifying metagenomics-based abundances using internal standards has substantial benefits. Theoretically, addition of internal standards could compensate for errors resulting from non-

quantitative steps [99]. *E. coli* is one of the most used spike-in strains, in part because it is well-studied and easy to recover in sequencing [103]. However, ideally, we want the internal standard to contain a set of diverse taxa, so that it well represents the diversity in microbial communities. We investigated the performance of the Zymo community as the internal standard for hospital-associated environmental samples, along with qPCR for the 16S rRNA gene.

Unfortunately, the efficiency of implementing the Zymo standard in quantitative metagenomics was drastically impeded by the limited resolution of taxonomic classification. We tried two approaches: Metaxa2 [124-126] coupled with the SILVA 132 SSU database [128, 146] and MetaPhlAn3 [130, 131], which uses a collection of marker genes. The taxonomic resolution varied substantially across different taxa. For samples containing only the Zymo standard, 85% of the small subunit rRNA reads were attributable by Metaxa2, while only 48.14% of the metagenomes were recognized by MetaPhlAn3. Within the attributable portion, MetaPhlAn3 performed better regarding specificity; all reads were classified at the species level, while Metaxa2 retained a decent amount of information at higher levels, with the ratio of genus/family-level and species-level classifications ranging from 0.18 to 11.28.

Foreseeably, this issue will be alleviated as reference databases and taxonomic assignment tools continue to advance. Currently, advantages of internal standards are mainly reflected when species-level identification is the major focus. For instance, clinical samples usually target pathogenic species whose core pangenomes are relatively well represented in databases. In this case, the biases from uneven representation of species can also be corrected based on the performance of closely related internal-standard species. However, if information at genus or higher levels is of consideration, internal-standard techniques become non-applicable, as we are

not able to distinguish internal-standard taxa from other species within the same genus (or at higher levels), which is the basis of making calculations and corrections. Coupling with qPCR, instead, is more appropriate (Fig. 3.3c). Environmental communities are typical examples where coupling with qPCR stands out because environmental microorganisms are not usually well represented at the species level. Of 87 samples in our study, strikingly, MetaPhlAn3 only recognized an average of 19.68% of the metagenome at the species level. The classification rate slightly increased to 38.40% using Metaxa2, which substantially improved to 87.24% when genus level was included.

3.4.4 Accessible sample features can predict required sequencing effort

To enable more informed decision-making before MetaSeq, we conducted a quasi-meta-analysis, using the limited number of existing hospital-related environmental metagenomics studies [14, 15, 90, 94, 101, 112]. We recruited 956 shotgun samples (874 from 6 previous studies and 82 from this study) (Table S3.1). Using these data, we linked accessible features (e.g., location, building, sampling method) to the required sequencing effort given a targeted coverage, leveraging machine learning-based models and Nonpareil (Fig. 3.6a).

Relationships between Nonpareil diversity (Nd, unit: log-bp) and metadata features were first explored (Fig. 3.6a: stage one). Nd is an index measuring the complexity of a microbial community regarding "sequence space", which correlates with classic bin-based diversity indices (e.g., Shannon index) for bacteria [52, 111]. Though not passing the normality test (Shapiro-Wilk Test, $p = 3.338e\text{-}16$) [147], normal distribution was still the best-fit distribution of our dataset, followed by logistic distribution, upon investigation by Cullen and Frey Graph and R package “fitdistrplus” (Fig. 3.5a, Fig. S3.9). Presumably, the deviation from normality will decrease as

sample sizes increase. For 90% of samples, Nd was within 2 orders of magnitude (15.4-20.0, natural log scale), suggesting a common range for hospital-associated environmental samples, which is valuable for reference when designing future studies. Notably, this Nd level was among the lowest across 6 different environments including animal hosts, fresh water, and soil (Fig. 3.5b) [111].

We further examined the influences of sample type (sink versus surface), sampling method and sample pooling on Nd. No significant difference was observed between sink and surface samples (Fig. 3.5c). Within sink samples, Nd was significantly different across sampling methods (Anova, $p = 7.18e-06$). Specifically, samples collected by swabs seemed to have a smaller diversity than those by the other methods (Tukey's post-hoc test; samples without a clear collection method stated in the original paper were assigned as "Sink"). Note that even though sink samples are generally from the same location, the confounding effects introduced by sub-locations (e.g., sink basin, pipe edge, p trap) cannot be ruled out. Similarly, within surface samples, though Nd of wipes was significantly larger than that of swabs (unpaired t-test), confounding effects remain (e.g., researchers tend to use wipes for large-area and high-biomass locations, like floors, which often contain more diverse communities) (Fig. 3.5d). Though weak, we noticed a trend of diversity increase after sample pooling (Fig. 3.5e), raising the alarm that more caution should be taken when increasing biomass by pooling samples. The practice of sample pooling assumes that the pooled samples share some core features, whose biomass will be increased past the detection limit. This may be true of certain sample types, e.g., host-associated microbiomes, but is unlikely to be true of built-environment samples that lack a conserved core [114]. Further investigations are needed should more data become available, as

the sample size was quite limited for some groups (e.g., n(pooled monitor) = 2, n(not-pooled monitor) = 4). In the interim, we recommend seeking other methods, such as a high-yield DNA extraction, before resorting to sample pooling, as the resulting sample characteristics may be different from individual samples.

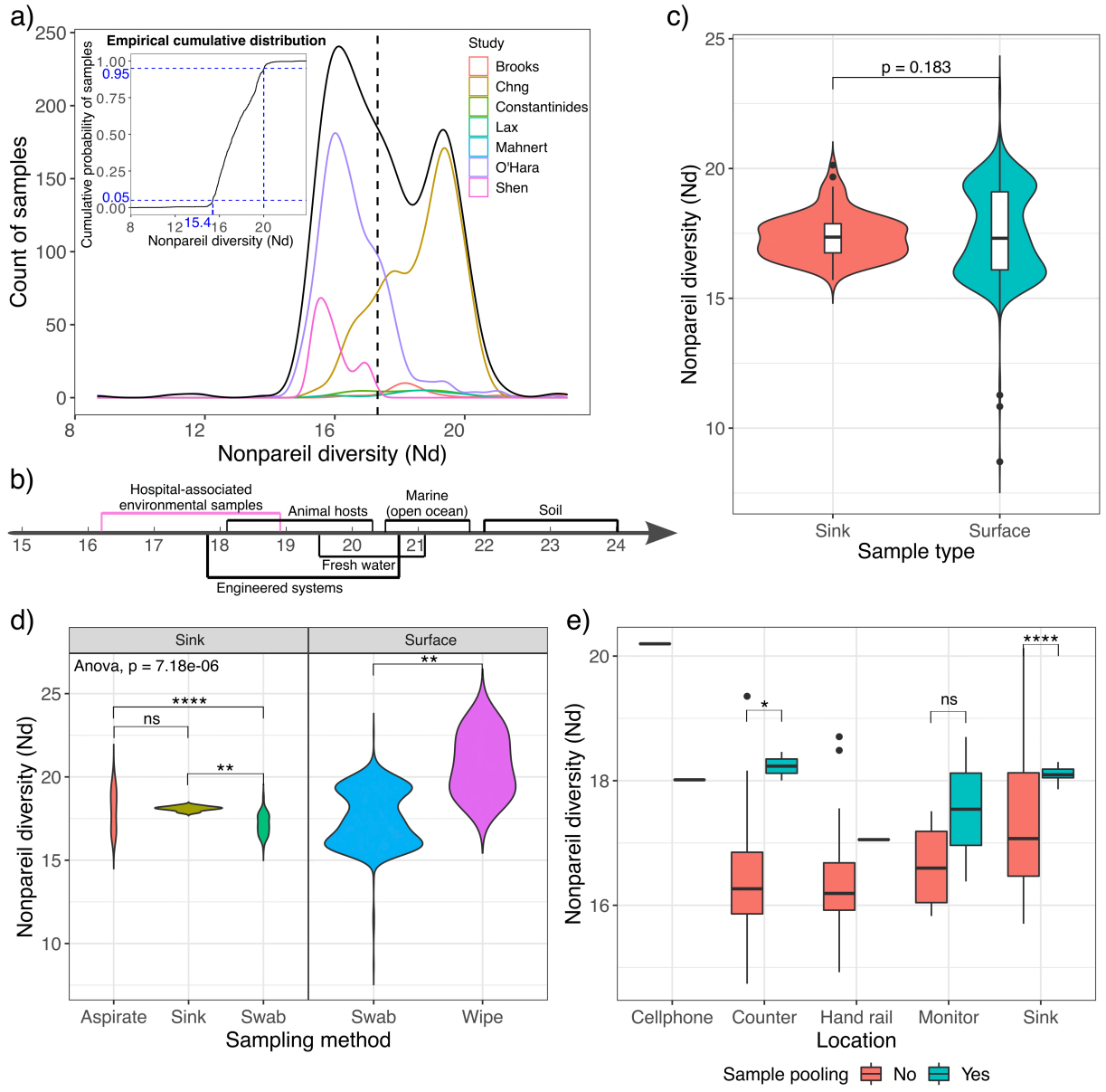


Fig. 3.5 Relationships between Nonpareil diversity and metadata features for hospital-associated surface samples. a) Overall distribution of Nonpareil diversity (black) and distributions for individual studies. b) Interquartile range of Nonpareil diversity for microbiome samples from different environments. This study is color coded in orchid. Effects of c) sample type, d) sampling method, and e) sample pooling on Nonpareil diversity. Significance codes are as follows: $p > 0.05$ (ns), $0.01 < p \leqslant 0.05$ (*), $0.001 < p \leqslant 0.01$ (**), $p \leqslant 0.001$ (***).

To further harness the reference potential of Nd, we built models to predict Nd from metadata features based on machine learning algorithms. Eight predictor variables (location, building, study, country, touch frequency, sample type, sampling method, sample pooling) were included based on data availability, MIxS-BE standards, and previous experience (Table S3.1) [112, 119, 148]. Nd, the response variable, was first converted from a numerical variable to a nominal variable. Three conversion schemes were tried, with the intervals being 2.5, 1.0, and 0.5 (number of categories being 2, 5, 11, respectively). Random sampling was adopted to split the entire dataset into training and testing datasets at the ratio of 4:1. Implementing repeated cross-validation (5 folds, 5 times) on the training dataset, 9 algorithms were examined to optimize the classification performance, including random forest, stochastic gradient boosting, and support vector machines. Algorithms were evaluated according to 4 metrics (area under curve, Kappa, and balanced accuracy on both training and testing datasets) [149]. Overall, no difference was observed among the tested algorithms. Random forest was selected due to its slightly better performance from a holistic perspective and capability of ranking the predictor variables.

The model accuracy positively correlated with the interval size. At 2.5, the accuracy on the training dataset was as high as 87.69%, and slightly lower on the testing dataset (82.60%). The accuracy dropped as the classification demand rose. The mean balanced accuracy on the testing dataset was 64.08% and 61.38% when intervals were 1.0 and 0.5, respectively. Considering that Nd was converted from a continuous variable, we examined the misclassifications and found that most of them fell into nearby categories. We thus calculated the mean balanced accuracy ± 1 category and observed a substantial improvement. Specifically, mean balanced accuracy of 87.06% and 77.17% can be achieved for 5- and 11-category classifications, respectively.

Considering the sparsity of the currently available dataset and the challenge of multiclass classification, this model demonstrated a reasonable degree of accuracy, which should improve as sample sizes and available features grow.

The variable importance ranking generated by random forest separated the predictor variables into 3 groups (Fig. 3.6b). Location, building, and study were the top 3 variables with the highest importance, followed by country, touch frequency, and sample type, while sample pooling and sampling method hardly impacted the classification. Group-wise, this ranking was generally consistent with the explanatory power described by linear regression (Fig. S3.10) [119]. That "study" ranked as one of the most important variables indicated the existence of biases towards individual studies in the current dataset (e.g., "batch effects" related to respective sampling, processing, sequencing, and analysis), which was also observed by a previous meta-analysis of indoor microbiota [119]. Interestingly, despite its high importance, the model performance had almost no drop after excluding "study" (> 95% for all conversion schemes), justifying making predictions without involving artificial metadata features like "study". It is worth noting that importance of the other variables (e.g., building, country) was raised after this exclusion (Fig. S3.11). To find the features necessary for making a comparably accurate prediction, we further examined the performance of models after gradually reducing the number of predictor variables, and found that using only 2, "location" and "building", the new model achieved 95% accuracy regardless of interval sizes tested.

With Nd and metadata features connected, the required sequencing effort at a targeted coverage was then inferred (Fig. 3.6a: stage two). Upon fitting the data, we revealed a linear relationship between the natural log of estimated sequencing effort at 95% coverage (ln(LRstar)) and Nd,

with the equation being ln(LRstar)=1.14*Nd+1.21 (Adjusted R-squared = 0.6012, p < 2.2e-16) (Fig. 3.6c). This is theoretically backed up by previous findings that sequencing effort depends on the diversity level and the genome size, and that the latter can be ignored for most microbial communities, particularly bacterial communities, since the differences in genome size are usually no more than one order of magnitude [52]. Instructions to make calculations between sequencing effort and other coverage levels are provided at https://github.com/jxshen311/workflow_metagenomic_environmental_surveillance/tree/main/nonpareil/example_SeqEffort%26Coverage.

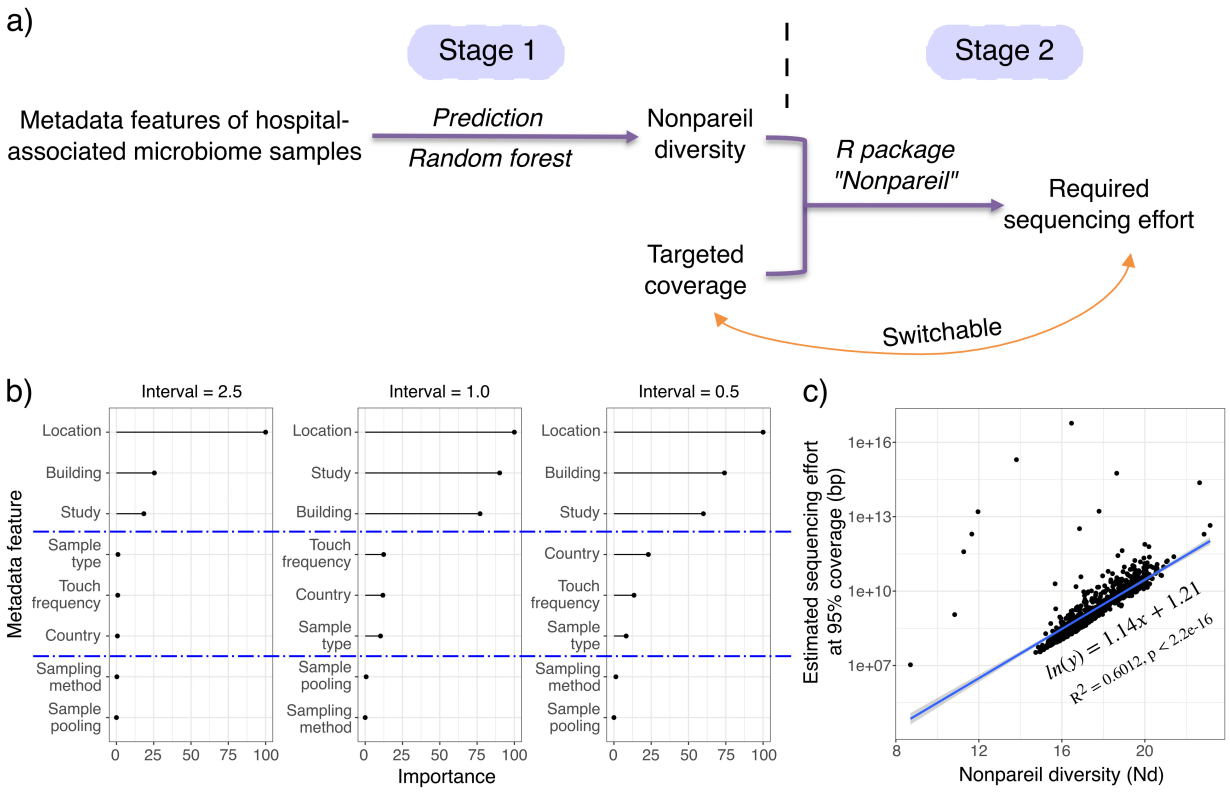


Fig. 3.6 Required sequencing effort can be predicted by accessible sample features and targeted coverage. a) Workflow of making the prediction. b) Variable importance rankings by random forest. c) The natural log of estimated sequencing effort at 95% coverage is linearly correlated with Nonpareil diversity.

3.5 Discussion

Although sequence-based environmental surveillance of microbial communities for better management of public health has been appealed for and utilized, best practices of the workflow have not been systematically studied to ensure proper interpretations of sequencing results to aid in infection risk assessment [50, 99]. This study introduces a well-structured and informed metagenomics-based workflow towards the goal of being appropriate for low-biomass, viability, quantification, and resource estimation. We recommend adopting liquid-liquid extraction to improve DNA yield and only incorporating whole-cell filtration when non-bacterial proportion is large. Despite its imperfection, we suggest including PMA treatment, and involving cultivation when demanding comprehensive profiling. We further recommend integrating internal standards for quantification, and additionally qPCR when we expect poor taxonomic classification. We also introduce a machine learning-based model to predict required sequencing effort from accessible sample features. The model helps make full use of sequencing resources and achieve desired outcomes.

While using realistic samples in testing simulates conditions the workflow may face in practical applications, it comes with side effects. Our aggregation sample had a small fraction of non-bacterial organisms (~1%). Thus, the conclusion that whole-cell filtration does not increase the bacterial proportion and signal of rare taxa to a statistically significant degree is probably only applicable to samples with similar characteristics, representing 84.90% among the 874 samples from hospital-related environmental studies used in the quasi-meta-analysis [14, 15, 90, 94, 101, 112]. However, a few samples did contain a decent proportion of eukaryotes. Specifically, 132 samples harbored more than 1% eukaryotic reads, and strikingly more than half reads were

attributed to eukaryotes for 20 samples. Moreover, samples collected from high-touch surfaces were more likely to have higher proportions of eukaryotes than low-touch surfaces and sinks. Of the 104 sink samples, the maximum percentage of eukaryotes was 0.1%. Therefore, filtration is probably unnecessary for most environmental samples (especially sink samples) and may be beneficial for part of high-touch surface samples (Fig. S3.12).

Despite being semi-quantitative and entailing considerable uncertainty, involving PMA takes us a step closer to understanding viability, particularly for low biomass samples whose complexity is also relatively low [103]. Notably, the overall uncertainty comes not only from PMA treatment but also from the metagenomics pipeline itself, such as biases from DNA extraction kits and taxonomic assignment tools [60, 150]. In addition to PMA, alternative metrics have been proposed, including methods based on RNA (reflects active transcription), peak-to-trough ratio (PTR) (reflects active replication), and nuclease digestion (e.g., benzonase). As stated in a systematic evaluation, while 16S rRNA transcript-based amplicon sequencing semi-quantified viability of synthetically constructed simple communities (*Escherichia coli* and *Streptococcus sanguinis*), it is inappropriate for realistic complex communities [151]. PTR has been demonstrated as an efficacious metric to estimate microbial growth rates in both human (e.g., skin, fecal) and environmental (e.g., marine, sludge) datasets by several studies [152-155]. However, a study based on freshly collected marine prokaryotes raised concerns as they observed poor correlations between PTR and growth rates for most marine bacterial populations ($r \sim -0.26–0.08$), except for the rapidly growing γ-Proteobacteria ($r \sim 0.63–0.92$) [156].

Some overlap exists between methods for viability determination and those for depleting eukaryotic DNA. For example, osmotic lysis followed by PMA treatment is recommended to

remove human DNA in saliva samples [157]. However, recommended methods depend on the sample type. PMA is not recommended for sputum samples, where nuclease-based methods (e.g., digest with benzonase) showed an equal or better performance [158]. Benzonase has also been applied to skin microbiome samples with desired outcomes [159]. In general, factors impacting method performance include percentage and composition (e.g., extracellular DNA, DNA in largely lysed or partially compromised cells) of targets to be removed (i.e., eukaryotes and dead bacteria), as well as characteristics of background matrix (e.g., viscosity). For instance, saliva and sputum consistently contain ≳ 90% human DNA [157, 158], while this percentage is very diverse for hospital-associated environmental samples (Fig. S3.12). Filtration failed to exclude human DNA in saliva likely because extracellular DNA was the dominant component rather than cells [157]. For sputum samples where cells are lysed and DNA is no longer protected, nucleases might be quite effective in depleting extracellular DNA, whereas PMA efficacy could be hindered by the viscosity of the matrix) [158]. In contrast, in environments where cells gradually decay due to harsh conditions (e.g., desiccation), more DNA attributable to dead cells would still have a partially compromised membrane; PMA, as a small molecule, may be more effective in penetrating the damaged cell membrane and depleting the DNA. For eukaryotic depletion, it may be beneficial to further unravel the underlying mechanisms influencing the efficacy of different methods in different sample types and characteristics.

Nevertheless, for viability assessment, instead of focusing on this viable/dead dichotomy, perhaps more critically, we should keep in mind that “viability” is rather an intermediate or methodological term, linking surveillance results to questions of interest (e.g., which bacteria are infectious) [99]. In the future, it is worth exploring whether the concept “viability” in the context

of interest is closer to intact cell membrane, active transcription, or active replication. Moreover, rather than optimize one single metric, integration of multiple methods has been proposed (e.g., using multi-omics techniques) [160]. Pursuing viability profiles using orthogonal methods would plausibly enable a more comprehensive understanding, but the cost-benefit ratio may be considerably high for multi-omics techniques. Integrating with cultivation, instead, provides an affordable alternative. Notwithstanding, it remains to be investigated how to properly interpret results generated by a combination of methods, as inconsistencies between disparate methods are common.

We applied multi-taxa internal standards and calculated PMA efficacy of spike-in taxa based on a reasonable assumption that the percentage of viable microbes in the Zymo community is negligible, resulting in a theoretical value of 1 (Fig. S3.8). While this internal standard can strongly reflect incomplete suppression of non-viable signals, potential toxicity of PMA might be underrepresented. Although no toxicity was observed at the PMA dose of our protocol in validation (Fig. S3.2), a customized internal-standard mixture featuring 0.5 as the designed PMA efficacy would be ideal for future studies [103]. As opposed to purchasing commercial products, we recommend utilizing the Zymo community as a reference for the taxonomic composition and constructing the mixture with live cultures in real-time, because viability (or membrane integrity when PMA is used) is difficult to maintain in manufacturing, shipping, and storage.

Continuous advancement of internal standards for quality control, as well as quantification and other features, is still one of the major hotspots in method optimization. A suitable internal standard should well balance representation and recognizability. Good representation means that the workflow impacts the spike-in and targeted microbes comparably (because of their

similarity). Good recognizability means that the spike-in can be easily distinguished from the targets. In this study, the Zymo community was selected due in large part to its representation, as it spans broadly the phylogenetic tree. Previous studies have selected internal standards based on a similar principle. For instance, the Zymo community and a 10-species mock community were chosen for gastrointestinal and stool samples, respectively [161, 162]. Peroxide-killed *Campylobacter sputorum* was used to quantify viable thermotolerant *Campylobacter* [44]. These internal standards are prone to be confounded with targets, thus posing challenges for bioinformatics to accurately identify and quantify taxa. To obtain good recognizability, exotic materials are sought. In the aforementioned example, the researchers chose 10 species that were generally absent from the stool of healthy individuals. The same criterion was followed by another gut microbiome study in which microbes from hypersaline environments, soil, and plants were utilized [110], as well as a study on Amazon River plume to which genomic DNA from *Thermus thermophilus* HB8 was applied [83]. Finding a completely exotic species is more challenging for environmental surveillance whose subjects are influenced by both human and environmental activities. As a potential solution, artificial DNA have been developed to ensure differentiation from the targets. Previous reports included sets of synthetic DNA, 16S rRNA genes, and chimeric DNA fragments, implemented in different venues of metagenomic and amplicon sequencing [109, 163, 164]. However, whether these exogenous (or even artificial) standards' behavior resembles that of the targets remains questionable. By and large, more systematic evaluation and optimization are needed to foster the development of internal-standard techniques that better balance representation and recognizability, or at least make their pros and

cons quantitatively accessible, both in general and for specific contexts. For example, it will be beneficial to conduct data-driven meta-analyses and curate databases to further inform the field.

Though the classification models performed well from a practical perspective, their accuracy with small intervals still merits improvement. Building a hierarchical classification model might be beneficial, as we observed a drastic increase in the accuracy when the interval size was enlarged. It is also likely that the available dataset is not good enough to train a model with very high accuracy. For example, there is clear evidence that the data were biased by the disparate sample sizes between studies. Moreover, we only managed to collect 7 common metadata features (excluding "study") without involving a substantial number of missing values, which raises the question of whether what we achieved has already reached the theoretical plateau of explanatory power of these features. If this is the case, standardized reporting of more high-quality metadata should be further promoted. Additionally, since normal distribution was the best-fit distribution of the current dataset, with seemingly missing pieces in the middle (Fig. 3.5a), fitting data into known distributions may be more explanatory as large sizes of data become accessible.

3.6 Conclusions

This study presents an improved workflow towards accurate and robust healthcare environmental surveillance using metagenomics. The workflow is appropriate for low-biomass samples, distinguishes viability, is quantitative, and enables estimation of necessary sequencing resources. We recommend liquid-liquid extraction, propidium monoazide treatment coupled with internal standards and absolute abundance profiling (e.g., using qPCR), and a machine learning-based

model for sequencing depth calculation. In addition, whole-cell filtration and cultivation may be valuable under particular circumstances.

This metagenomics-based environmental surveillance workflow is particularly useful in infection prevention and disinfection assessment. Although we focus on microbial surveillance of built environments, especially hospital-associated surfaces, the workflow developed in this study can be adapted to other contexts with similar characteristics. For example, the multifaceted lessons learned from this study will benefit the continuing development of microbiome-based clinical testings from body sites (e.g., skin), such as methods to increase low-biomass signals and determine viability [21]. Moreover, the experience gained in overcoming challenges unique to environmental microbiomes (e.g., quantitative metagenomics with poor taxonomic classifications) is also useful to studies on other environments, such as wastewater and air.

3.7 Supplementary information

3.7.1 Supplementary figures

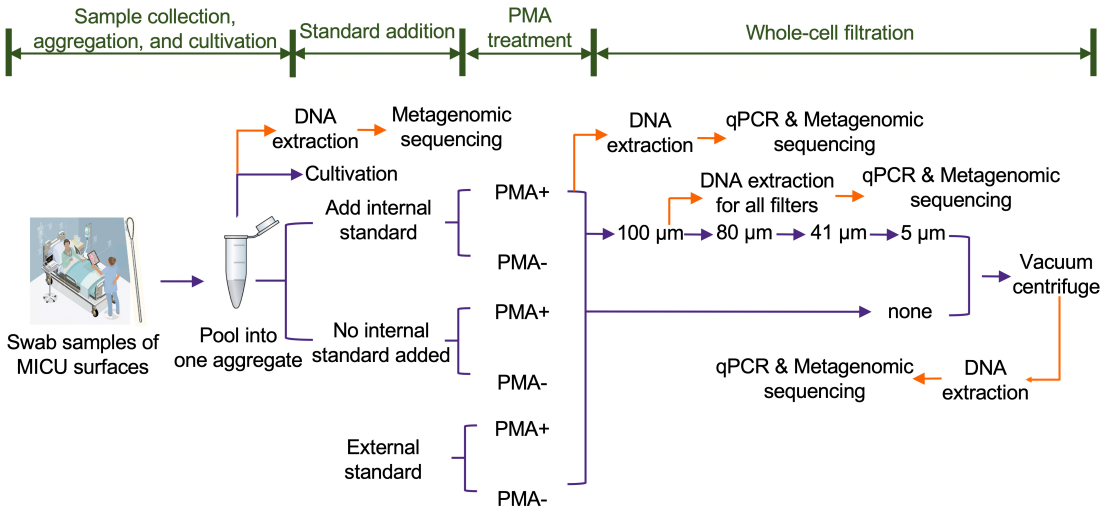
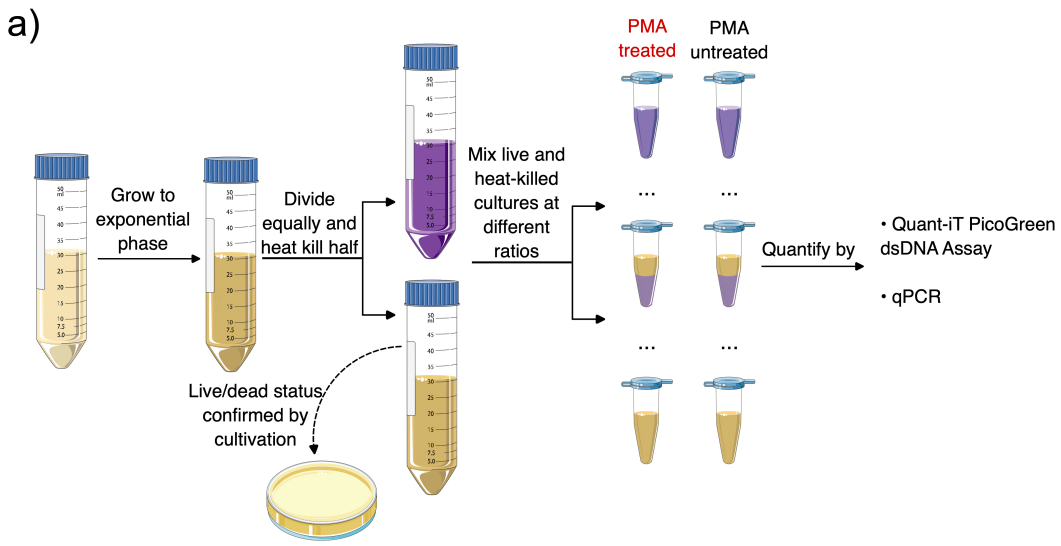
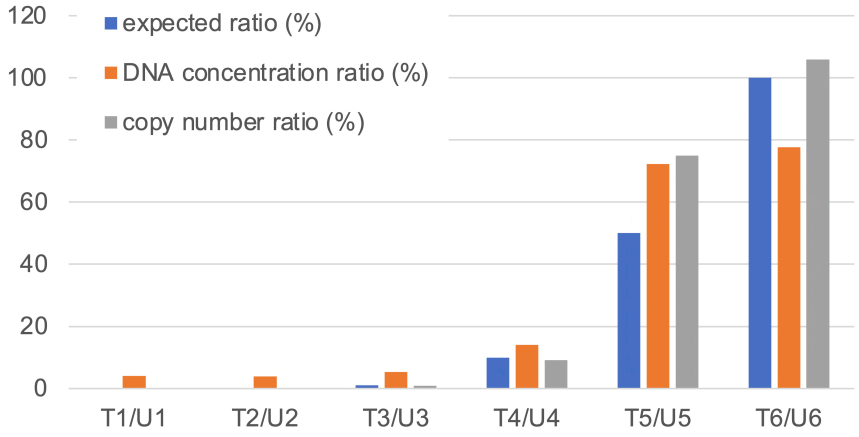


Fig. S3.1 Experimental pipeline for assessing and optimizing techniques in sample treatments.



b)

$$\frac{T}{U} = \frac{\text{PMA } treated}{\text{PMA untreated}}$$



Expected ratio (%)	0	0.1	1	10	50	100
DNA concentration ratio (%)	4.11	3.94	5.39	14.12	72.27	77.72
Copy number ratio (%)	0.07	0.14	0.92	9.08	75.01	105.98

* Copy number ratio is calculated by assuming the efficiency equals 100%.

Fig. S3.2 a) Experimental pipeline and b) result of the PMA validation experiment. We want to acknowledge Servier (https://smart.servier.com/) for providing the following icons (licensed under CC-BY 3.0): microtube-closed-blue, falcon-50ml-pink, and petri-dish-yellow. Colors of the first two icons were changed to adapt to our needs.

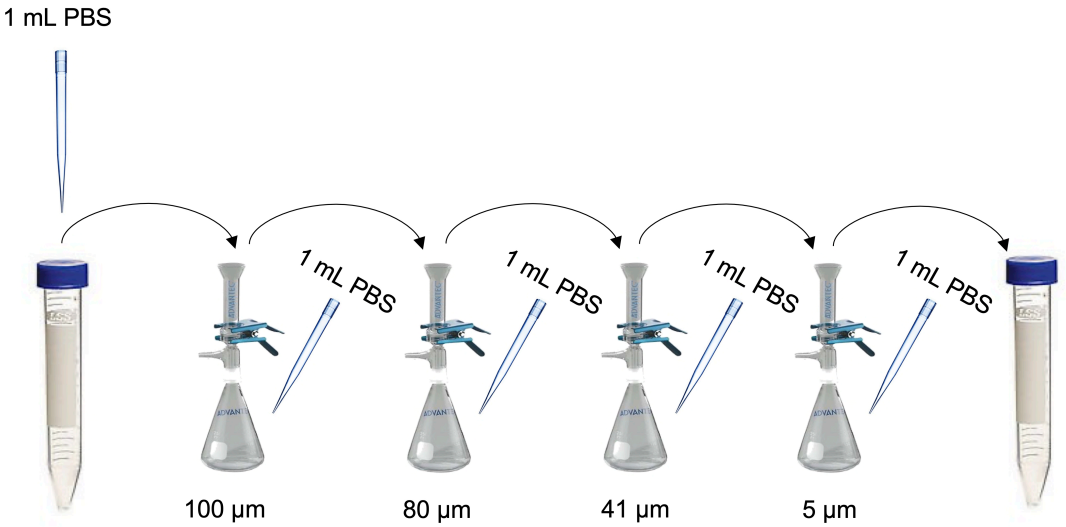


Fig. S3.3 Schematic of the whole-cell filtration workflow.

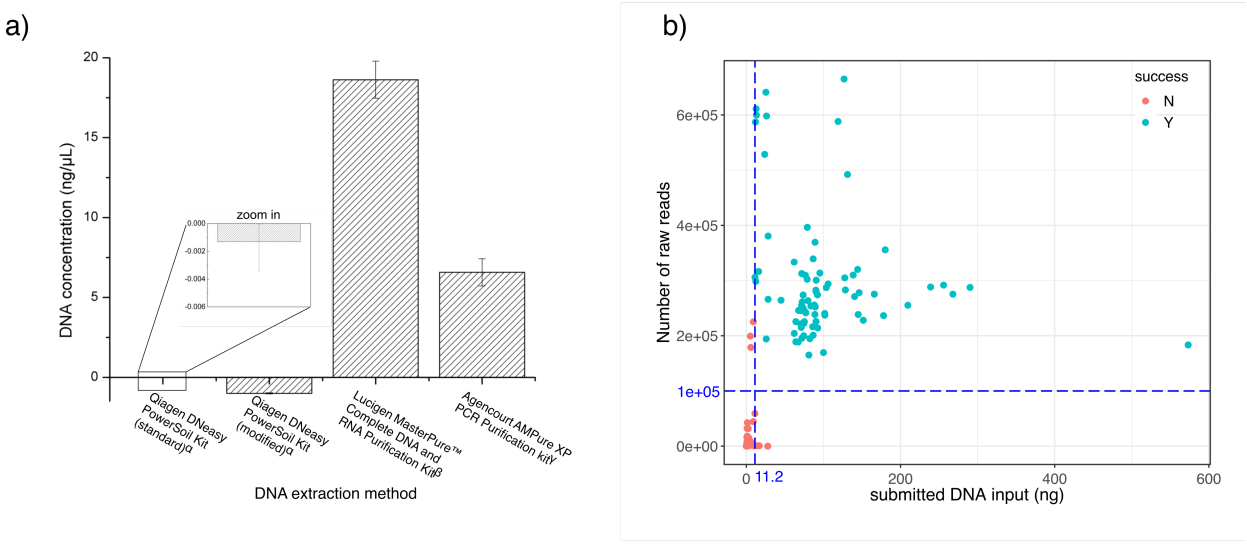


Fig. S3.4 a) DNA yields (quantified by NanoDrop) of low-biomass surface-associated samples extracted by different methods. Extraction kits marked with α are column-based, with β are magnetic bead-based, and with γ use liquid-liquid extraction. b) Practical relationship between reads number and submitted DNA input.

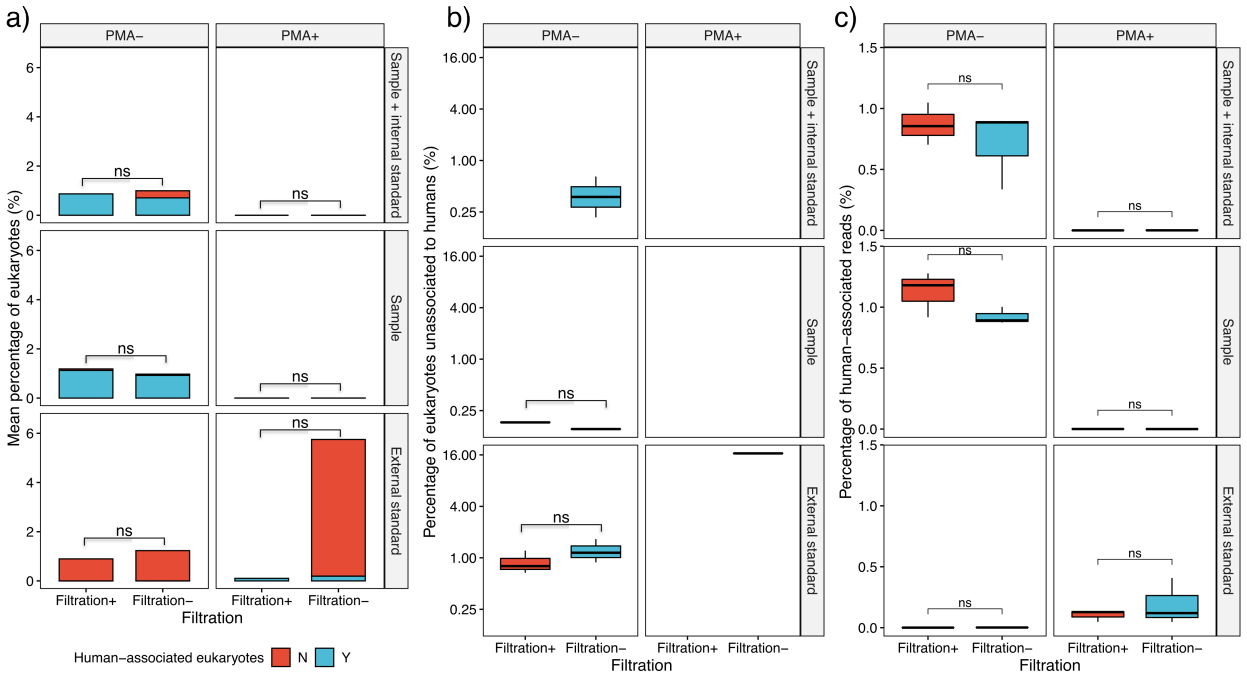


Fig. S3.5 Effects of whole-cell filtration on detected proportions of a) overall eukaryota reads, b) eukaryotes unassociated to humans, and c) human−associated reads. Significance was determined by paired *t*-tests with BH corrections.

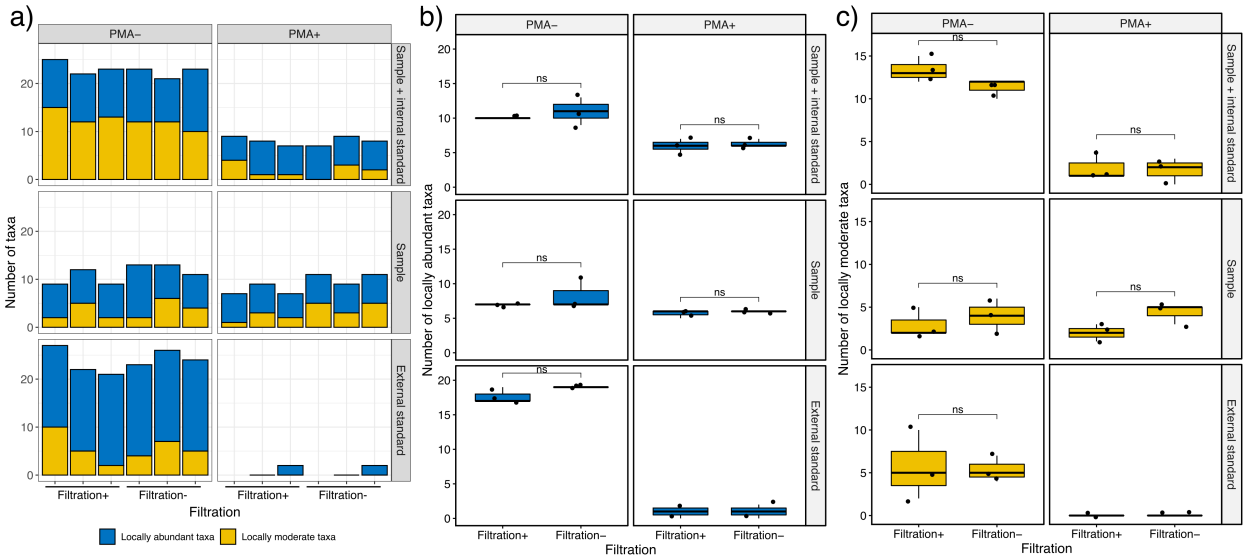


Fig. S3.6 Whole-cell filtration did not have a significant impact on a) the overall number of recoverable taxa, b) the detectable number of locally abundant taxa or c) moderate taxa. Significance was determined by paired *t*-tests with BH corrections. Locally abundant, rare, and moderate taxa were defined as taxa with a representation of ⩾1%, <0.01%, and < 1% & ⩾ 0.01% within a sample.

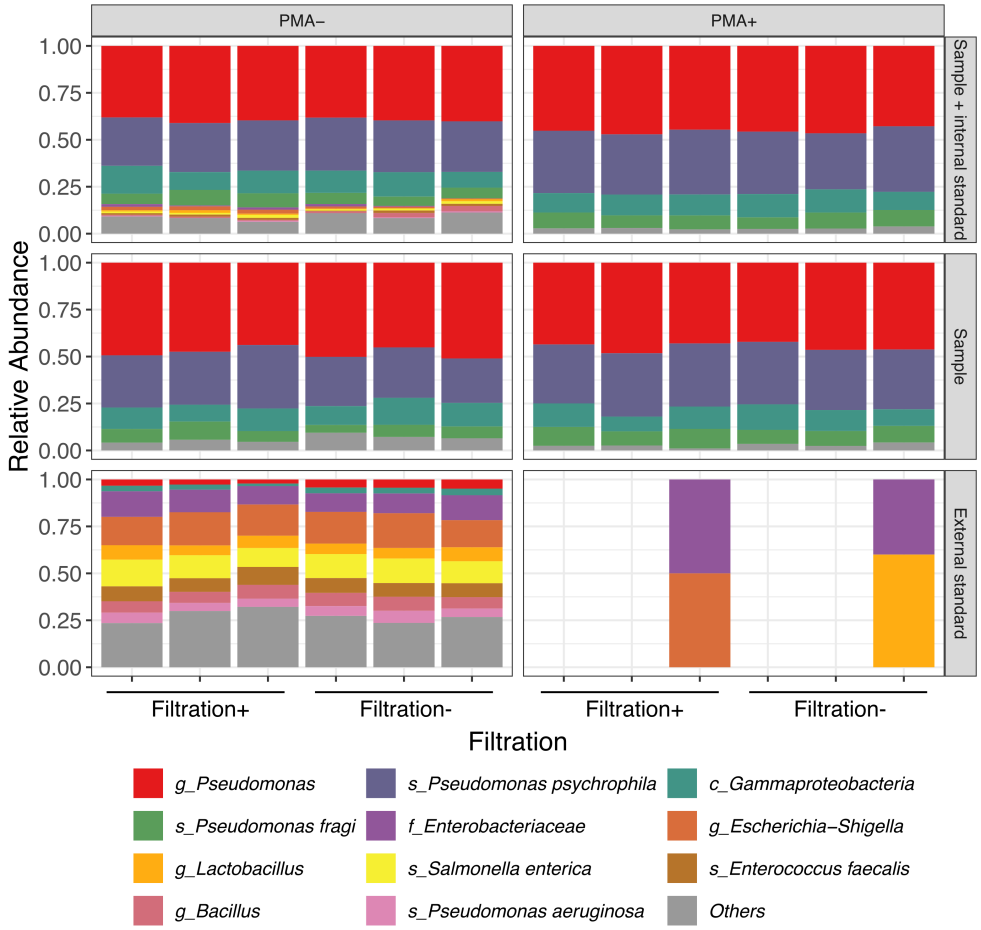


Fig. S3.7 Relative abundance of the top abundant taxa (average abundance > 1%) for samples with and without filtration.

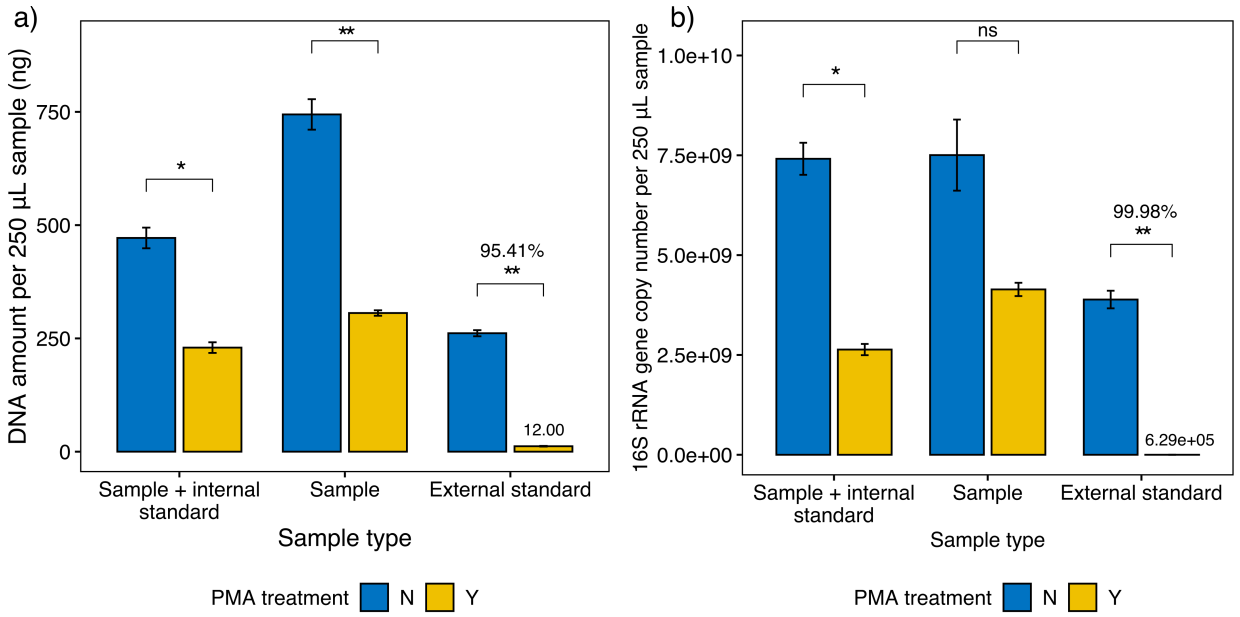


Fig. S3.8 Biomass reduction with PMA treatment for samples (with and without internal standards spiked in) and external standards, according to a) DNA quantity and b) 16S rRNA gene copy number. Significance was determined by paired t tests with BH corrections.

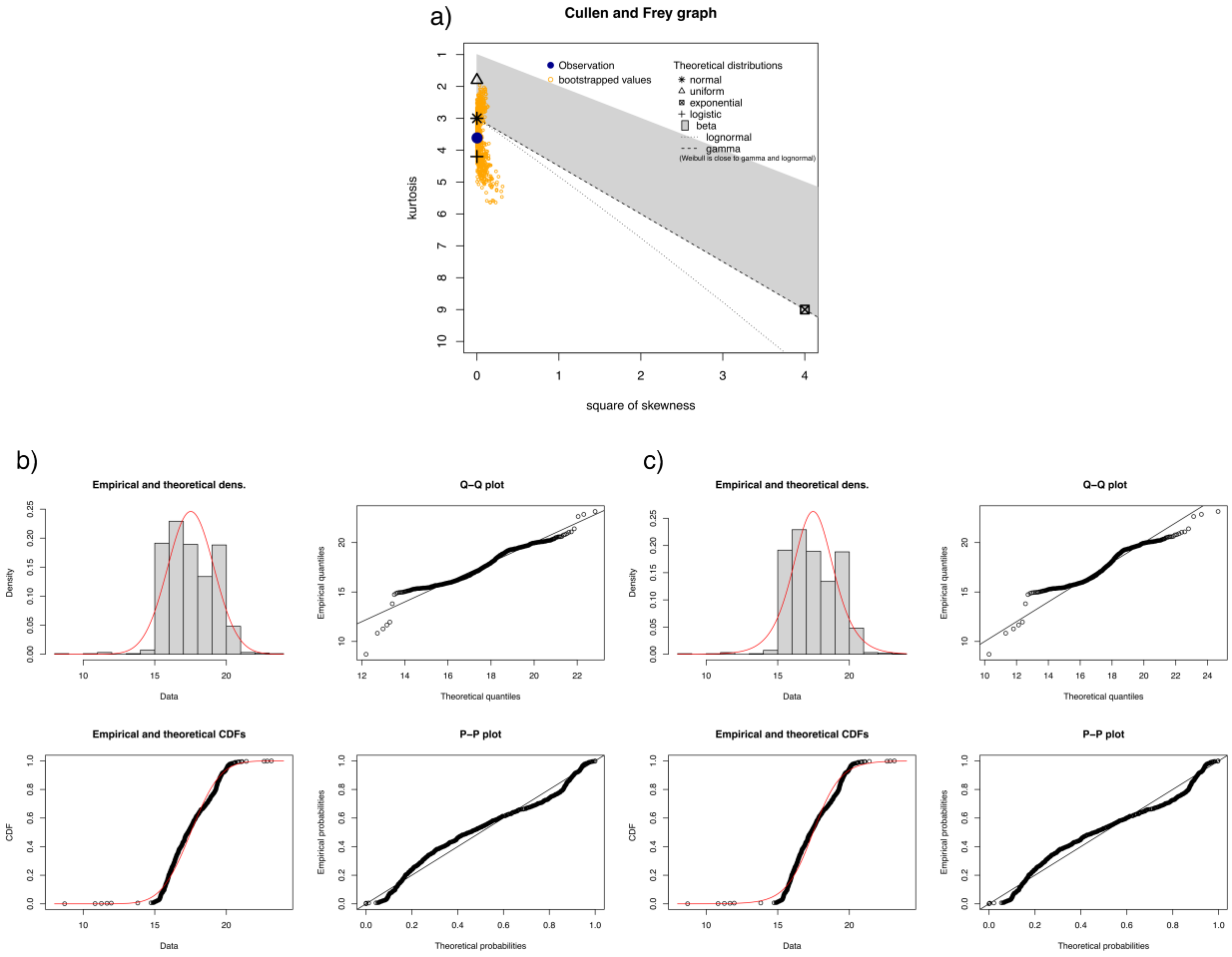


Fig. S3.9 a) Cullen and Frey graph showing the distance from theoretical distributions to the observation. b) Fit for the normal distribution. c) Fit for the logistic distribution.

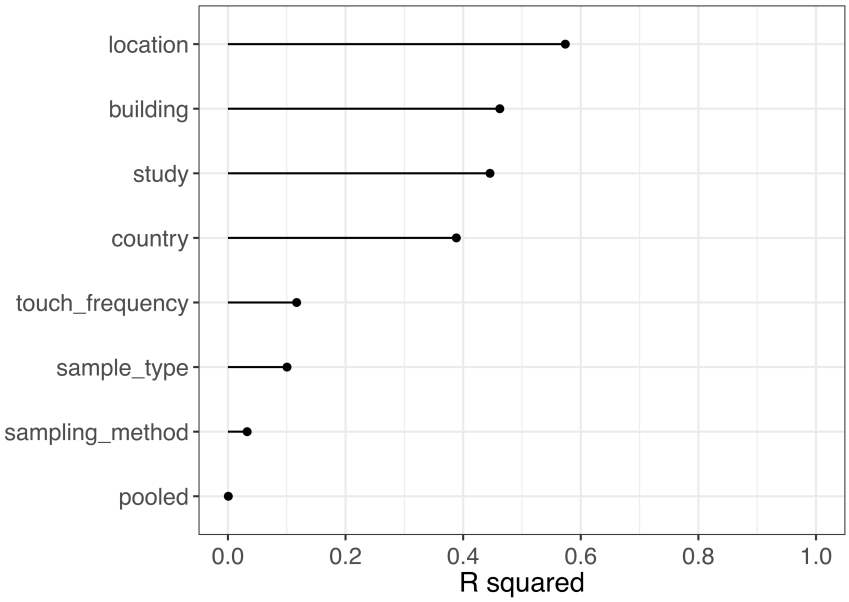


Fig. S3.10 Ranking of variables based on their explanatory power according to the R-squared value of a linear regression model.

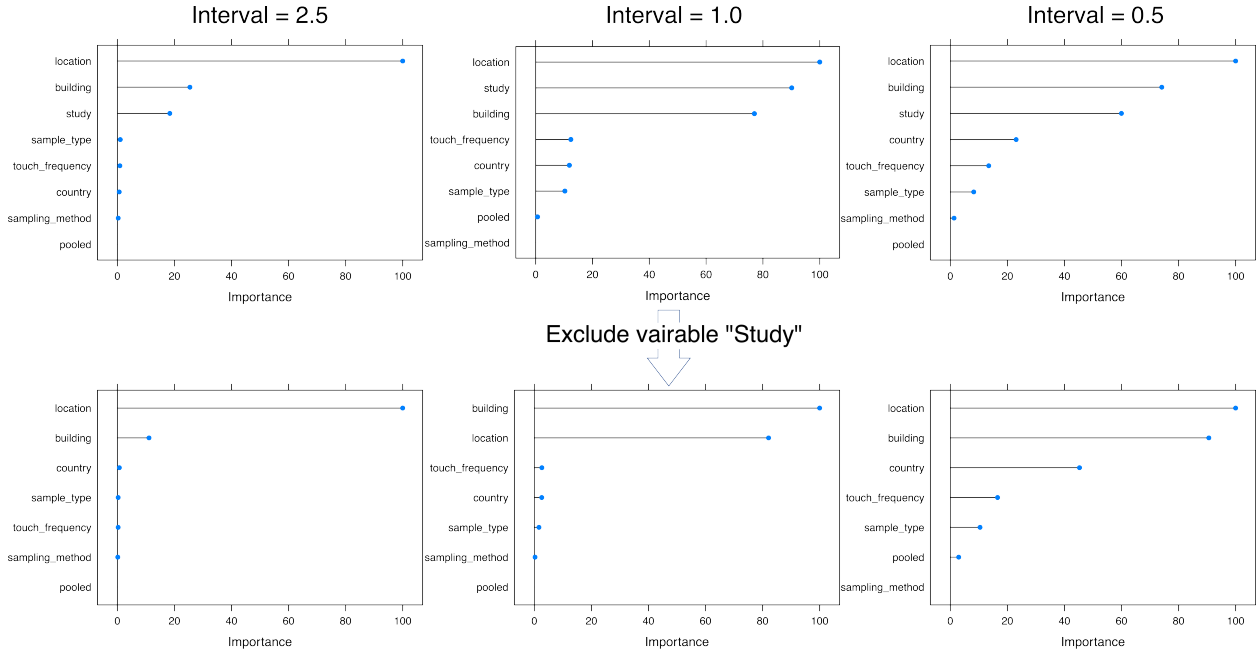


Fig. S3.11 Variable importance rankings with and without the variable "study" based on random forest classifications.

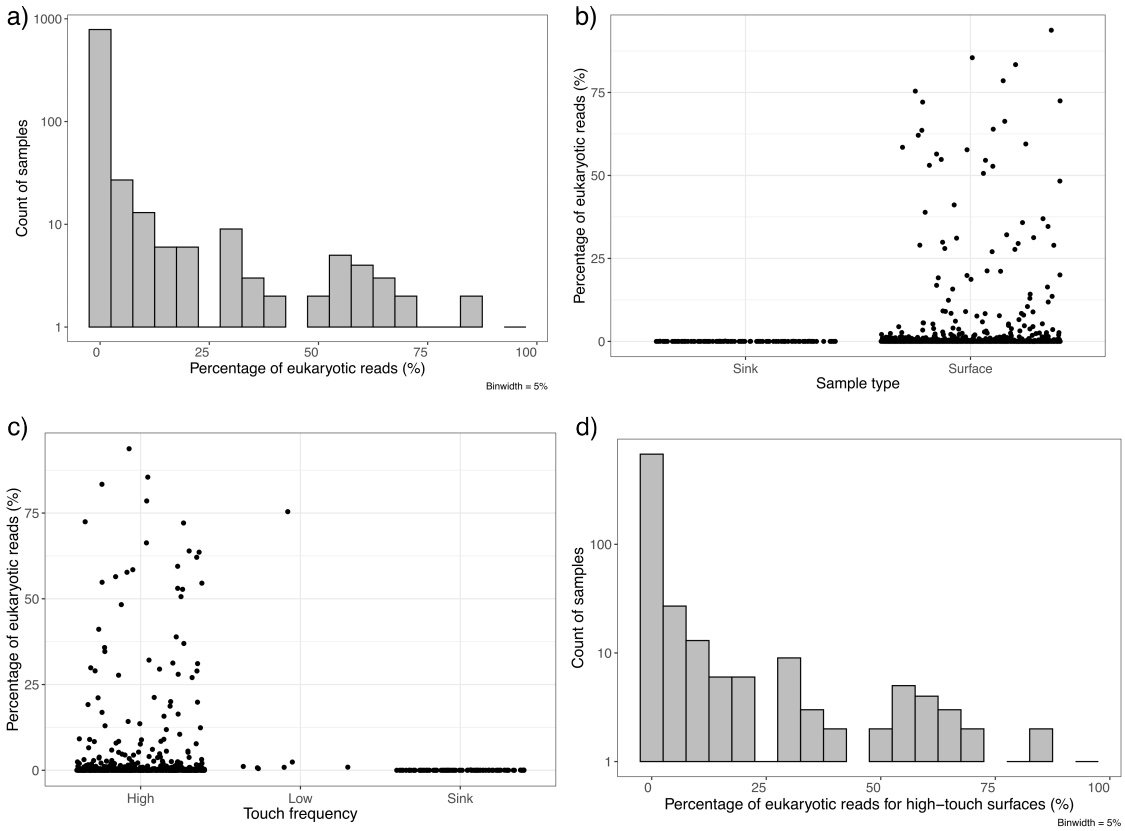


Fig. S3.12 a) Distribution of the percentage of eukaryotic reads among 874 samples from hospital-related environmental studies. b) Surface samples, especially c) high-touch surface samples are more likely to contain higher proportions of eukaryotic reads. d) Distribution of the percentage of eukaryotic reads among 763 samples from high-touch surfaces. Note that the percentage of eukaryota reads may be underestimated because we did not know whether these fastq files accessed from databases had been processed with tools like kneaddata or not.

3.7.2 Supplementary tables

Table S3.1 Dataset of hospital-related environmental metagenomic samples used in the machine learning models.

File: Table S3.1_prediction_dataset.xlsx

Table S3.2 Sample collection details.

File: Table S3.2_sampling_condition.xlsx

Table S3.3 Statistics of the genome-centric approach.

File: Table S3.3_genome-centric_approach_statistics.xlsx

Chapter 4 Hospital environments harbor chlorhexidine tolerant bacteria potentially linked to chlorhexidine persistence in the environment

4.1 Abstract

Background: While chlorhexidine has been used as an antiseptic for decades with little concern for resistance, growing numbers of cases of reduced susceptibility to chlorhexidine are leading to growing concern. Chlorhexidine resistance has even been associated with resistance to clinical antibiotics and healthcare-associated infection outbreaks. The concern is magnified considering the possible dissemination of antimicrobial resistance genes. However, scrutiny of chlorhexidine resistance is still very limited and largely restricted to clinical isolates. While the hospital environment is known to harbor nosocomial pathogens and antibiotic-resistant bacteria, systematic evaluations of chlorhexidine resistance in the hospital environment is lacking.

Results: We investigated chlorhexidine resistance in the hospital environment by combining controlled microcosm experiments and field surveys. Our study revealed that chlorhexidine decreased in concentration while persisting on indoor surfaces. Surface materials and disinfection and cleaning practices impacted the patterns of chlorhexidine persistence. *Ex situ*, we found that clinically relevant bacteria could survive 24 h chlorhexidine digluconate (CHG) exposure on surfaces in the concentration range commensurate with CHG persistence. *In situ*, we found that CHG tolerance was widespread in a medical intensive care unit (MICU) environment. Particularly, we highlight that sinks were a critical reservoir regarding the absolute bioburden, CHG tolerance, and bacteria with exceedingly high CHG minimum inhibitory concentrations (≥ 512 μg/mL). Additionally, we encourage more attention to indoor air as a transport mechanism

for resistant organisms, as evidenced by their presence on doorsills. We also found efflux to be the dominating antimicrobial resistance mechanism, especially RND efflux pumps.

Conclusions: Chlorhexidine remains an effective antiseptic at its application concentration (2%) in most cases. However, the study highlights the importance of active monitoring of its resistance development, particularly in the previously neglected surrounding environment. The study advances our understanding of chlorhexidine persistence on indoor surfaces, the resistance profile of environmental isolates in a MICU, as well as potential resistance mechanisms of environmental bacteria. Moreover, this study will contribute to improving clinical application of chlorhexidine, establishing potential antiseptic stewardship, and preventing antibiotic-resistant infections in healthcare facilities.

4.2 Introduction

The emergence of resistance to antimicrobials is a critical concern in healthcare settings, severely limiting our ability to prevent and treat antibiotic-resistant infections. Chlorhexidine digluconate (CHG) has been used as a topical antiseptic both to disinfect the skin of the patient and the hands of the healthcare providers in the medical intensive care units (MICU). It acts against a wide array of bacteria by disrupting phospholipids in the cell membrane of bacteria [165], and has been used for decades with little concern for resistance. While daily bathing with CHG for patients has served as an effective routine for infection prevention at therapeutic doses [31, 166, 167], exposure of hospital-associated bacteria to sublethal concentrations of chlorhexidine can lead to reduced susceptibility to chlorhexidine and other clinically-important antibiotics (e.g., erythromycin, clindamycin) [22]. Such conditions are likely met in the patients' surrounding

environment post application of CHG, in part due to its degradation by heat, light, and other chemicals [168-170].

Recently, decreased susceptibility to CHG has been observed in various bacterial taxa (e.g., *Acinetobacter* spp*., Pseudomonas* spp*., Enterobacter* spp., and *Enterococcus* spp.), including important pathogens (e.g., *Klebsiella pneumoniae,* methicillin-resistant *Staphylococcus aureus* [MRSA]), as has cross-resistance to colistin, raising concerns about unintended consequences of CHG application in healthcare settings [171]. Studies have obtained good correlations between chlorhexidine and antibiotic susceptibility in both minimum inhibitory concentrations (MICs) and minimum bactericidal concentrations (MBCs) among Gram-negative bacteria, and MBCs among Gram-positive bacteria [172]. Reports of hospital infection outbreaks attributable to chlorhexidine resistance are also emerging (e.g., *Serratia marcescens* outbreaks [23, 173-175]). The concern is magnified considering the inductive effect of widely used antimicrobial chemicals (e.g., triclosan, CHG) on the dissemination of antibiotic resistance genes [176, 177], and the possibility that mobile antibiotic resistance elements could transfer from environmental microbes to human pathogens. In a veterinary hospital, multi-drug resistant *Serratia* isolates with similar resistance profiles were detected from animals and chlorhexidine solutions (environmental sources) used across the hospital. Furthermore, the authors identified an IncHI2 multi-drug resistance plasmid that is shared by these *Serratia* isolates and an *Enterobacter hormaechei* strain previously recovered from the hospital environment, with the plasmid's transferring capability confirmed by conjugation experiments [23].

Triclosan, a widely used antimicrobial precedent to chlorhexidine, has largely been phased out, in part due to extensive research demonstrating its promotion of antibiotic resistance [171, 177,

178]. However, the research scrutiny on chlorhexidine is still very limited. Existing studies about chlorhexidine resistance are largely restricted to human subjects (e.g., clinical isolates from patients, especially those on the WHO priority list of antibiotic-resistant bacteria), while little attention has been paid to the surrounding environment, which can serve as a source of antibiotic-resistant bacteria and hospital-acquired infections. To date, chlorhexidine susceptibility of environmental microbes is only sporadically reported, and these reports are often associated with outbreak tracing [23, 179-181], while systematic surveillance is lacking.

We systematically studied chlorhexidine resistance in the built environment by combining controlled microcosm experiments and field surveys. Using controlled microcosm experiments, we revealed chlorhexidine persistence patterns on plastic, metal, wood surfaces under common disinfection and cleaning practices. We found that the application of peracetic acid induced immediate reduction of CHG on surfaces. After 24 h, CHG concentrations decreased significantly on metal surfaces and cleaned or disinfected plastic and wood surfaces. Applying bacterial cultures on CHG-containing surfaces, we found that clinically relevant bacteria can survive 24 h CHG exposure on surfaces in the range of concentrations at which CHG persists. We further conducted a field survey in a MICU and investigated how widespread CHG tolerance was *in situ.* Despite CHG primarily being applied to patient skin, chlorhexidine tolerance was identified in multiple spatially distinct locations in the MICU. Particularly, sinks were a critical reservoir regarding the absolute bioburden, CHG tolerance, and bacteria with exceedingly high CHG MICs (≥ 512 μg/mL). Moreover, indoor air deserves more attention in addition to high-touch surfaces as a transport mechanism for resistant organisms. Efflux was the dominant antimicrobial resistance mechanism in the environmental isolates, especially RND efflux pumps.

4.3 Methods

4.3.1 Chlorhexidine persistence on surfaces

We investigated the chlorhexidine persistence on 3 surface materials commonly found in hospital (laminate wood, plastic [high-density polyethylene; HDPE], metal [stainless steel]) in response to 6 different cleaning practices a hospital surface typically experiences (no cleaning, clean with water, disinfect with ethanol, bleach, peracetic acid, or benzalkonium chloride) (Fig. S4.1).

Experimental setup

Each surface was cut into 48 4”×8” coupons and divided into 1”×1” sections by drawing markers. The coupons were UV sterilized for 15 minutes before applying 2% chlorhexidine digluconate (CHG) solution (w/v, Sigma-Aldrich, St. Louis, MO, USA). The concentration 2% was selected to represent commercial CHG wipes used in patient bathing [182, 183]. We first folded a Kimwipe (Kimberly-Clark, Irving, TX, USA) in half 3 times to get a square about 2”×2” and 8 layers thick. We then pipetted 1 mL CHG onto the Kimwipe and wiped in a zigzag pattern along the long axis of the surface, such that the whole surface was covered after 5 passes. Wiping was repeated with a second Kimwipe from the opposite direction. Once the surface dried, an initial CHG measurement was taken in triplicate, representing 0 min. Each 1”×1” section was swabbed using a nonsterile cotton swab premoistened in 0.8 mL deionized water. The swabbing was conducted with consistent pressure and speed in 2 directions (up and down, left and right) for 15 s total. The swab was then stored in the deionized water for extraction and CHG concentration measurement.

Simulated cleaning practices were applied on CHG-containing surfaces using the same wiping technique. A group of surfaces without cleaning served as baseline controls. The wiping solutions included deionized water, 70% ethanol (Decon Laboratories, King of Prussia, PA, USA), 10% bleach (Clorox, Oakland, CA, USA), 0.2% peracetic acid (Sigma-Aldrich, St. Louis, MO, USA), and 0.26% benzalkonium chloride (MP Biomedicals, Santa Ana, CA, USA; C8-18) [184]. The CHG concentrations were measured in triplicate at 5 additional timepoints (10 min, 1 h, 3 h, 6 h, and 24 h) following the same method. The coupons were kept in light-proof compartments when not being wiped or swabbed. This was repeated for each disinfectant and each surface for a total of 8 replicates.

Swab extraction and CHG concentration measurement

CHG on swabs was extracted into water following an adaptation of CDC's Swab Extraction Method [114, 185]. The Falcon tube was vortexed at maximum speed for 1 min and shaken at 180 rpm and 25 °C for 10 min, followed by careful squeezing against the tube wall and removal of the swab. CHG concentration was determined by a previously reported colorimetric assay with modifications [31, 186, 187]. We mixed 500 µL CHG solution with 500 µL hexadecyltrimethylammonium bromide (0.01 g/mL in deionized water, Sigma-Aldrich, St. Louis, MO, USA) and 200 µL sodium hypobromite (Aqua Solutions, Deer Park, TX, USA). Absorbance of the mixture was measured at 262 nm using a spectrophotometer ().

Determination of CHG recovery rates from surfaces by swab sampling

We examined the CHG recovery rates from different surface materials using the same swabbing technique and accounted for these rates in calculating the CHG concentration on surfaces. After UV sterilization, we pipetted 100 µL 2% CHG onto a 1"×1" square for a total of 10 replicates

each surface type. Additionally, 100 µL 2% CHG was directly pipetted onto a swab to investigate the recovery rate of the swab extraction.

4.3.2 Viability assay of bacterial isolates on CHG-containing surfaces

We tested the survival of bacterial isolates on plastic surfaces after exposure to different concentrations of CHG for 24 h in triplicate. Four bacterial isolates (*Escherichia coli* [ATCC 25922], *Klebsiella pneumoniae* [ATCC 13883], *Klebsiella variicola* [20-20012], and *Staphylococcus aureus* [ATCC 29213]) were selected as model strains. *Klebsiella variicola* was an in-house strain obtained from our collaborator at Rush University. Plastic was selected as a representative surface because it is a common material of equipment in the immediate vicinity of patients and healthcare workers (e.g., bedrail, nurse call, keyboard, mouse). Five different CHG concentrations (31.00, 15.50, 3.10, 0.31, 0 [positive control] µg/cm^2) were tested to cover the range of CHG persisting on surfaces under disinfection and cleaning.

To prepare bacterial inocula, the second passage of the stock culture were grown at 37°C to the late exponential phase and harvested at 3000 rpm for 1 min. The pellet was then resuspended in phosphate-buffered saline (PBS). The bacterial cell density in the inocula was determined by spread plating. To prepare surface coupons, HDPE plastic was cut into 2”×2” squares, UV sterilized for 15 min, and placed in a petri dish. CHG was applied to the center 1”×1” of the coupon by pipetting 100 µL solution and spreading evenly with the pipette tip. Subsequently, 100 µL isolate suspension was applied using the same technique [123]. An additional coupon inoculated with 100 µL PBS was included as the negative control. Surfaces were kept in dark at room temperature.

After 24 h, bacteria were sampled from surfaces and the surviving numbers were counted. Each surface was first dry swabbed for 5 s and wet swabbed twice for 7.5 s each time after rinsing in 1 mL PBS. Finally, the swab was squeezed against the tube wall to expel any residual liquid [123]. Samples were diluted and 100 µL of each dilution was spread plated onto TSA plates. The plates were incubated at 37°C for 24 hours to count the colonies.

4.3.3 Field survey in a Medical Intensive Care Unit

Sample collection

We collected 219 swab samples from 7 different locations (e.g., bedrails, nurse calls, doorsills, keyboards, light switches, sink; Table S4.4, Table S4.5) in the MICU at Rush University Medical Center over 2 sampling events (February and July, 2018) (Fig. S4.2), including negative field controls and negative media controls. The MICU was selected because patients there are likely to have highly resistant infections and be highly susceptible to infection, which gives this environment a higher probability but also a lower threshold for acceptable risk of containing antibiotic-resistant organisms. Sampling locations were selected to capture a variety of surfaces reflecting different ways in which people interact with the surface [17] and time-integrated aggregates of what was in the air (i.e., doorsills) [188]. Sites were further selected to facilitate comparison with other hospital surveillance studies [189] and to keep a relatively balanced number of samples from each touch frequency group. Touch frequency categorization included bedrail and nurse call as high-touch, keyboard as medium-touch, and light switch as low-touch surfaces. Doorsills and sinks were labeled as themselves. Touch frequency designations were based on anecdotal evidence from experienced personnel and prior evidence from the literature [190-193].

Swabbing was conducted following CDC's Environmental Surface Sampling Swab Contact Method. Briefly, we used three sterile Nylon Flocked Dry Swabs (COPAN Diagnostics) premoistened with PBS to increase the recovery of biomass. Surfaces were swabbed in their entirety three times with consistent pressure and speed. Each time, we rotated the swabs and switched their order. Swabs were stored in 15 mL tubes with PBS and transferred to the lab on ice within 24 hours for immediate processing. Temperature, relative humidity, isolation level, and patient mobility for each patient room were recorded. The surface area of each sampling location was measured (Table S4.6).

Swab extraction

Biomass on swabs was extracted into PBS following an adaptation of CDC's Swab Extraction Method and [114]. Briefly, the Falcon tubes with swabs inside were vortexed at maximum speed for 10 s, shaken at 180 rpm and 25 °C for 10 min, and then vortexed again for 10 s, followed by careful squeezing against the side of the tube and removal of the swabs (Fig. S4.3).

Sample cultivation and initial screening

The samples were then subjected to cultivation and screening following a protocol modified from [194] (Fig. S4.4). Inocula were diluted to obtain between 30 and 300 colonies per plate. 100 µL of inoculum was spread plated on tryptic soy agar supplemented with 4 mg/L itraconazole (TSAI) and incubated at 25˚C for 4 days. When observable growth was detected, morphologies of colonies were characterized using standardized ontologies for streamlined analysis, and a representative number of colonies belonging to each morphology type were picked and stored in glycerol at -80˚C. These plates were then replica plated onto blood agar and chlorhexidine agar plates for initial screening [195]. Chlorhexidine agar plates contained 10.56 µg/mL chlorhexidine

(equivalent to 18.75 µg/mL CHG, Sigma-Aldrich, St. Louis, MO, USA). *Staphylococcus aureus* ATCC 25923, *Staphylococcus epidermidis* ATCC 12228, and *Pseudomonas putida* 56A10 (in-house ID) [196] were used as negative control strains in this chlorhexidine screening process. The chlorhexidine concentration was selected for initial tolerance screening because it was reported as the minimum effective concentration associated with significantly decreased colony counts of Gram-positive bacteria on the skin of patients who were bathed daily with CHG in the MICU. In addition, all the patient skin isolates recovered in the study had MICs below 18.75 µg/mL CHG [31].

Antimicrobial susceptibility testing

We used a simplified agar dilution method to confirm chlorhexidine tolerance observed in the initial screen [197]. All steps were strictly followed except that the inoculum was not standardized to 10^4 CFU per 5- to 8-mm diameter spot. Briefly, agar dilution plates with different concentrations of CHG (0, 4, 8, 16, 32, 64, 128, 256, 512 µg/mL) were made with Mueller-Hinton Agar (MHA; Sigma-Aldrich, St. Louis, MO, USA) and chlorhexidine digluconate solution (20% w/v, Sigma-Aldrich, St. Louis, MO, USA). Each isolate was first inoculated from frozen stock in 100 µL TSB in a sterile 96-well clear round-bottom not-treated microplate (Corning, NY, USA) and incubated at 37 °C for 24 h, yielding F1. F1 was subcultured in 100 µL TSB with a 96-pin microplate replicator (transfers 1 µL inoculum each pin, Boekel Scientific, Feasterville-Trevose, PA, USA) and incubated for another 24 h, yielding F2. F2 was inoculated onto surfaces of the agar dilution plates in triplicate with the 96-pin replicator, starting from the lowest concentration to the highest. MICs were determined after incubating the plates at 37 °C for 20 h. Negative controls and 2 quality control (QC) strains were included in every batch. The 2 QC strains were both MRSA with CHG

MICs 1-2 and 8 µg/mL, respectively. The QC strains were obtained from our collaborator at Rush University, whose MICs were validated by them and the CDC.

Because no consensus on standardized clinical breakpoints is available for chlorhexidine [198], we discuss chlorhexidine resistance based on epidemiological cut-off values proposed by previous studies (Table S4.3) [171, 199, 200]. The values were defined as the upper limit of the normal MIC distribution for a given antimicrobial and wild-type species. Clinical implications of these values are unclear, but they are considered helpful in resistance monitoring. In this study, CHG was used for MIC measurements, and chlorhexidine MICs were used as a general term for MICs reported by previous studies where the chemical was not explicitly stated.

4.3.4 Genomic exploration of isolates recovered in the MICU

Whole genome sequencing

We selected 43 putative *Pseudomonas* isolates based on morphology recovered from the first sampling event for whole genome sequencing. Out of the 43 isolates, 27 were from the sinks, 9 from the doorsills, 5 from the bedrails, 1 from the light switch, and 1 from the nurse call. Whole genome sequencing was conducted by the Broad Institute using a fluorescent dye-based method followed by paired-end short-read (2×150 bp) sequencing on an Illumina HiSeq platform as previously described [196].

Genome assembly

The Binary Alignment Map (BAM) files were first converted to Sequence Alignment Map (SAM) files using SAMtools v1.10.1 [53]. The quality profiling, read filtering, red pruning, and adaptor trimming were performed using Fastqc v 0.11.5. Genomes were assembled using

SPAdes v3.14.1 [201] and assembly quality statistics obtained using QUAST v 5.0.2 (https://github.com/ablab/quast). The taxonomic assignment was performed using two algorithms: autoMLST (https://automlst.ziemertlab.com/analyze) and KmerFinder 3.2 (https://cge.cbs.dtu.dk/services/KmerFinder/). The inconsistencies in results were resolved using GTDB-Tk 2.0.0.

Identification of antimicrobial resistance genes

The complete open reading frames (ORFs) were first predicted with Prodigal v2.6.3. The predicted protein sequences were then analyzed based on the reference data from the Comprehensive Antibiotic Resistance Database (CARD, v3.2.2) and antibiotic resistance genes extracted using the Resistance Gene Identifier (RGI, v5.2.1, https://card.mcmaster.ca/) [202, 203].

4.3.5 Statistical analysis

Statistical analyses and data visualization were conducted in R (v4.0.4) [138] with packages such as ggplot2 [139], and dplyr [204]. Differences between groups were determined by t-tests with the BH adjustment. $p \leqslant 0.05$ was defined as statistically significant. Significance codes are as follows: $p > 0.05$ (ns), $0.01 < p \leqslant 0.05$ (*), $0.001 < p \leqslant 0.01$ (**), $0.0001 < p \leqslant 0.001$ (***), and $p \leqslant 0.0001$ (****).

4.4 Results and Discussion

4.4.1 Chlorhexidine persists on surfaces under disinfection and cleaning

We investigated the chlorhexidine persistence on 3 common types of hospital surfaces (laminate wood, plastic [high-density polyethylene; HDPE], metal [stainless steel]) in response to interactions with 6 typical cleaning practices (no cleaning, clean with water, disinfect with ethanol, bleach, peracetic acid, or benzalkonium chloride) (Fig. 4.1, Fig. S4.1).

We first explored the CHG recovery rates from different surface materials in sampling and measurement, and the CHG transfer rates from the Kimwipe to the surfaces in application (Table S4.1, Table S4.2). When we directly applied CHG onto swabs, 85.46% was recovered in the measurement. For the tested surfaces, metal had a comparably high recovery rate (83.04%), indicating that almost all the CHG on a metal surface can be collected by swabbing. In contrast, plastic and wood both retained CHG in sampling, with recovery rates of 69.79% and 55.58%, respectively. Upon wiping, we found that only 0.13-0.20% CHG was transferred onto the surfaces. In comparison to the recovery rates from surfaces by swab sampling, the transfer rates showed an opposite trend across surface materials. While metal had the highest recovery rate in sampling, it captured the least CHG in application. Furthermore, the values were within the range that disinfectants released from ready-to-use Towelettes [205], validating that our application technique mimics the usage of commercial disinfectant Towelettes.

Overall, CHG persisted on surfaces, displaying subtle differences in persistence patterns across surface materials and cleaning practices. We investigated the immediate impact of wiping a surface on the CHG persistence by comparing the CHG concentrations at 0 and 10 min (i.e., 0.17

h) using one-tailed paired *t*-tests. Wiping with peracetic acid had the largest immediate impact, although only the difference on plastic surfaces was statistically significant ($p = 0.01$). This is supported by previous evidence that chlorhexidine can be inactivated at a low pH and in the presence of anionic molecules [206-208]. In contrast, none of the other disinfectants tested are known to have interactions with chlorhexidine that will result in its degradation. Benzalkonium chloride and ethanol have been shown to have synergistic or additive antimicrobial properties when applied with chlorhexidine, indicating that their interaction will not cause chlorhexidine degradation [209-213].

The longevity of chlorhexidine on a surface in response to cleaning and disinfection was examined by comparing the concentrations at 10 min and 24 h using one-tailed paired t-tests. When wiping was not conducted, the amount of CHG did not decrease on wood and plastic surfaces, whereas a significant decrease was observed on metal surfaces. Moreover, across all different cleaning practices, the decrease patterns were similar to that of no wiping on metal surfaces. However, compared with no wiping, cleaning plastic or wood surfaces with water or disinfectants induced a significant reduction of CHG concentrations after 24 h. Our findings are consistent with previous indications that chlorhexidine degradation might be catalyzed by metals and that chlorhexidine is unstable under acidic and alkaline conditions (pH 3.5 to 6.5 and pH > 8.5) [214].

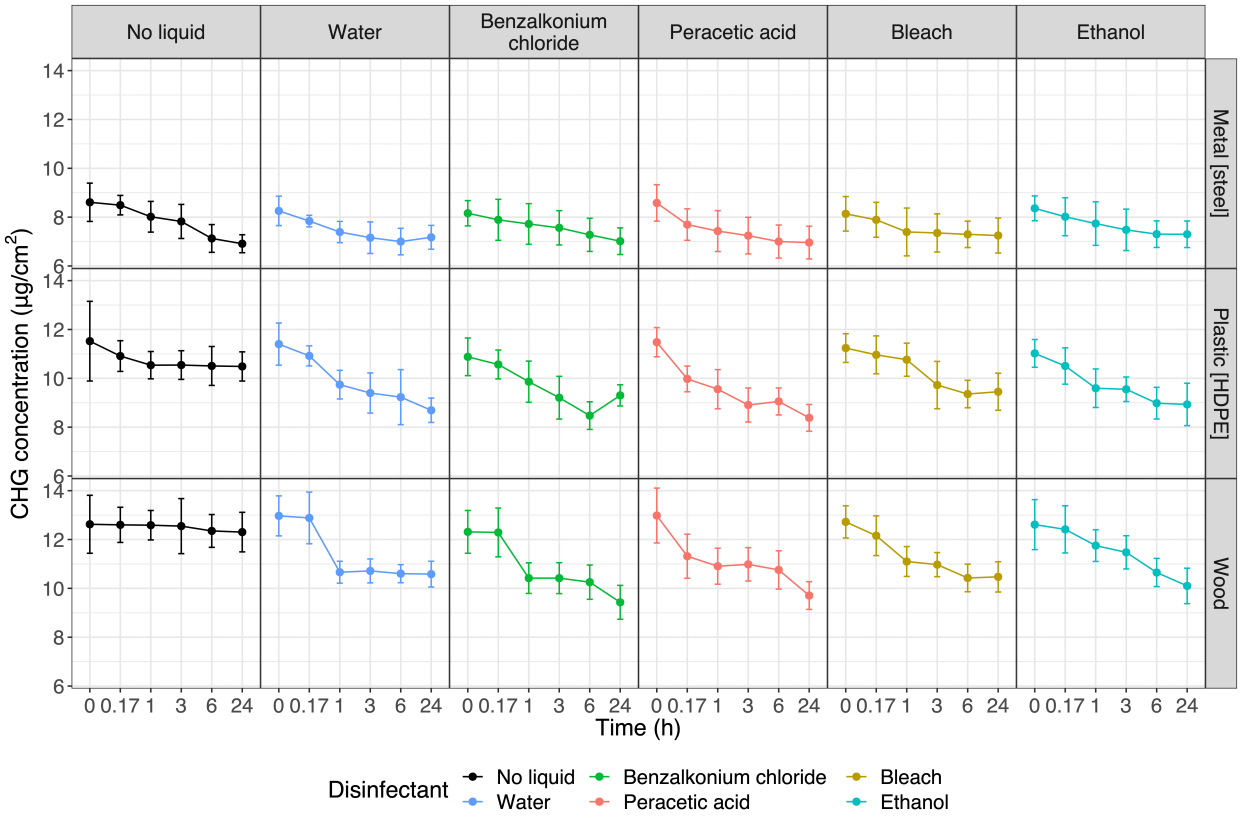


Fig. 4.1 Concentration of chlorhexidine digluconate on surfaces over time in response to cleaning and disinfection.

4.4.2 Clinically relevant bacteria can survive 24 h CHG exposure on surfaces

We examined the survival of 4 clinically relevant bacterial isolates (*Escherichia coli*, *Klebsiella pneumoniae*, *Klebsiella variicola*, *Staphylococcus aureus*) on plastic surfaces after exposure to CHG for 24 h. Plastic was selected as a representative surface because it is a common material of equipment in the immediate vicinity of patients and healthcare workers (e.g., bedrail, nurse call, keyboard, mouse). Five concentrations (31.00, 15.50, 3.10, 0.31, 0 µg/cm^2) were tested to cover the range of CHG persisting on surfaces under disinfection and cleaning.

After 24 h, bacteria survived on CHG-containing surfaces across all tested concentrations for *Klebsiella variicola*, *Escherichia coli*, and *Staphylococcus aureus* (6.10×10^2 – 1.28×10^6 CFU/cm^2,

log10 reduction ranging from 0.08 to 2.84). Although much fewer *Klebsiella pneumoniae* survived at 15.5 and 31 μg/cm^2 CHG, it is possible that >17 (around the scale of 10^2) CFU/cm^2 will survive at CHG concentrations of 8.4-11.5 μg/cm^2 (range of CHG persisting on plastic surfaces). These raise concerns about potential development of antimicrobial resistance in bacteria on surfaces that possibly harbor chlorhexidine residue (e.g., hospital environments where patients are bathed with CHG).

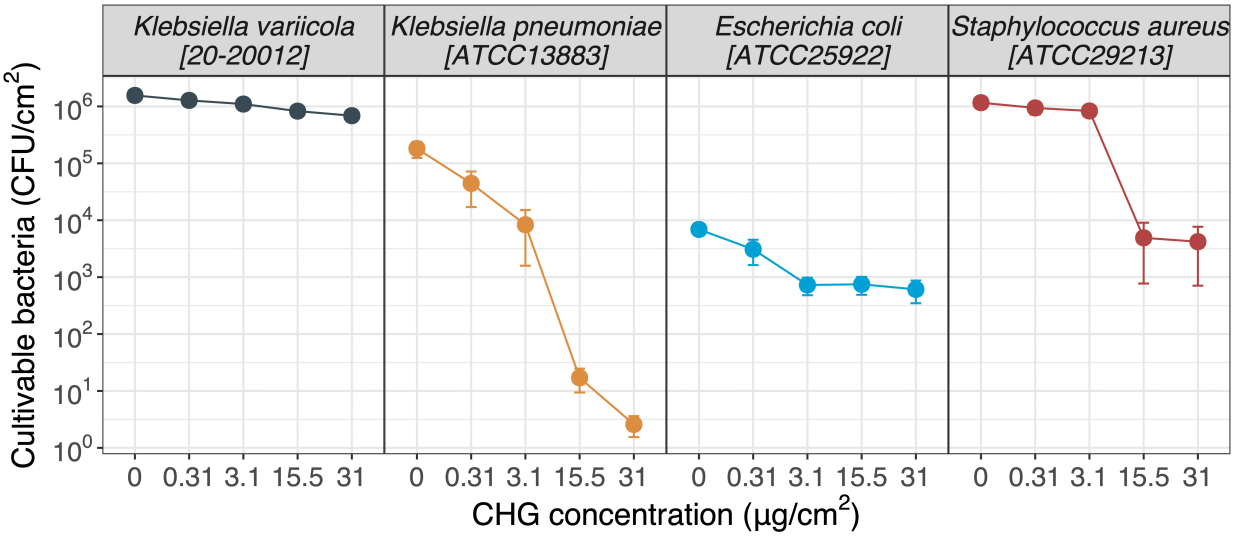


Fig. 4.2 Cultivable bacteria (CFU/cm^2) after 24 h exposure to CHG on plastic surfaces.

4.4.3 Sink is a critical reservoir for highly CHG-tolerant bacteria in MICU

Our microcosm experiments indicated that the environment may serve as a reservoir for CHG- or antibiotic-resistant bacteria. To determine how widespread CHG tolerance is *in situ*, we further conducted a field survey in a MICU. Cultivation yielded a total of 1415 isolates of 345 different phenotypic morphologies (Fig. S4.5), of which 306 isolates (21.6%) were confirmed to grow at 18.75 μg/mL CHG at 25 °C. These isolates were described as CHG tolerant in this study.

Additionally, 203 isolates (14.3%) were suspected to grow, and all the 509 isolates (36.0%) were subjected to antimicrobial susceptibility testing to determine their MICs to CHG.

Unit area surface bioburden in the MICU is significantly associated with touch and cleaning frequencies. The rarely cleaned doorsills had significantly higher bioburden than high-touch bedrails and nurse calls, followed by medium-touch keyboards (unpaired *t*-tests, log10 scale) (Fig. 4.3a). This trend remained considering the absolute bioburden in the entirety of the surfaces (Fig. S4.6). The absolute bioburden of the low-touch light switches was the lowest. However, they displayed the highest unit area bioburden among dry surfaces. This is probably due to the high uncertainty induced by switches' small area. No effect of sampling seasons and patient room isolation status (contact isolation or not) was observed (Fig. S4.7).

Despite chlorhexidine primarily being applied to patient skin, chlorhexidine tolerance was identified in multiple spatially distinct locations within the MICU. The percentages of bacteria exhibiting tolerance to 18.75 μg/mL CHG at 25 °C were not significantly different among dry surfaces (Fig. 4.3b). Most bacteria on dry surfaces were not CHG tolerant, with the mean percentages ranging from 0% (keyboards inside patient rooms, light switches in the hallway) to 13.2% (nurse calls). Doorsills were the only dry surfaces where bacteria with CHG MICs higher than 18.75 μg/mL at 37 °C were detected (Fig. 4.3c, 32 μg/mL - 1 isolate, 64 μg/mL - 4 isolates). There was also a slight trend of a subpopulation of doorsill samples harboring high proportions of CHG tolerant bacteria (3 out of 59 samples had percentages >90%). As biomass on doorsills indicates what is circulating in the air [215-217], we suggest increasing assessments on the indoor air, especially regarding its potential as a transport mechanism for resistant organisms (e.g., chlorhexidine resistance).

Sinks are a critical reservoir both regarding the absolute bioburden (Fig. S4.6) and CHG tolerance (Fig. 4.3b). The absolute bioburden and the proportion of CHG-tolerant bacteria from sink were significantly higher than all the dry surfaces to a great extent ($p \leq 0.0001$), indicating that water availability is an impactful factor. Its impact on antimicrobial adaptation is probably stronger than other factors such as touch frequency in the built environment. Moreover, 37 bacterial isolates with high MICs were detected from sink drains in patient rooms, medication rooms, and communication space (128 μg/mL - 17 isolates, 256 μg/mL - 10 isolates, 512 μg/mL - 10 isolates) (Fig. 4.3c). Of the bacterial isolates tolerant to 18.75 μg/mL CHG at 25 °C from the sinks, 70.6% (127 out of 180 isolates) were β-hemolytic (a highly predictive indication of pathogenicity [218]), including these with high MICs (24 out of 29 isolates, Fig. 4.3d). The prevalent detection of CHG-tolerant bacteria and potential pathogens in sinks is concordant with previous observations that hospital sink drains serve as a reservoir for nosocomial pathogens and antibiotic-resistant bacteria [219-221]. With previous evidence showing that microbes can be transmitted from sink to patients [98], our findings further emphasize the importance of sinks in preventing healthcare-associated infections and combatting antimicrobial resistance.

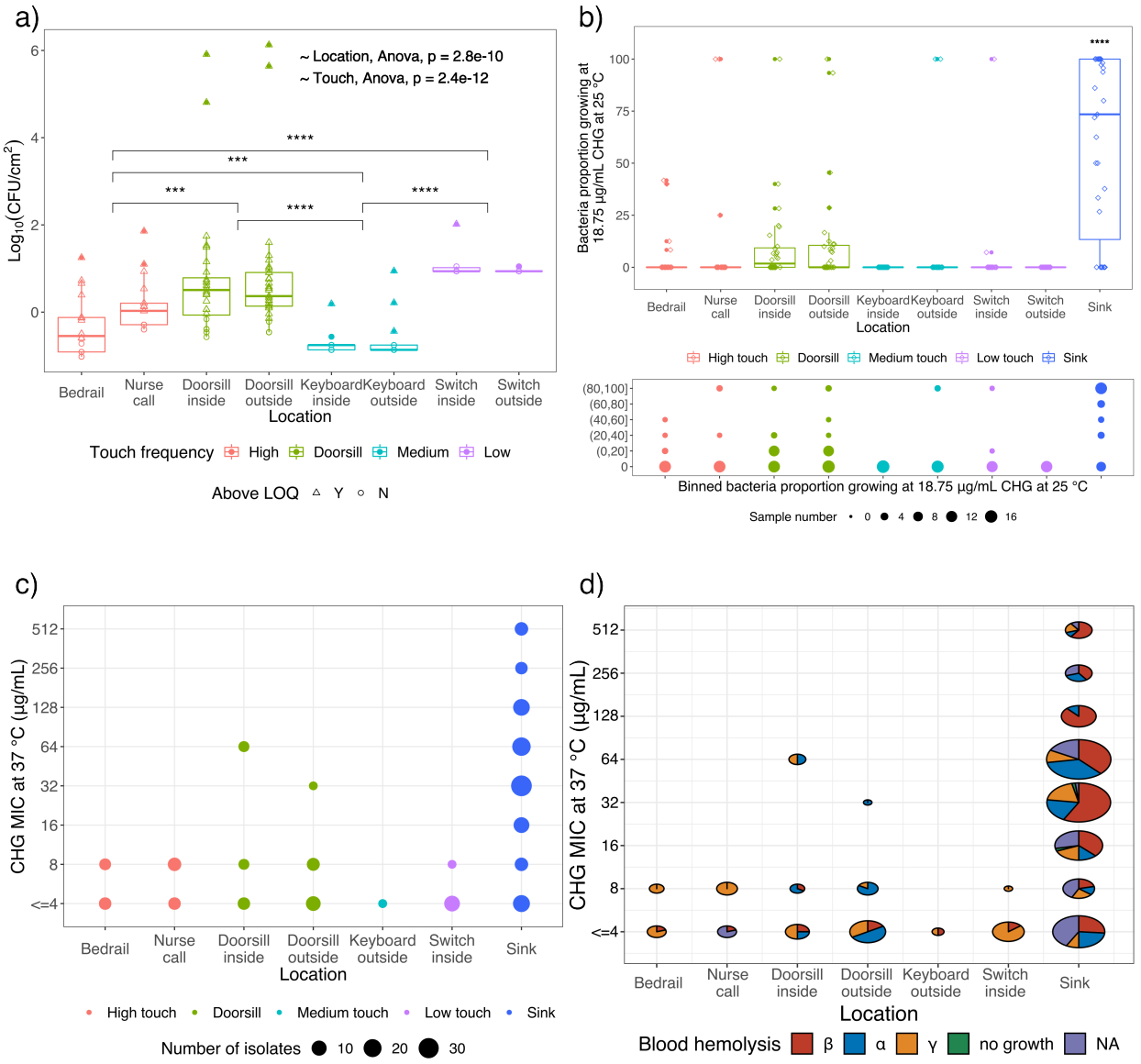


Fig. 4.3 Chlorhexidine tolerance and bioburden of bacteria in a MICU. a) Unit area bioburden of bacteria across different touch frequencies on dry surfaces. CFU stands for colony forming unit. To avoid mis-interpreting data with small colony counts, we empirically set the limit of quantification (LOQ) of our culture efforts as colony counts = 3. Values below this threshold was recoded as half of LOQ (i.e., 1.5). **b)** Proportion of bacteria tolerant to 18.75 µg/mL CHG at 25 °C across different touch frequencies on dry surfaces and sink. The proportion of sink samples was significantly higher than that of all the other touch groups according to unpaired *t*-tests ($p \leqslant 0.0001$ for all comparisons). Of the bacterial isolates growing at 18.75 µg/mL CHG at 25 °C, distribution of **c)** their CHG MICs at 37 °C and **d)** the corresponding blood hemolysis status.

As MICs are naturally expected to vary between species, a rigorous comparison of CHG MICs necessitates taxonomic classification and is thus not yet possible for all our isolates. Nevertheless, bacteria isolated in our study had CHG MICs near the high end of the historical records (especially those from sinks) [171, 172, 222-227]. This somewhat contrasts with existing observations in the literature. A meta-analysis concluded that there was no evidence of increased Minimum Bactericidal Concentration (MBC) to chlorhexidine over time for clinical and laboratory isolates of common surgical site infection-causing microbes (e.g., *Enterobacteriales*, *Staphylococci*, *Streptococci*, *Pseudomonas*, *Acinetobacter*) [222]. A three-armed cluster randomized decolonization trial found no increase of chlorhexidine resistance in the antiseptic bathing groups among 790 bacterial isolates from clinical samples of patients in ICU [198]. However, a recent review showed an increased resistance to chlorhexidine over time for *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, and *Acinetobacter baumannii* [223]. Although no population level resistance increase was detected for *Escherichia coli*, *Enterobacter faecalis*, *Staphylococcus aureus*, or *Candida albicans* [223], subpopulations with reduced susceptibility were identified for *E. faecalis* and *C. albicans*. While it is not yet clear whether our isolates represent a dramatic increase in tolerance compared to previously observed clinical isolates, our results strongly suggest that the MICU environment harbors relatively CHG-tolerant cultivable organisms.

Chlorhexidine resistance surveillance efforts were rarer in environmental sourced isolates than clinical, animal, laboratorial isolates, and the MICs of environmental isolates were generally considered lower. The highest MICs of the reported environmental isolates included 2 μg/mL for *A. baumannii* from the drain [228], 18.4 μg/mL for *E. coli* from the drain [228], 8 μg/mL for *E.*

faecalis from the dust in pig breeding facilities [180] and municipal, poultry, livestock, hospital sewage [181], 78.2 µg/mL for *P. aeruginosa* (3 isolates: 1 from the laundry siphon at a household yard, 1 from the tables and handrail in an ICU cubicle, 1 from the hand wash in an ICU hall) [179], and 16 µg/mL for *E. faecium* from municipal sewage [181]. In addition, although MIC was not measured, a *S. marcescens* strain was isolated from chlorhexidine solutions used in a veterinary hospital in 2017, indicating that extremely chlorhexidine resistant bacteria could persist in hospital environments sporadically [23]. Unfortunately, this isolate was not preserved, impeding further examination of its resistance mechanism (e.g., whether its resistance was heritable or just transient). Furthermore, chlorhexidine solutions have been linked to *S. marcescens* outbreaks in both human [174, 229] and veterinary [23, 173] hospitals. The paucity of data precludes a systematic evaluation of chlorhexidine tolerance in environmental isolates, but what data are available indicate that environmental isolates may be comparably tolerant than their clinical counterparts (e.g., as observed for *P. aeruginosa*).

Taken together, we highlight the need for more environmental surveillance of chlorhexidine resistance. Particularly, in addition to high-touch surfaces, sinks and indoor air deserve more attention in monitoring and intervention in the context of antimicrobial resistance and transmission of organisms in the hospital.

4.4.4 Efflux is the dominating antimicrobial resistance mechanism in our isolates

Of the 43 bacterial isolates that were whole genome sequenced, 19 unique species were identified, with *Cupriavidus pauculus*, *Micrococcus luteus*, and *Stenotrophomonas maltophilia* being the 3 most frequently isolated species (Fig. S4.8). Alarmingly, *S. maltophilia* is an emerging global opportunistic pathogen associated with multidrug resistance. *S. maltophilia*

exhibits resistance to a wide range of antibiotics, including β-lactam antibiotics, macrolides, and cephalosporins. Nosocomial infections associated with *S. maltophilia* are of particular concern to immunocompromised patients due to their high fatality/case ratio [230]. *C. pauculus* is reported to be associated with intravascular catheter-related bloodstream infections [231]. While *M. luteus* is often thought to be non-hazardous to humans, there have been cases of *M. luteus* bloodstream infections [232].

Among these isolates, we detected 89 unique antibiotic resistance genes (ARGs), belonging to 30 unique ARG families, 41 drug classes, and 6 antibiotic resistance mechanisms (Fig. 4.4a). An ARG family is a group of ARGs with similar ontological relationships [202]. *adeF, qacG,* and *qacJ* were the 3 major ARGs identified, with 75 hits, 34 hits, and 24 hits respectively. Antibiotic efflux was found to be the dominant resistance mechanism (associated with 86.03% ARGs), followed by antibiotic inactivation (7.82%), target replacement (3.35%), and target protection (2.79%) (Fig. 4.4a-b). Four efflux gene families were identified (i.e., resistance-nodulation-cell division [RND], small multidrug resistance [SMR], major facilitator superfamily [MFS], ATP-binding cassette [ABC] efflux pump systems), and RND efflux pump is dominating regarding the detection frequency (96 hits). Moreover, SMR and RND efflux pumps were the 2 most prevalent ARG families across different species (identified in 11 and 10 out of 14 listed species, respectively; Fig. 4.4c). Three or more efflux pump systems were detected in 4 species, indicating potential multidrug resistance.

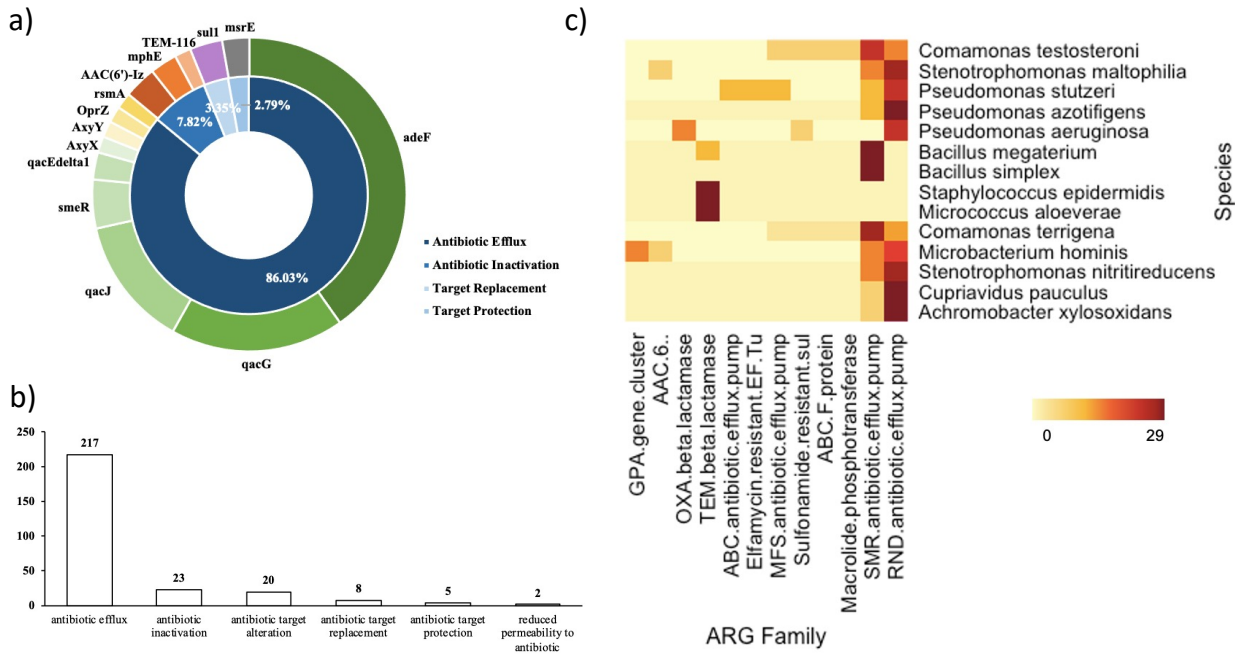


Fig. 4.4 Genetic profiles of antibiotic resistance for 43 environmental isolates. a) Major ARGs and associated resistance mechanisms identified in 43 bacterial isolates recovered from the MICU environment. Only the top 14 ARGs with number of hits greater than 2 were shown. b) Distribution of antibiotic resistance mechanisms out of 265 ARG hits. Note that one ARG can sometimes be associated with multiple resistance mechanisms. c) Distribution of ARG families across species. Only ARG families present in more than 1 species were shown.

4.4.5 Limitations and future directions

One enduring limitation in the field of chlorhexidine susceptibility studies is unstandardized reporting of MICs. For example, the type of chlorhexidine chemicals (e.g., chlorhexidine digluconate, chlorhexidine acetate) used in MIC assays was often not reported, impeding reviews and meta-analyses [222]. Consequently, existing reviews and meta-analyses usually did not normalize MICs to one chemical type. This reduces the accuracy of comparisons between our findings and previous observations.

We tested chlorhexidine persistence on surfaces using a controlled microcosm experiment. While this is informative in mechanistic exploration, further investigation of chlorhexidine

concentrations *in situ* is needed to better understand the fate of chlorhexidine in the environment (e.g., both local hospitals and downstream wastewater treatment plants).

We will sequence the isolates to further identify the taxonomy of the isolates, confirm their pathogenicity, and explore their genomic resistance mechanisms. With this information accounted, we will compare our findings to historical records. The ARGs detected are restricted to those represented in CARD, and thus novel genes will not be identified. In the future, it is worth further exploring chlorhexidine resistance mechanisms in our isolates with functional metagenomics [233]. It would also be beneficial to confirm the function of these putative resistance genes via knockouts or cloning, as well as assessing cross-resistance to medically relevant antibiotics.

In addition to current cross resistance investigations which mostly focus on a few strains, larger scale surveillance of relationships between chlorhexidine tolerance and broader antimicrobial resistance is needed. Ideally, studies should use isolates recovered *in situ* and test in a similar environment (e.g., on surfaces rather than in rich media).

While our findings indicate that sinks and indoor air may serve as reservoirs for CHG-resistant or pathogenic bacteria, more efforts are needed to lead our findings to a quantitative assessment of infection risk. Due to the lack of identification of the taxonomy for all isolates, detection of a pathogen in our study provides limited information about its load. It remains to be investigated whether the pathogens and antimicrobial-resistant bacteria we cultivated are metabolically active or dormant *in situ*. The number of bacteria that humans are actually exposed to when interacting with the built environment (e.g., using the sink) must also be investigated.

Furthermore, to evaluate chlorhexidine resistance in the built environment holistically, studies in other building types should be carried out in the future (e.g., homes, athletic facilities, veterinary facilities) to address the question: Are our findings concerning chlorhexidine-tolerant bacteria unique to the MICU where we collected samples, to hospitals where chlorhexidine is intensively used, or do they apply widely to other built environments?

Finally, critical evaluation of chlorhexidine use as part of an integrated strategy for healthcare-associated infection prevention should be undertaken, leading to antiseptic stewardship [171]. Combining lessons learned in healthcare settings with broader built environments, the antiseptic stewardship could ultimately be expanded to different aspects of human life.

4.5 Conclusion

Our study revealed that chlorhexidine decreased in concentration while persisting on indoor surfaces. Surface materials and disinfection and cleaning practices impacted the patterns of chlorhexidine persistence. *Ex situ*, we found that clinically relevant bacteria could survive 24 h CHG exposure on surfaces in the concentration range that CHG persisted. *In situ*, we found that CHG tolerance was widespread in MICU environment. Particularly, we highlight that sinks were a critical reservoir regarding the absolute bioburden, CHG tolerance, and bacteria with exceedingly high CHG MICs (≥ 512 μg/mL). Additionally, we encourage more attention to indoor air as a transport mechanism for resistant organisms, as evidenced by their presence on doorsills. We also found efflux to be the dominant antimicrobial resistance mechanism in the sequenced environmental isolates, especially RND efflux pumps.

Chlorhexidine remains an effective antiseptic at its application concentration (2%) in most cases. However, the trend of increased and widespread resistance cannot be ignored, and we should actively monitor the resistance development, particularly in the previously neglected surrounding environment.

This study advances our understanding of chlorhexidine persistence on indoor surfaces, the resistance profile of environmental isolates in a MICU, as well as potential resistance mechanisms of environmental bacteria. Moreover, this study will contribute to the improvement of clinical application of chlorhexidine and the prevention of antibiotic-resistant infections in healthcare facilities.

4.6 Supplementary information

4.6.1 Supplementary figures

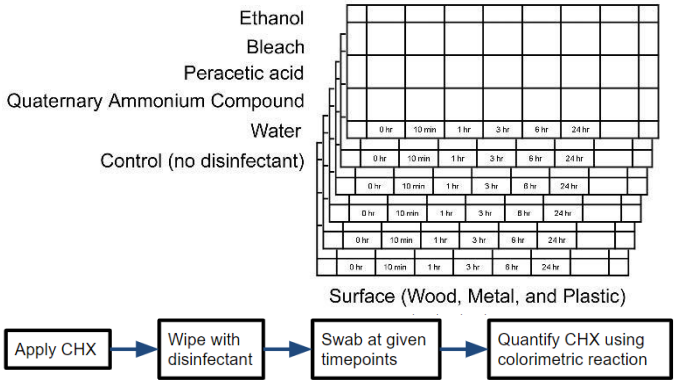


Fig. S4.1 a) Experimental workflow examining the chlorhexidine persistence on surfaces in response to different cleaning practices. b) Schematic of experiment representing one replicate.

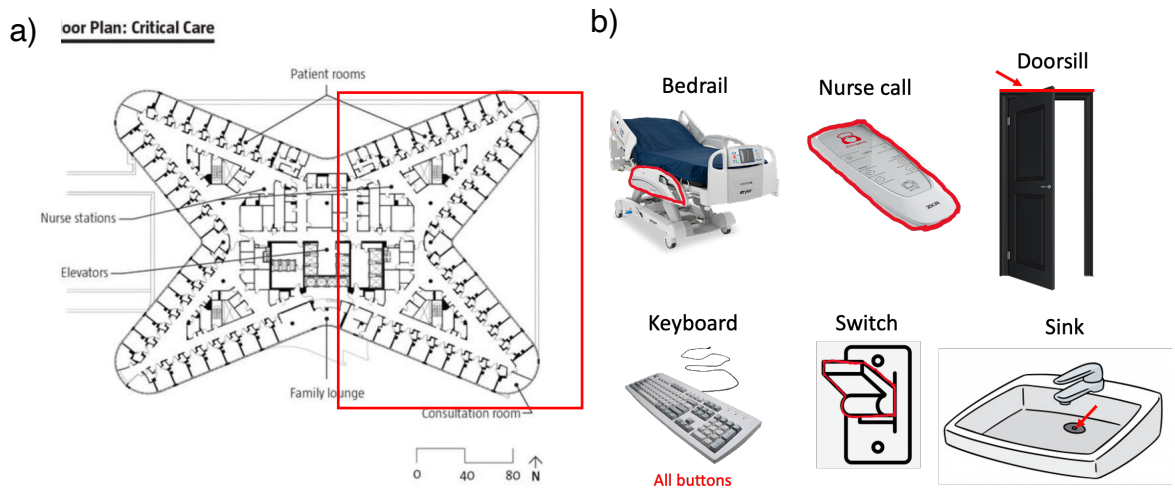


Fig. S4.2 a) Floor plan of the Medical Intensive Care Unit at Rush. b) Swabbed locations.

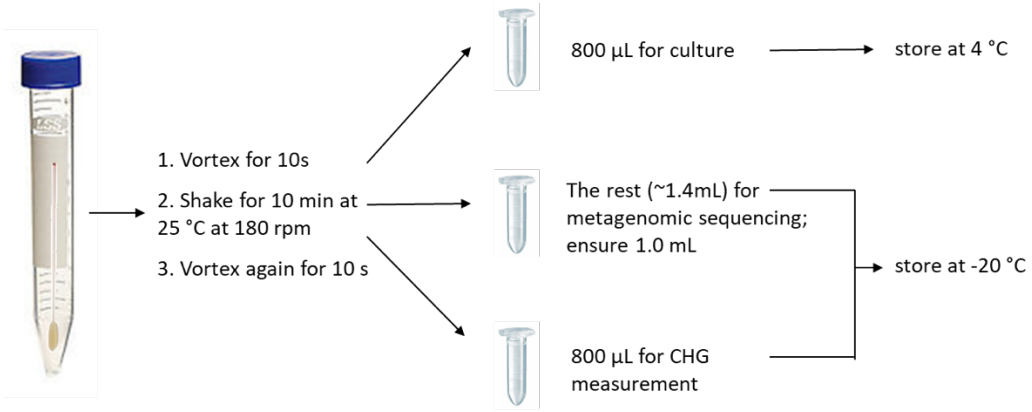


Fig. S4.3 Swab extraction process

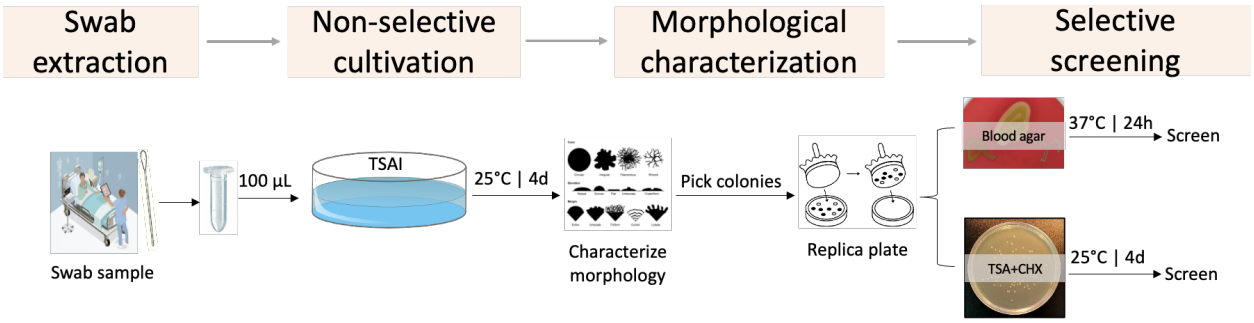


Fig. S4.4 Cultivation and screening pipeline. TSAI is tryptic soy agar (TSA) supplemented with 4 mg/L itraconazole. TSA+CHX is TSA containing 10.56 µg/mL chlorhexidine (equivalent to 18.75 µg/mL CHG).

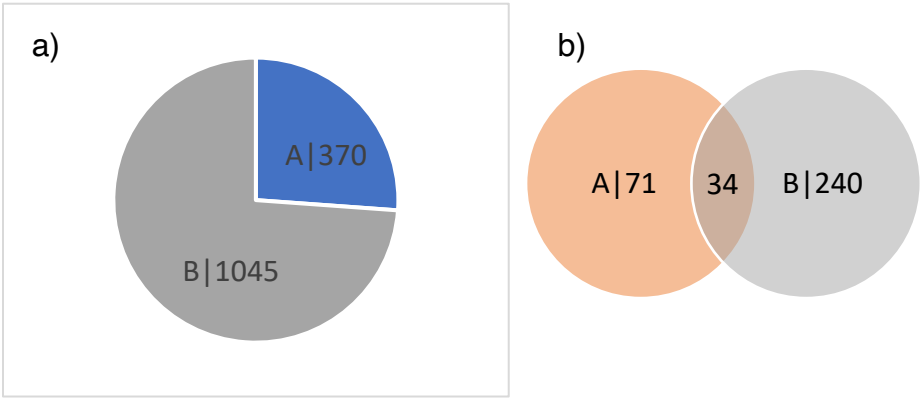


Fig. S4.5 a) Distribution of the number of isolates recovered from cultivation. b) Distribution of the number of unique morphology types. A and B are two different sampling events.

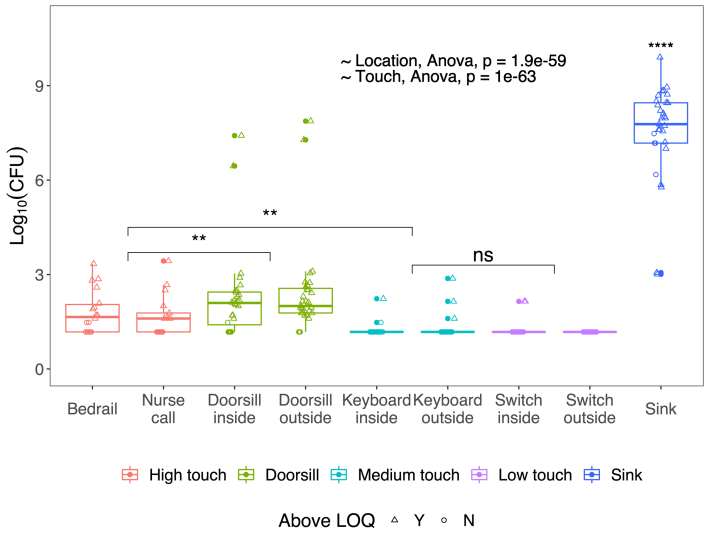


Fig. S4.6 Absolute bioburden of bacteria (log_{10}CFU) in the entirety of the surfaces and by swabbing the beginning end of the sink pipe. The bacteria CFUs were not normalized by sampling area. Differences between groups were determined by unpaired *t*-tests with the BH adjustment.

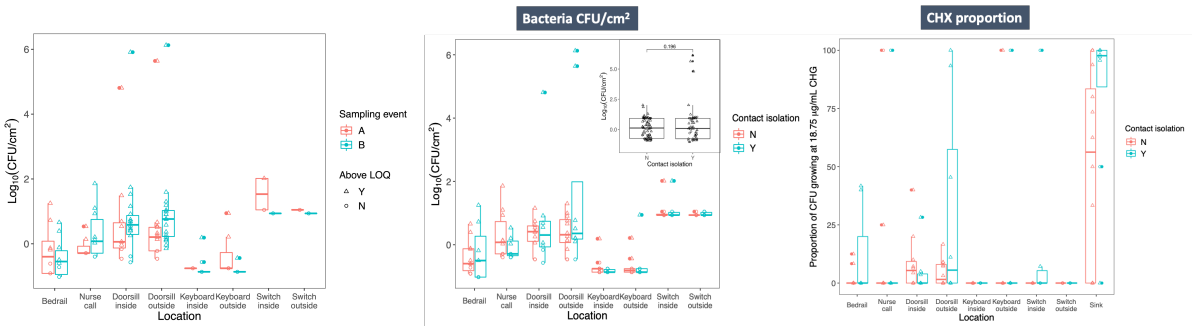


Fig. S4.7 No effect of sampling seasons and patient room isolation status (contact isolation or not) was observed. a) Unit area bioburden of bacteria across 2 sampling seasons on different surfaces. b) Unit area bioburden of bacteria across different patient room isolation levels on surfaces. c) Bacteria proportion growing at 18.75 µg/mL CHG at 25 °C across different patient room isolation levels.

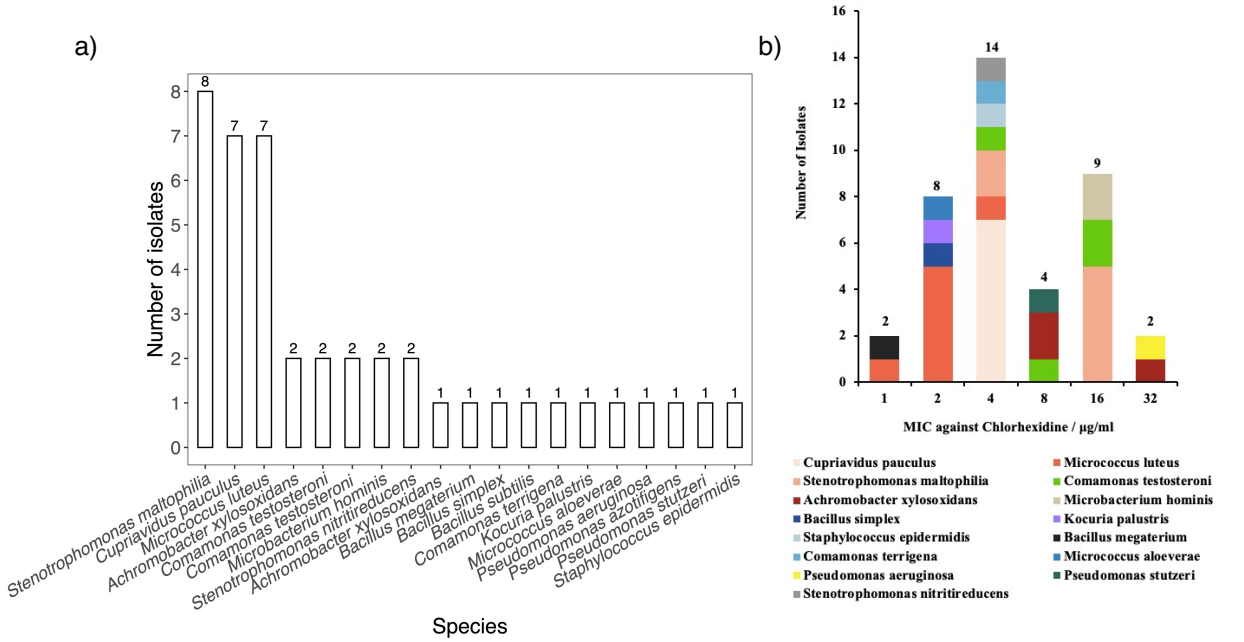


Fig. S4.8 a) Distribution of unique species identified. b) Distribution of CHG MICs across different species. CHG MICs of 39 environmental isolates were measured by broth microdilution method, with those of 4 isolates not tested because the morphological screening did not pass. Isolates of the same species may show different MIC values. Each MIC group may contain multiple unique species.

4.6.2 Supplementary tables

Table S4.1 Recovery rate of CHG on different surface materials. The recovery rate measured for surfaces was the integrative rate, representing the overall impacts of swabbing on surfaces, extracting CHG from the swab, and all the loss in the process.

Surface	CHG applied (µg)	CHG recovered (µg)	CHG recovered (%)	Standard Deviation (%)
Wood	2000	1111.56	55.58	12.86
Metal	2000	1660.87	83.04	9.13
Plastic	2000	1395.81	69.79	4.42
Swab	2000	1709.28	85.46	12.50

Table S4.2 Transfer rate of CHG from Kimwipe to surfaces by wiping.

Surface	CHX transferred (µg)	Standard Deviation (µg)	CHX transferred (%)	Standard Deviation (%)
Wood	81.96	27.52	0.2049	0.0688
Metal	53.87	15.39	0.1347	0.0385
Plastic	72.64	20.23	0.1816	0.0506

Table S4.3 Epidemiological cut-off value for CHG resistance.

Bacterial genus/species	Epidemiological cut-off value for CHG resistance (MIC; µg/mL)	Citation
E. coli	64	
K. pneumoniae	64	
Enterobacter spp.	16	
Salmonella spp.	32	
P. aeruginosa	50	
S. aureus	8	[171]
Staphylococci	4 (resistance or reduced susceptibility)	[199]
E. faecium	32	
E. faecalis	64	[171]

Table S4.4. Sampling locations within the MICU at Rush University Medical Center

Main occupant	Main Space Type	Associated Space type (Space Type Used for Mark)	Sampling location
Patient	Patient room	Patient room	Doorsill inside
			Keyboard inside
			Light switch inside
			Nurse Call
			Bedrail (left-hand side)
			Sink drain inside
	Communicating space		Doorsill outside
			Keyboard outside
			Light switch outside
Healthcare worker	Medication room	Medication room	Doorsill inside
			Sink drain
	Communicating space		Doorsill outside
	Staff-only restroom	Staff-only restroom	Doorsill inside
			Sink drain
	Communicating space		Doorsill outside
All	Communicating space	Communicating space	Doorsill inside
			Doorsill outside
			Sink drain
	Communicating space	Public restroom	doorsill inside
	Public restroom		doorsill outside
			sink drain

Table S4.5. List of Abbreviations

Sampling location	Abbreviation	Space type	Abbreviation
Doorsill inside	DI	Patient room	PT
Keyboard inside	KI	Medication room	MD
Light switch inside	SI	Staff-only restroom	ST
Nurse Call	CALL	Communicating space	CM
Bedrail (left-hand side)	BR	Public restroom	PB
Sink drain	SINK		
Doorsill outside	DO		
Keyboard outside	KO		
Light switch outside	SO		

Table S4.6. Surface area of sampling locations

Location	Area (cm^2)	Location	Area (cm^2)	Specification	Location	Area (cm^2)
Patient Room		**Medication room**		**Public restroom**		
BR	552.18	DI	128.52	PB_1	DI	111.59
CALL	130.98	DO	128.52		DO	111.59
DI	195.58	**Staff-only restroom**		PB_2	DI	10.19
DO	195.58	DI	111.59		DO	112.11
KI	384.59	DO	111.59			
KO	384.59	**Communicating space**				
SI	6.06	DI	19.02			
SO	6.06	DO	190.35			

Chapter 5 RefDeduR: A text-normalization and decision-tree aided R package enabling accurate and high-throughput reference deduplication for large datasets

5.1 Abstract

As the scientific literature grows exponentially and research becomes increasingly interdisciplinary, accurate and high-throughput reference deduplication is vital in evidence synthesis studies (e.g., systematic reviews, meta-analyses) to ensure the completeness of datasets while reducing the manual screening burden. Existing tools fail to fulfill these emerging needs, as they are often labor-intensive, insufficient in accuracy, and limited to clinical fields. Here, we present RefDeduR, a text-normalization and decision-tree aided R package that enables accurate and high-throughput reference deduplication. We modularize the pipeline into text normalization, three-step exact matching, and two-step fuzzy matching processes. We also introduce a decision-tree algorithm, consider preprints when they co-exist with a peer-reviewed version, and provide actionable recommendations. Therefore, the tool is customizable, accurate, high-throughput, and practical. RefDeduR provides an effective solution to perform reference deduplication and represents a valuable advance in expanding the open-source toolkit to support evidence synthesis research.

5.2 Introduction

As research becomes increasingly interdisciplinary, searching multiple platforms (e.g., PubMed, Web of Science) is vital to ensure the completeness of retrieved datasets in evidence synthesis

studies, especially for systematic reviews and meta-analyses [234]. This makes reference deduplication a key step between search and screening [235]. Deduplication is usually accessible as a module in multiple tools along the evidence synthesis pipeline, including search platforms (e.g., Ovid), reference management software (e.g., EndNote, Zotero, synthesisr), and screening assistance tools (e.g., Covidence, Rayyan, Metta [235], SRA-DM [236], revtools [237]). Despite wide availability, existing modules are often labor-intensive, insufficient in accuracy, and limited to certain fields or databases (primarily clinical), probably because deduplication is only one element among their multiple functions. While still useful for small datasets, due to these limitations, many approaches quickly become impractical in the fast-growing era of big data. Already, systematic reviews take tens of weeks. It has been shown that a systematic review takes an average of 164 full-time equivalent days in environmental science [27] and 67.3 weeks for a five-person team in medical fields [28]. Any processes that can accelerate computation and decrease manual labor are thus valuable improvements.

To address these challenges, we developed an R package, RefDeduR, specializing in reference deduplication. The pipeline is modularized into text normalization, three-step exact matching, and two-step fuzzy matching, making it highly customizable. With finely tuned text cleaning and normalization, RefDeduR's high-confidence exact matching process outperforms many tools, even with their fuzzy matching procedure included. We semi-automate the time-consuming and error-prone manual review process in fuzzy matching by introducing a decision-tree algorithm, making it high-throughput while still maintaining accuracy. Additionally, we propose to use the inflection point of the similarity distribution curve as the cutoff threshold, making the pipeline more practical. The tool also takes into account preprints and conference proceedings, discarding

them when a peer-reviewed version is present. This is and will be increasingly important with the rise of preprint servers. Last but not least, as a free open-source package, RefDeduR is highly interoperable with other software (including both commercial and open-source), will support the development of future tools (just like the various packages RefDeduR is built upon [237-241]), and more broadly will contribute to the prosperity of the evidence synthesis field.

Users can access RefDeduR on GitHub (https://github.com/jxshen311/RefDeduR) and view further documentation and examples on the website (https://jxshen311.github.io/RefDeduR/). Below, we demonstrate the functionality of RefDeduR with an example pipeline (see Fig. 5.1 for a flowchart of the recommended pipeline with key functions listed).

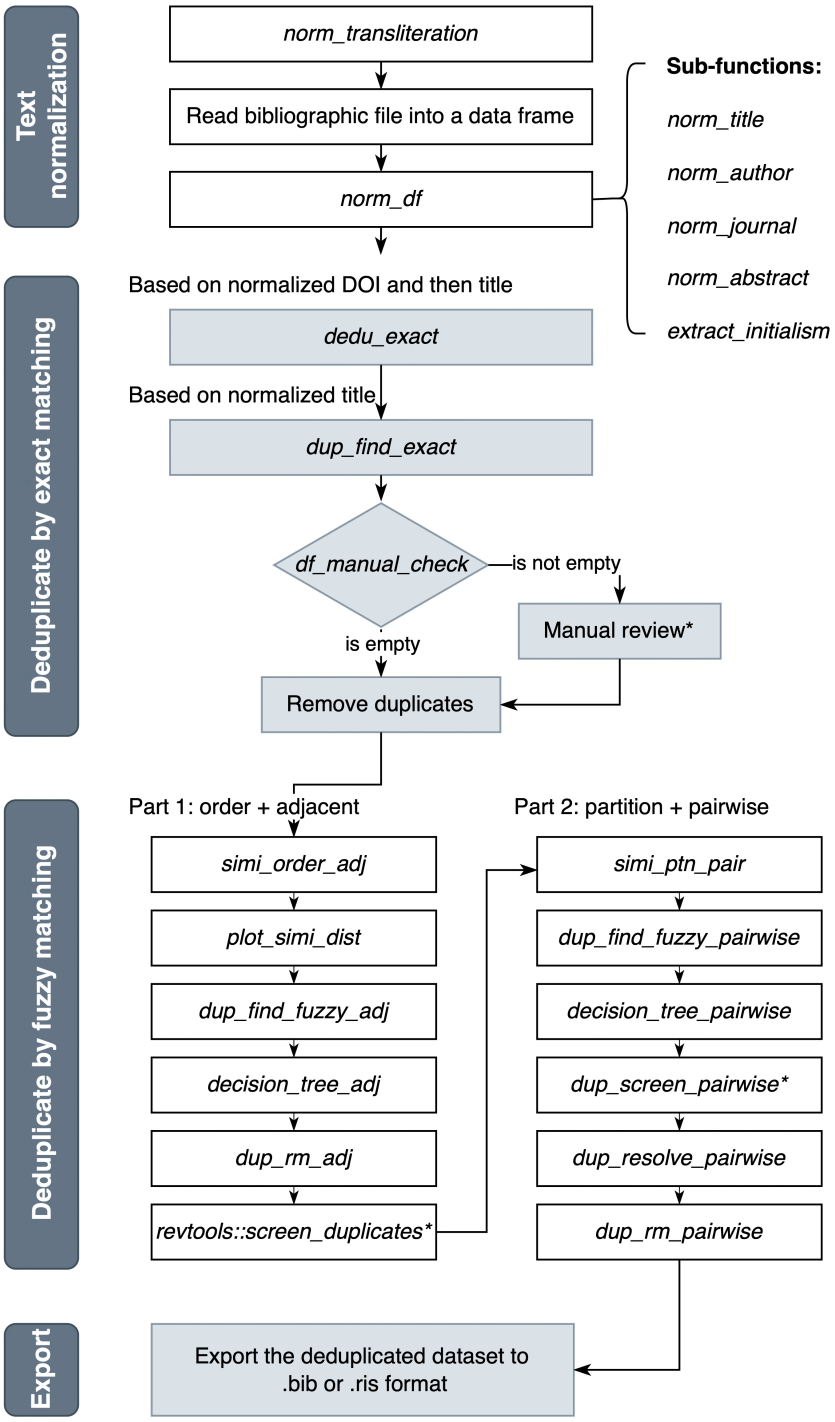


Fig. 5.1 A flowchart of RefDeduR's recommended pipeline with key functions listed. Functions are italicized to differentiate from the non-italicized descriptions. Processes involving manual review are marked with an asterisk. Refer to R documentation of each function and the tutorial (https://jxshen311.github.io/RefDeduR/articles/RefDeduR_tutorial.html) for more details.

5.3 Functionality with an example pipeline

5.3.1 Example dataset

The example dataset contains all bibliographic records (n = 6384) retrieved in a systematic review on indoor surface microbiome studies. We conducted the systematic search on 2022-01-10 through 3 platforms, and 1268, 3386, and 1733 records were retrieved from PubMed, Scopus, and Web of Science, respectively. The search terms consist of 4 key concepts: 1) indoor, 2) bacterial community/microbiome, 3) sequence-based, and 4) surface. Table 5.1 summarizes the number of records with missing values for relevant fields.

Table 5.1 Number of records with missing values for relevant fields

Field	title	author	journal	abstract	year	DOI
Number of records with missing values	0	3	50	8	20	221

5.3.2 Module 1: Text normalization

Transliterate non-ASCII characters

Before reading bibliographic files (usually BibTeX or RIS format) into data frames (the standard format for R datasets), we transliterate non-ASCII characters to the ASCII format using the function `norm_transliteration`. The default transliteration process includes: (1) transliterate common Greek letters to their names (e.g., “α” to “alpha”, “β” to “beta”) and (2) transliterate accented characters to ASCII equivalents (e.g., “á” to “a”, “ä” to “a”). Users can also perform transliteration based on customized rules or any other rules supported by `stringi::stri_trans_general`.

The transliteration increases the chance of successful deduplication by exact matching (**Module 2**). Additionally, it reduces noise when partitioning the dataset by the first two letters of `first_author_last_name_norm` at the fuzzy matching step (**Module 3**). For example, a record titled "Carriage and population genetics of extended spectrum **β**-lactamase-producing *Escherichia coli* in cats and dogs in New Zealand" sometimes has the title "Carriage and population genetics of extended spectrum **beta**-lactamase-producing *Escherichia coli* in cats and dogs in New Zealand". Author names "Álvarez-Fraga, L. and Pérez, A." are sometimes written as "Alvarez-Fraga, L. and Perez, A.".

Read bibliographic file into a data frame

We leverage function `read_bibliography` from the R package revtools [237] to read the transliterated bibliographic file (we recommend using BibTeX; see Github tutorial for reasons) into a data frame.

Text cleaning and normalization

Before deduplication, we perform multiple finely tuned text cleaning steps on the dataset. Text cleaning includes not only standard text normalization such as converting letters to lowercase, but also tailored operations in response to patterns we observed, such as removing trademark "(TM)" in title, removing English stop words in journal, and removing publisher/citation information in abstract (`norm_` functions, e.g., `norm_title`). Furthermore, we extract helper columns which we will use downstream (e.g., `journal_initialism` and `first_author_last_name_norm`). While the functions responsible for normalizing individual columns facilitate flexibility, a wrapper function, `norm_df`, is also provided to streamline the process. This function takes a data frame as the input and performs all

normalization operations required in the downstream reference deduplication. Similarly, finer text normalization increases the probability of successful deduplication at the exact matching stage, where both accuracy and confidence are assured.

5.3.3 Module 2: Deduplicate by exact matching

Following text normalization, we deduplicate by exact matching based on (1) normalized DOI (coded as `doi_norm`), (2) title, and (3) normalized title (`title_norm`) in order. DOI is chosen as the first metric because it is decisive (i.e., single selectivity). However, deduplication efficacy is limited by its medium/low applicability due to missing values. Hence, title is then employed, as it is high in both selectivity and applicability. Selectivity of a field equals 1 – (1/the number of unique field values). A field with high selectivity is one whose values are shared by only a limited number of records. A decisive field is a special case, where any non-null value is unique [235]. Applicability equals the number of non-null values divided by the number of total records, meaning that a high applicability field has few null values.

We use the function `dedu_exact` to automatically identify and remove duplicates. The user specifies one or multiple fields of the dataset according to which deduplication is conducted. If multiple fields are specified, deduplication will be performed one-by-one in order. Only records with non-null values will be investigated and the most recent version will be retained at removal. We recommend using `dedu_exact` for high-confidence fields such as normalized DOI and title.

Since completeness is crucial in evidence syntheses, we prioritize specificity over sensitivity for RefDeduR, and thus introduce a verification mechanism for fields that the user might consider

less confident. Exemplified in the recommended pipeline, we use `dup_find_exact` to locate potential duplicates according to normalized title and check the outcome based on `first_author_last_name`. If the detected duplicate sets have different values in the verification field (in this case, `first_author_last_name`), they will be output for manual review. Typically, the number of duplicate sets requiring manual review is small at this step (e.g., in the example dataset, only 1 set needs to be reviewed). Note that incorporating the verification mechanism for normalized title is particularly conservative. If verification is not needed, the user can incorporate normalized title into `dedu_exact`. Moreover, although not included in the standard pipeline, the user can utilize the functions to further search for duplicates by exact matching other fields, such as normalized abstract.

5.3.4 Module 3: Deduplicate by fuzzy matching

Once we remove all duplicates by the high-confidence exact matching processes, we proceed to fuzzy matching. Fuzzy matching is performed by calculating string similarity based on Levenshtein edit distance.

Two major practical challenges of making the fuzzy-matching process both accurate and high-throughput are (1) to choose a sensible cutoff threshold for the similarity score and (2) to reduce burden of manual review and accelerate the step. In RefDeduR, we propose two strategies to address these challenges. For challenge 1, we examine the similarity distribution plots and use the inflection point of the curve as the cutoff threshold. This value serves as a starting point to further finely-tune the threshold (discussed in more detail below). For challenge 2, we introduce a decision tree that incorporates multiple fields (e.g., title, author, year, journal, first author) to semi-automate the “manual review” step (illustrated in Fig. S5.1). This is especially helpful for

large datasets, in which case the number of duplicate sets requiring manual review could be unfeasibly high (e.g., revtools outputs ~1,400 duplicate sets for manual confirmation when treating this example dataset). Moreover, in addition to ensuring that the most recent record is retained at duplicate removal, we recognize the increasing prevalence of preprints and conference proceedings and remove them as well when co-existing with a peer-reviewed version.

To improve the computational efficiency, we divide this process into 2 parts: (1) order the records and compare only between the adjacent rows, and (2) perform pairwise comparisons between records within the same group after partitioning.

Part 1: order + adjacent

First, we calculate string similarity between adjacent rows for columns `title_norm` and `abstract_norm` using the function `simi_order_adj`. By default, the dataset is ordered alphabetically by `title_norm` before calculation, but the user may choose another field. Second, we plot similarity distributions of `title_norm` and `abstract_norm` by `plot_simi_dist` (Fig. 5.2) to choose cutoffs. The plots suggest a cutoff score of 0.7 or 0.6 for the title and 0.3 for the abstract. For demonstration purpose, we use 0.7 and 0.3 here. The selected cutoffs are then passed to `dup_find_fuzzy_adj` to locate potential duplicates. The function outputs 2 data frames: (1) the input data frame with `match` column added and (2) a data frame listing `id` of duplicate pairs (`id_dup_pair_adj`). The decision tree is introduced next to semi-automate the "manual review" process. The function `decision_tree_adj` generates and adds decisions to `id_dup_pair_adj`. There are 3 possible levels of decisions: duplicate, not duplicate, and check. If the decision is not duplicate, the `match` column will be

modified. To ensure a high accuracy, especially a low false positive rate, check is kept in the decision tree to signal manual confirmation. Finally, we deduplicate accordingly for different scenarios. For the duplicate, we remove duplicates directly by `dup_rm_adj`. For the check, we leverage `revtools::screen_duplicate`, leading to a graphical interface to interactively screen the duplicate pairs [237].

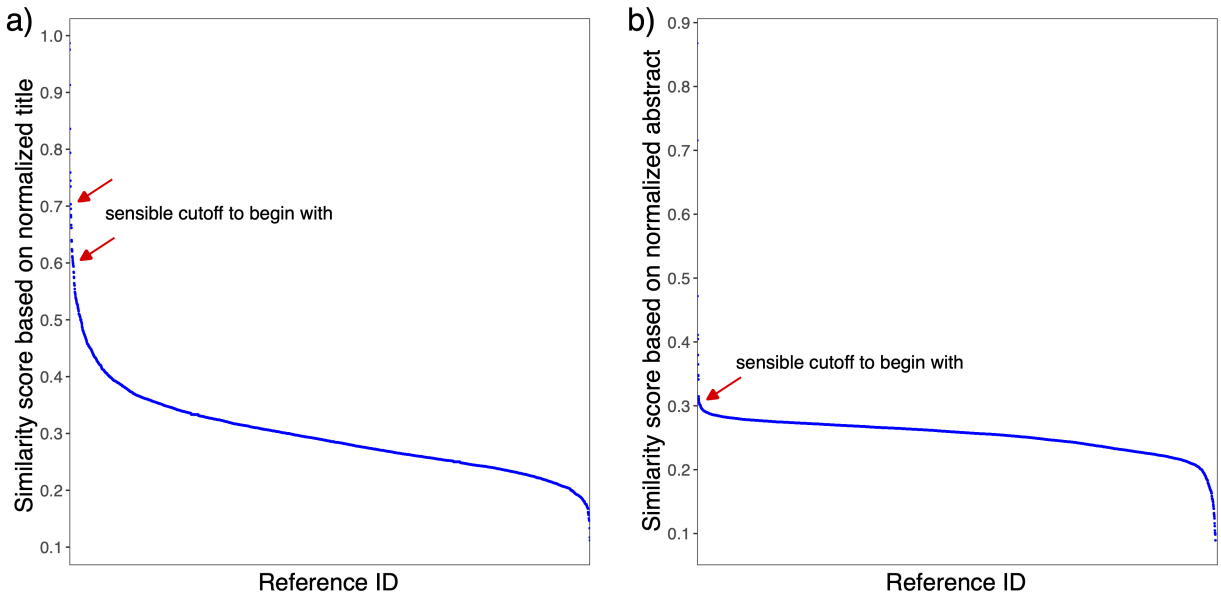


Fig. 5.2 Distribution of string similarity scores of the example dataset based on a) normalized title and b) normalized abstract. Sensible cutoffs to begin with are marked by red arrows.

Part 2: partition + pairwise

We further deduplicate according to pairwise string similarity between all records within the same partitioned group. Following procedures similar to those in part 1, we first calculate string similarity for columns `title_norm` and `abstract_norm` using the function `simi_ptn_pair`. The difference is that we partition the dataset rather than order it. We recommend using the first two letters of `first_author_last_name_norm` as the

partitioning metric, but the user has the flexibility to choose another parameter (e.g., year). We found the default metric more efficient than year for datasets that are skewed towards recent years. This is probably the case for many evidence synthesis studies, as the literature tends to grow exponentially. In addition, with the prevalence of preprints, partitioning by year becomes less accurate. Because the dataset is partitioned, results are now stored in lists as opposed to data frames in part 1. Likewise, we then locate potential duplicates by `dup_find_fuzzy_pairwise`. The cutoff thresholds can be inherited from part 1. To avoid over-deleting unique records, we suggest tightening the cutoff of abstract similarity to 0.7 (or 0.6) in this step, as opposed to 0.3 in part 1, where the risk is mitigated by the more restricted ordering. Decision tree is applied subsequently by `decision_tree_pairwise` and we can call `dup_screen_pairwise` to output the duplicate pairs that are labeled as check for manual review. Finally, we call `dup_resolve_pairwise` to resolve the check decisions to either duplicate or not duplicate accordingly and remove duplicates by `dup_rm_pairwise`.

We recommend using the inflection point as a data-driven threshold selection method based on the assumption that duplicates constitute a relatively small portion of the total after text normalization and exact matching. Thresholds from previous deduplication methods are largely empirical or anecdotal. For example, litsearchr uses “titles that are more than 95% similar, or abstracts that are more than 85% similar” but does not provide a theoretical basis for these cutoffs [242]. Compared with thresholds inherited from previous experience, this method is more quantitative. However, the selection of cutoff thresholds is based on visual observation of the distribution curves and does not actually need to be very precise. For example, the end result is not impacted by using 0.7 or 0.6 for the first threshold (Table S5.1). Instead, numbers around

that area all work similarly, partly due to the buffering functionality of the decision tree. Generally, a larger similarity score leads to higher specificity and may cause false negatives, while a smaller similarity score leads to higher sensitivity and may cause false positives. Regardless, all the 8 scenarios we tested had accuracy ≥ 99.94% (Table S5.1; title similarity: 0.7 ~ 0.5; abstract similarity: 0.7 ~ 0.3). Specifically, the recommended schemes (S1 & S2) resulted in 0 false positive and 1 false negative. If the cutoff of abstract similarity was not raised to 0.7 (or 0.6) in part 2 (i.e., still used 0.3, S3 & S4), all duplicates would be identified at the cost of having 4 false positives. Towards the conservative direction, if 0.7 (or 0.6) was used throughout the procedures for both title and abstract similarity (S5 & S6), a false positive rate of 0 was maintained, and the number of false negatives was slightly increased to 3 for 0.7 and 2 for 0.6. Lowering the thresholds to 0.55 or 0.5 caused both false positives and false negatives (S7 & S8), but none exceeded 2.

5.3.5 Export the deduplicated dataset

We can leverage `write_bibliography` from R package revtools [237] to export the deduplicated data frame into a BibTeX or RIS file. Alternatively, R packages synthesisr (https://CRAN.R-project.org/package=synthesisr) and `RefManageR` (https://CRAN.R-project.org/package=RefManageR) also contain similar functions.

5.4 Discussion

5.4.1 Benchmarking

We benchmarked RefDeduR against existing tools using the example dataset. After manual curation, 3828 records were retained in the unique subset among the raw dataset with 6384

records (Table S5.2). With this manually curated dataset as the benchmark set, deduplication performance was then assessed between March and September 2022. We considered deduplication modules from a variety of tools along the evidence synthesis pipeline, including search platforms (Ovid), reference management software (EndNote X20 v20.2, Zotero v6.0.9, Mendeley desktop v1.19.8 and synthesisr v0.3.0), and screening assistance tools (Covidence, Rayyan, Metta [235], SRA-DM [236], and revtools v0.4.1 [237]). Ovid, Metta, and SRA-DM were excluded after a preliminary examination because their functionality was restricted to clinical databases. Deduplication of the dataset was then performed for the other software. We used default settings for Endnote X20, Covidence, Zotero, Mendeley and Rayyan, and chose the highest-performance scenario for revtools and synthesisr since they provided multiple options in their documentation. Version information is not available for Covidence and Rayyan but the operation date as well as other details are described at https://github.com/jxshen311/RefDeduR_benchmark.

Quantitative evaluation was not obtained using Mendeley or Rayyan due to the lack of an option to automatically resolve detected duplicates. Mendeley desktop found 1457 sets of duplicates. Rayyan detected 3369 potential duplicates out of the raw dataset. Both are unfeasibly labor intensive with manual resolution as the only option. Moreover, according to the official announcement, Mendeley desktop has been discontinued and will be gradually replaced by the web-centric version, Mendeley Reference Manager, which does not currently support reference deduplication. Mendeley desktop also had a worse performance than Covidence in a previous comparison study based on a smaller dataset (n = 3130), which can be used as an indirect indicator [243]. While Rayyan claims that it automatically detects and resolves 100% duplicate

articles, none of the flagged duplicates were automatically deleted in the example dataset. Despite failure to make a quantitative comparison, we uploaded the dataset that had gone through RefDeduR's deduplication for Rayyan to further perform deduplication. It found 8 duplicates, all of which were deemed false positives by manual review. This indicates that Rayyan has no higher sensitivity and lower specificity than RefDeduR.

The remaining tools were evaluated quantitatively. We first identified false positives and false negatives, and then calculated accuracy, sensitivity, and specificity accordingly as described previously (Table S5.3) [243]. All these tools except Zotero support automatic resolution, for which user-developed scripts are available as interim workarounds. For example, a developer with the user name "marcelparciak" posted a Java script that automates the clicking of "Merge X items" button for 100 times with a second waiting time in between on the Zotero forum (https://forums.zotero.org/discussion/40457/merge-all-duplicates). Although in practice, the script stopped frequently (after ~ 10 clicks) and required a manual restart, we resolved all duplicates after approximately a day.

RefDeduR deduplicated the dataset to 3829 records, with only one record missed in comparison to the benchmark set, while all the other tools missed substantially more duplicate records for exclusion (i.e., false negatives) (Fig. 5.3). Notably, with only the exact matching module, RefDeduR has already outperformed all the benchmarked tools, possibly due to the optimized text normalization. All the tools had 100% specificity for the example dataset except revtools (99.97%), which mis-identified one record as duplicate (i.e., false positive). This is surprising for Endnote and Zotero since their specificity was previously reported to be only 89% and 99%, respectively [243]. This means 208 and 20 false positives out of a dataset with 3130 references.

It is possible that the software has enhanced their deduplication capability since this previous study that used Endnote X9 and was conducted between December 2018 and January 2020. Alternatively, variation between the two benchmark sets may also contribute to the different outcomes. In contrast to searching 3 platforms (PubMed, Scopus, and Web of Science) in our study, the benchmark set in the previous study was retrieved from Ovid only and was mostly clinically focused.

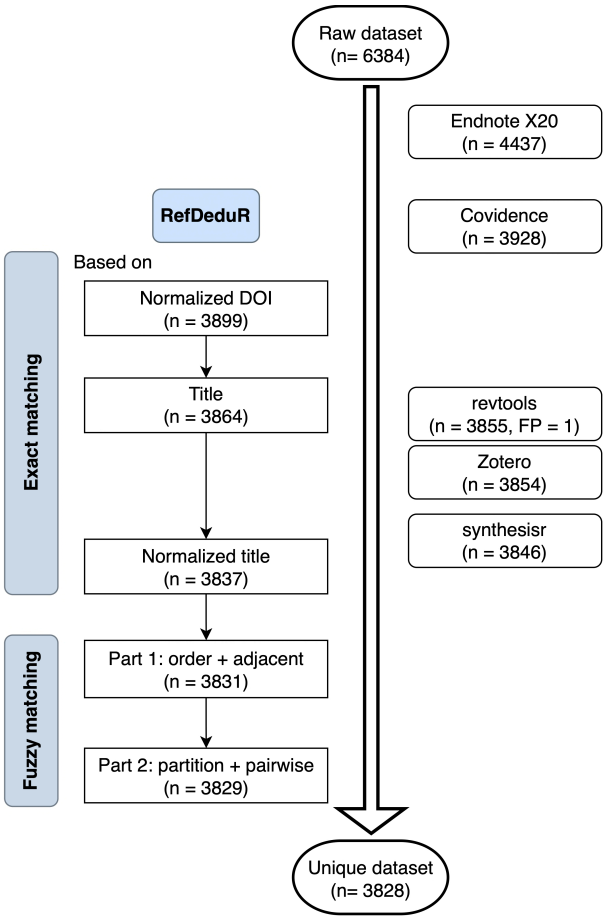


Fig. 5.3 Comparison of deduplication performance between RefDeduR and existing tools. Performance of intermediate compartments is also displayed for RefDeduR. Default settings were used for Endnote X20, Covidence, and Zotero. The tools offering multiple options (e.g., revtools and synthesisr) are represented by the highest-performance one. Except revtools, all the other tools have 0 false positives (FP) (i.e., 100% specificity) for the example dataset.

RefDeduR balances risk of misidentification and manual burden via modularizing the entire process and combining the automatic decision-tree algorithm with the attenuated manual review. In contrast, the other tools rely heavily on users to check the flagged duplicates by design, thus limiting their robustness. Following the recommended pipeline of RefDeduR, only 3 duplicate sets needed to be reviewed before achieving the manually assured result (1 in the exact matching step and 2 in part 1 of the fuzzy matching step). However, the user had to screen 1460 duplicate sets for revtools and 556 for synthesisr. The numbers of duplicate sets requiring manual review were not disclosed by Endnote X20, Covidence, and Zotero. Nevertheless, they were approximated to be 974, 1228, and 1265 assuming all were duplicate pairs. In addition, because of the decision tree, RefDeduR's manual workload is less susceptible to the change of fuzzy-matching thresholds. For instance, after lowering the similarity threshold from 0.7 to 0.6, 22 more duplicate sets were inputted into the decision tree (from 35 to 57), while the manual workload only increased by 6 (from 2 to 8) (Table S5.1).

5.4.2 General recommendations, potential limitations, and future directions

We recommend following the example pipeline when using RefDeduR, but users are offered the flexibility to build a custom pipeline. For example, if the user is satisfied with the number of output records, they may stop after the exact-matching module or part 1 of the fuzzy-matching module, since these modules are substantially faster (< 1 min) than part 2 of the fuzzy-matching module (~20 min for an exhaustive similarity calculation). Alternatively, combining the RefDeduR's intermediate or final output with other tools (e.g., Rayyan, revtools) could be a reassuring operation to further increase the possibility of achieving a desired outcome. In this way, the heavy manual review burden of these tools could also be relieved. For instance, using

RefDeduR's intermediate output from part 1 of the fuzzy-matching module as the input for Rayyan led to only 10 duplicates for manual review, as opposed to 3369 if the raw dataset was inputted.

Future versions of RefDeduR will further expand automated matching to focus on accuracy without increasing the burden on the user. For example, the decision tree could be expanded to include more fields (e.g., volume, issue and page). Due to the difficulty of obtaining a manually confirmed unique dataset, RefDeduR was only benchmarked using one dataset at this stage. Although synthetic datasets do not fully reflect the complexity of real-world datasets, we will further examine the performance of RefDudeR using mock datasets. We will continue improving the tool as more data become available. For instance, it is interesting to further explore the impact of threshold selection on software performance and the efficacy of training machine learning models for duplicate classification.

In conclusion, RefDeduR provides an effective solution to perform reference deduplication and represents a valuable advance in expanding the open-source toolkit to support evidence synthesis research. It will also support the development of future tools, just like the packages RefDeduR is built upon (e.g., revtools).

5.5 Data availability

Users can access RefDeduR on GitHub (https://github.com/jxshen311/RefDeduR) and view further documentation and examples on the website (https://jxshen311.github.io/RefDeduR/). A step-by-step tutorial is available at https://jxshen311.github.io/RefDeduR/articles/RefDeduR_tutorial.html. Source code,

supplementary data, and additional descriptions about the benchmarking analysis are available at https://github.com/jxshen311/RefDeduR_benchmark.

5.6 Supplementary information

5.6.1 Supplementary figures

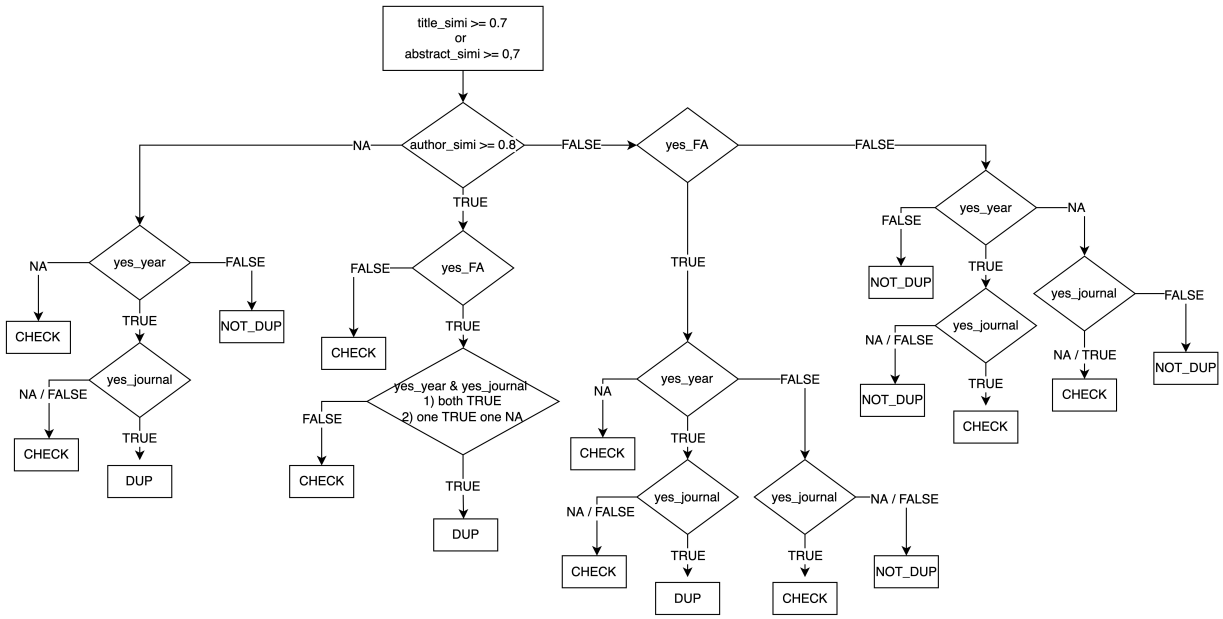


Figure S1 A detailed illustration of the decision tree used in fuzzy-matching processes.

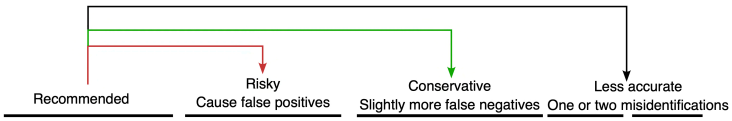
List of abbreviations:

Abbreviation	Explanation	Abbreviation	Explanation
title_simi	title similarity	yes_journal	journal initialism is identical or one of the records is a preprint or conference proceeding
abstract_simi	abstract similarity		
author_simi	author similarity	CHECK	decision is "check"
yes_FA	normalized first_author_last_name is identical	DUP	decision is "duplicate"
yes_year	year is identical or one of the records is a preprint or conference proceeding	NOT_DUP	decision is "not duplicate"

Fig. S5.1 A detailed illustration of the decision tree used in fuzzy-matching processes.

5.6.2 Supplementary tables

Table S5.1 Comparison of RefDeduR's performance for different similarity thresholds.



	Scenario	S1	S2	S3	S4	S5	S6	S7	S8
Number of records after exact matching		3837	3837	3837	3837	3837	3837	3837	3837
Fuzzy matching part 1: threshold for title similarity		0.7	0.6	0.7	0.6	0.7	0.6	0.5	0.55
Fuzzy matching part 1: threshold for abstract similarity		0.3	0.3	0.3	0.3	0.7	0.6	0.5	0.55
Number of duplicate pairs entering the decision tree		30	41	30	41	9	26	86	39
Automatic decision	check	2	4	2	4	1	3	4	3
	duplicate	5	5	5	5	3	4	4	4
	not duplicate	23	32	23	32	5	19	78	32
Manual decision for "check"	duplicate	1	1	1	1	1	1	1	1
	not duplicate	1	3	1	3	0	2	3	2
Number of records after part 1		3831	3831	3831	3831	3833	3832	3832	3832
Fuzzy matching part 2: threshold for title similarity		0.7	0.6	0.7	0.6	0.7	0.6	0.5	0.55
Fuzzy matching part 2: threshold for abstract similarity		0.7	0.6	0.3	0.3	0.7	0.6	0.5	0.55
Number of duplicate pairs entering the decision tree		5	16	92	98	5	16	42	28
Automatic decision	check	0	4	24	25	0	4	9	6
	duplicate	2	2	7	7	2	2	4	3
	not duplicate	3	10	61	66	3	10	29	19
Manual decision for "check"	duplicate	0	0	0	0	0	0	0	0
	not duplicate	0	4	24	25	0	4	9	6
Number of records after part 2 (Final outcome)		3829	3829	3824	3824	3831	3830	3828	3829
Performance statistic	False positive	0	0	4	4	0	0	2	1
	False negative	1	1	0	0	3	2	2	2
	Accuracy (%)	99.98	99.98	99.94	99.94	99.95	99.97	99.94	99.95
	Specificity (%)	100	100	99.9	99.9	100	100	99.95	99.97
	Sensitivity (%)	99.96	99.96	100	100	99.88	99.92	99.92	99.92

Table S5.2 A complete list of the manually curated benchmarking dataset.

File: Table S5.2_manually_curated_benchmarking_dataset.xls

Table S5.3 Performance statistics of different tools. The performance metrics include false positives, false negatives, accuracy, sensitivity, and specificity.

Method	Unique references retained	Duplicates flagged	FP	FN	Accuracy (%)	Sensitivity (%)	Specificity (%)
Endnote X20	4437	1947	0	609	90.46	76.17	100
Zotero	3854	2530	0	26	99.59	98.98	100
Covidence	3928	2456	0	100	98.43	96.09	100
revtools	3855	2529	1	28	99.55	98.9	99.97
synthesisr	3846	2538	0	18	99.72	99.3	100
RefDeduR	3829	2555	0	1	99.98	99.96	100

Chapter 6 Conclusions

The diverse research presented in this dissertation enhances our knowledge of diversity and antimicrobial resistance in indoor microbiomes as well as improves our capability for future exploration. By developing improved techniques and investigating previously overlooked areas, my work contributes to linking the understanding of indoor microbiomes to the promotion of human health and environmental sustainability.

In Chapter 2, I discussed 3 critical conceptual and technical aspects that need to be incorporated throughout the metagenomic environmental surveillance process (i.e., viability assessment, taxonomic resolution, and quantitation). I highlight the importance of building solid conceptual frameworks and identifying rational limits to facilitate the application of techniques. I also propose the usage of internal standards as a promising approach to overcome analytical bottlenecks introduced by low biomass samples and the inherent lack of quantitation in metagenomics. In addition, I emphasize that the field should promote crucial standardizations ranging from sampling protocols to data analysis, curation, and presentation, to increase internal consistency and external compatibility [50, 89].

Subsequently, in Chapter 3, I developed an improved and easily followable workflow for metagenomics-based environmental surveillance. Best practices were suggested in this well-structured workflow. Briefly, I recommend adopting liquid-liquid extraction to improve DNA yield and only incorporating whole-cell filtration when the non-bacterial proportion is large. I suggest including propidium monoazide treatment coupled with internal standards and absolute abundance profiling for viability assessment, and involving cultivation when demanding

comprehensive profiling. I further recommend integrating internal standards for quantification, and additionally qPCR when we expect poor taxonomic classification. I also introduce a machine learning-based model to predict required sequencing effort from accessible sample features. The model helps make full use of sequencing resources and achieve desired outcomes. This metagenomics-based environmental surveillance workflow is particularly useful in infection prevention and disinfection assessment. Although I focus on microbial surveillance of built environments, especially hospital-associated surfaces, the workflow developed in this study can be adapted to other contexts with similar characteristics. Moreover, the experience gained in overcoming challenges unique to environmental microbiomes (e.g., quantitative metagenomics with poor taxonomic classifications) is also useful to studies on other environments, such as wastewater and air.

In Chapter 4, I investigated chlorhexidine resistance in the hospital environment by combining controlled microcosm experiments and field surveys. The study revealed that chlorhexidine decreased in concentration while persisting on indoor surfaces. Surface materials and disinfection and cleaning practices impacted the patterns of chlorhexidine persistence. *Ex situ*, I found that clinically relevant bacteria could survive 24 h chlorhexidine digluconate (CHG) exposure on surfaces in the concentration range commensurate with CHG persistence. *In situ*, I found that CHG tolerance was widespread in a MICU environment. Particularly, I highlight that sinks were a critical reservoir regarding the absolute bioburden, CHG tolerance, and bacteria with exceedingly high CHG minimum inhibitory concentrations (≥ 512 μg/mL). Additionally, I encourage more attention to indoor air as a transport mechanism for resistant organisms, as evidenced by their presence on doorsills. We also found efflux to be the dominating

antimicrobial resistance mechanism, especially RND efflux pumps. This chapter advances our understanding of chlorhexidine persistence on indoor surfaces, the resistance profile of environmental isolates in a MICU, as well as potential resistance mechanisms of environmental bacteria. Moreover, this chapter will contribute to the improvement of clinical application of chlorhexidine and the prevention of antibiotic-resistant infections in healthcare facilities.

In Chapter 5, I developed an R package, RefDeduR, that performs accurate and high-throughput reference deduplication for large datasets to promote broader applications of evidence synthesis research (e.g., in indoor microbiomes). RefDeduR is text-normalization and decision-tree aided. I modularize the pipeline into text normalization, three-step exact matching, and two-step fuzzy matching processes. I also introduce a decision-tree algorithm, consider preprints when they co-exist with a peer-reviewed version, and provide actionable recommendations. Therefore, the tool is customizable, accurate, high-throughput, and practical. RefDeduR provides an effective solution to perform reference deduplication and represents a valuable advance in expanding the open-source toolkit to support evidence synthesis research.

Based on my results, I believe it is worth aiming at the following aspects in future research.

1. Indoor microbiome research benefits from better experimental, computational, and presentational standards. Experimentally, continuous advancement of internal standards for quality control, as well as quantification and other features, is still one of the major hotspots in method optimization. Computationally or analytically, we should continue creating quality control metrics. For instance, in companion with internal standards, evaluation metrics are needed. They might be analogous to sequencing coverage and

depth in metagenomics, or calculation schematics exemplified in [45]. For rigor and reproducibility, more standardized and detailed reporting is required in various sub-areas, such as high-quality metadata (e.g., sampling location, sampling methods, sampling date) and the type of antimicrobial chemicals used in MIC assays (e.g., chlorhexidine digluconate or chlorhexidine acetate).

2. In addition to optimizing techniques, concepts should be paired with applicational contexts and reevaluated as perceptions grow. For example, it is worth exploring whether the concept of "viability" in the context of interest is closer to intact cell membrane, active transcription, or active replication. For viability assessment, instead of focusing on this viable/dead dichotomy, perhaps more critically, we should keep in mind that "viability" is rather an intermediate or methodological term, linking surveillance results to questions of interest (e.g., which bacteria are infectious) [99].

3. It is beneficial to curate indoor microbiome databases, comprising metadata features, sequences, diversity and antimicrobial resistance profiles, etc. Such high-quality databases will profoundly facilitate the implementation of artificial intelligence technologies to decipher complex and high-dimensional microbiome data. They will also aid in systematic reviews and meta-analyses. Moreover, organized knowledge and recommendations in the databases will inform future research design.

4. More and larger scale investigations are essential to expand our knowledge to previously neglected areas (e.g., chlorhexidine tolerance in environments), and to provide high-quality data points in evidence synthesis and database curation.

5. It remains to be investigated how to properly interpret results generated by a combination of methods. Given the complexity, high dimensionality, and noise of microbiome data, combining multiple and ideally orthogonal techniques is usually desired. Consequently, interpretating inconsistencies between disparate methods is an imminent challenge to overcome.

Through this dissertation and future research efforts, I hope that our understanding of the microbial systems underlying human and environmental health will be continuously advanced, and better risk assessment, decision making, intervention strategies will be continuously developed to promote human health and environmental sustainability.

Reference

1. **Antimicrobial** [https://en.wikipedia.org/w/index.php?title=Antimicrobial&oldid=1016396187]
2. **Glossary | Infection Prevention & Control in Dental Settings | Division of Oral Health | CDC.** 2020.
3. **Benzalkonium chloride.** In *Meyler's Side Effects of Drugs (Sixteenth Edition).* Edited by Aronson JK. Oxford: Elsevier; 2016: 843-844
4. Green JL: **Can bioinformed design promote healthy indoor ecosystems?** *Indoor Air* 2014, **24:**113-115.
5. Dannemiller KC: **Moving towards a Robust Definition for a "Healthy" Indoor Microbiome.** *mSystems* 2019, **4**.
6. Chau CK, Tu EY, Chan DW, Burnett J: **Estimating the total exposure to air pollutants for different population age groups in Hong Kong.** *Environ Int* 2002, **27:**617-630.
7. Cincinelli A, Martellini T: **Indoor Air Quality and Health.** *Int J Environ Res Public Health* 2017, **14**.
8. Antova T, Pattenden S, Brunekreef B, Heinrich J, Rudnai P, Forastiere F, Luttmann-Gibson H, Grize L, Katsnelson B, Moshammer H, et al: **Exposure to indoor mould and children's respiratory health in the PATY study.** *J Epidemiol Community Health* 2008, **62:**708-714.
9. Fakunle AG, Jafta N, Okekunle AP, Naidoo RN: **Indoor microbiome and risk of lower respiratory tract infections among children under-five years: A meta-analysis.** *Indoor Air* 2020, **30:**795-804.
10. Fisk WJ, Eliseeva EA, Mendell MJ: **Association of residential dampness and mold with respiratory tract infections and bronchitis: a meta-analysis.** *Environ Health* 2010, **9:**72.
11. Quansah R, Jaakkola MS, Hugg TT, Heikkinen SA, Jaakkola JJ: **Residential dampness and molds and the risk of developing asthma: a systematic review and meta-analysis.** *PLoS One* 2012, **7:**e47526.
12. Gandara A, Mota LC, Flores C, Perez HR, Green CF, Gibbs SG: **Isolation of Staphylococcus aureus and antibiotic-resistant Staphylococcus aureus from residential indoor bioaerosols.** *Environ Health Perspect* 2006, **114:**1859-1864.
13. Davis MF, Iverson SA, Baron P, Vasse A, Silbergeld EK, Lautenbach E, Morris DO: **Household transmission of meticillin-resistant Staphylococcus aureus and other staphylococci.** *Lancet Infect Dis* 2012, **12:**703-716.
14. Chng KR, Li C, Bertrand D, Ng AHQ, Kwah JS, Low HM, Tong C, Natrajan M, Zhang MH, Xu L, et al: **Cartography of opportunistic pathogens and antibiotic resistance genes in a tertiary hospital environment.** *Nat Med* 2020, **26:**941-951.
15. Lax S, Sangwan N, Smith D, Larsen P, Handley KM, Richardson M, Guyton K, Krezalek M, Shogan BD, Defazio J, et al: **Bacterial colonization and succession in a newly opened hospital.** *Sci Transl Med* 2017, **9:**eaah6500.
16. Farhat M, Alkharsah KR, Alkhamis FI, Bukharie HA: **Metagenomic study on the composition of culturable and non-culturable bacteria in tap water and biofilms at intensive care units.** *Journal of Water and Health* 2018, **17:**72-83.

17. Hsu T, Joice R, Vallarino J, Abu-Ali G, Hartmann EM, Shafquat A, DuLong C, Baranowski C, Gevers D, Green JL, et al: **Urban Transit System Microbial Communities Differ by Surface Type and Interaction with Humans and the Environment.** *mSystems* 2016, **1**.
18. Kembel SW, Jones E, Kline J, Northcutt D, Stenson J, Womack AM, Bohannan BJ, Brown GZ, Green JL: **Architectural design influences the diversity and structure of the built environment microbiome.** *ISME J* 2012, **6:**1469-1479.
19. D'Souza AW, Potter RF, Wallace M, Shupe A, Patel S, Sun X, Gul D, Kwon JH, Andleeb S, Burnham CD, Dantas G: **Spatiotemporal dynamics of multidrug resistant bacteria on intensive care unit surfaces.** *Nat Commun* 2019, **10:**4569.
20. Ben Maamar S, Glawe AJ, Brown TK, Hellgeth N, Hu J, Wang JP, Huttenhower C, Hartmann EM: **Mobilizable antibiotic resistance genes are present in dust microbial communities.** *PLoS Pathog* 2020, **16:**e1008211.
21. Selway CA, Eisenhofer R, Weyrich LS: **Microbiome applications for pathology: challenges of low microbial biomass samples during diagnostic testing.** *The Journal of Pathology: Clinical Research* 2020, **6:**97-106.
22. Fruh R, Anderson A, Cieplik F, Hellwig E, Wittmer A, Vach K, Al-Ahmad A: **Antibiotic Resistance of Selected Bacteria after Treatment of the Supragingival Biofilm with Subinhibitory Chlorhexidine Concentrations.** *Antibiotics (Basel)* 2022, **11**.
23. Allen JL, Doidge NP, Bushell RN, Browning GF, Marenda MS: **Healthcare-associated infections caused by chlorhexidine-tolerant Serratia marcescens carrying a promiscuous IncHI2 multi-drug resistance plasmid in a veterinary hospital.** *PLOS ONE* 2022, **17:**e0264848.
24. Lee YH: **Strengths and Limitations of Meta-Analysis.** *Korean J Med* 2019, **94:**391-395.
25. Stone DL, Rosopa PJ: **The Advantages and Limitations of Using Meta-analysis in Human Resource Management Research.** *Human Resource Management Review* 2017, **27:**1-7.
26. Pollet T, Sprong H, Lejal E, Krawczyk AI, Moutailler S, Cosson JF, Vayssier-Taussat M, Estrada-Peña A: **The scale affects our view on the identification and distribution of microbial communities in ticks.** *Parasites and Vectors* 2020, **13**.
27. Haddaway NR, Westgate MJ: **Predicting the time needed for environmental systematic reviews and systematic maps.** *Conserv Biol* 2019, **33:**434-443.
28. Borah R, Brown AW, Capers PL, Kaiser KA: **Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry.** *BMJ Open* 2017, **7:**e012545.
29. Smith JD, Mitsakou C, Kitwiroon N, Barratt BM, Walton HA, Taylor JG, Anderson HR, Kelly FJ, Beevers SD: **London Hybrid Exposure Model: Improving Human Exposure Estimates to NO2 and PM2.5 in an Urban Setting.** *Environmental Science & Technology* 2016, **50:**11760-11768.
30. Stephens B, Azimi P, Thoemmes MS, Heidarinejad M, Allen JG, Gilbert JA: **Microbial Exchange via Fomites and Implications for Human Health.** *Current Pollution Reports* 2019, **5:**198-213.

31. Popovich KJ, Lyles R, Hayes R, Hota B, Trick W, Weinstein RA, Hayden MK: **Relationship between chlorhexidine gluconate skin concentration and microbial density on the skin of critically ill patients bathed daily with chlorhexidine gluconate.** *Infect Control Hosp Epidemiol* 2012, **33:**889-896.
32. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, Garcia-Cobos S, Kooistra-Smid AM, Raangs EC, Rosema S, Veloo AC, et al: **Application of next generation sequencing in clinical microbiology and infection prevention.** *J Biotechnol* 2017, **243:**16-24.
33. Valdez MK, Sexton JD, Lutz EA, Reynolds KA: **Spread of infectious microbes during emergency medical response.** *Am J Infect Control* 2015, **43:**606-611.
34. Xiao S, Li Y, Wong TW, Hui DSC: **Role of fomites in SARS transmission during the largest hospital outbreak in Hong Kong.** *PLoS One* 2017, **12:**e0181558.
35. Zhong Z-P, Solonenko NE, Gazitúa MC, Kenny DV, Mosley-Thompson E, Rich VI, Van Etten JL, Thompson LG, Sullivan MB: **Clean Low-Biomass Procedures and Their Application to Ancient Ice Core Microorganisms.** *Frontiers in Microbiology* 2018, **9**.
36. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS: **Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations.** *Trends in Microbiology* 2019, **27:**105-117.
37. Gomez-Silvan C, Leung MHY, Grue KA, Kaur R, Tong X, Lee PKH, Andersen GL: **A comparison of methods used to unveil the genetic and metabolic pool in the built environment.** *Microbiome* 2018, **6:**71.
38. Codony F, Dinh-Thanh M, Agustí G: **Key Factors for Removing Bias in Viability PCR-Based Methods: A Review.** *Current Microbiology* 2020, **77:**682-687.
39. Fittipaldi M, Nocker A, Codony F: **Progress in understanding preferential detection of live cells using viability dyes in combination with DNA amplification.** *Journal of Microbiological Methods* 2012, **91:**276-289.
40. Elizaquivel P, Aznar R, Sanchez G: **Recent developments in the use of viability dyes and quantitative PCR in the food microbiology field.** *J Appl Microbiol* 2014, **116:**1-13.
41. Emerson JB, Adams RI, Román CMB, Brooks B, Coil DA, Dahlhausen K, Ganz HH, Hartmann EM, Hsu T, Justice NB, et al: **Schrödinger's microbes: Tools for distinguishing the living from the dead in microbial ecosystems.** *Microbiome* 2017, **5:**86.
42. Wang Y, Yan Y, Thompson KN, Bae S, Accorsi EK, Zhang Y, Shen J, Vlamakis H, Hartmann EM, Huttenhower C: **Whole microbial community viability is not quantitatively reflected by propidium monoazide sequencing approach.** *Microbiome* 2020**:**in press.
43. Yang Y, Wan C, Xu H, Lai W, Xiong Y, Xu F, You X, Xu H, Aguilar ZP, Sun J, Wei H: **Development of a multiplexed PCR assay combined with propidium monoazide treatment for rapid and accurate detection and identification of three viable Salmonella enterica serovars.** *Food Control* 2012, **28:**456-462.
44. Pacholewicz E, Buhler C, Wulsten IF, Kraushaar B, Luu HQ, Iwobi AN, Huber I, Stingl K: **Internal sample process control improves cultivation-independent quantification of thermotolerant Campylobacter.** *Food Microbiol* 2019, **78:**53-61.

45. Fittipaldi M, Codony F, Adrados B, Camper AK, Morató J: **Viable Real-Time PCR in Environmental Samples: Can All Data Be Interpreted Directly?** *Microbial Ecology* 2011, **61:**7-12.
46. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N: **Shotgun metagenomics, from sampling to analysis.** *Nat Biotechnol* 2017, **35:**833-844.
47. Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, Perry T, Kao D, Mason AL, Madsen KL, Wong GKS: **Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics.** *Frontiers in Microbiology* 2016, **7**.
48. Zaheer R, Noyes N, Ortega Polo R, Cook SR, Marinier E, Van Domselaar G, Belk KE, Morley PS, McAllister TA: **Impact of sequencing depth on the characterization of the microbiome and resistome.** *Scientific Reports* 2018, **8:**5890.
49. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, Knight R, Knights D: **Evaluating the Information Content of Shallow Shotgun Metagenomics.** *mSystems* 2018, **3**.
50. Nayfach S, Pollard KS: **Toward Accurate and Quantitative Comparative Metagenomics.** *Cell* 2016, **166:**1103-1116.
51. Huson DH, Auch AF, Qi J, Schuster SC: **MEGAN analysis of metagenomic data.** *Genome Research* 2007, **17:**377-386.
52. Rodriguez RL, Konstantinidis KT: **Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets.** *Bioinformatics* 2014, **30:**629-635.
53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25:**2078-2079.
54. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, **26:**841-842.
55. Vezzulli L, Grande C, Tassistro G, Brettar I, Höfle MG, Pereira RPA, Mushi D, Pallavicini A, Vassallo P, Pruzzo C: **Whole-Genome Enrichment Provides Deep Insights into Vibrio cholerae Metagenome from an African River.** *Microbial Ecology* 2017, **73:**734-738.
56. Minich JJ, Zhu Q, Janssen S, Hendrickson R, Amir A, Vetter R, Hyde J, Doty MM, Stillwell K, Benardini J, et al: **KatharoSeq Enables High-Throughput Microbiome Analysis from Low-Biomass Samples.** *mSystems* 2018, **3:**e00218-00217.
57. Seitz V, Schaper S, Dröge A, Lenze D, Hummel M, Hennig S: **A new method to prevent carry-over contaminations in two-step PCR NGS library preparations.** *Nucleic Acids Research* 2015, **43:**e135-e135.
58. Silkie SS, Tolcher MP, Nelson KL: **Reagent decontamination to eliminate false-positives in Escherichia coli qPCR.** *Journal of Microbiological Methods* 2008, **72:**275-282.
59. Tamariz J, Voynarovska K, Prinz M, Caragine T: **The Application of Ultraviolet Irradiation to Exogenous Sources of DNA in Plasticware and Water for the Amplification of Low Copy Number DNA.** *Journal of Forensic Sciences* 2006, **51:**790-794.
60. McLaren MR, Willis AD, Callahan BJ: **Consistent and correctable bias in metagenomic sequencing experiments.** *eLife* 2019, **8:**e46923.

61. Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, Leite R, Elovitz MA, Parry S, Bushman FD: **Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota.** *Microbiome* 2016, **4:**29.
62. Lusk RW: **Diverse and Widespread Contamination Evident in the Unmapped Depths of High Throughput Sequencing Data.** *PLOS ONE* 2014, **9:**e110808.
63. Breitwieser FP, Lu J, Salzberg SL: **A review of methods and databases for metagenomic classification and assembly.** *Briefings in Bioinformatics* 2017, **20:**1125-1136.
64. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ: **Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data.** *Microbiome* 2018, **6:**226.
65. Martí JM: **Recentrifuge: Robust comparative analysis and contamination removal for metagenomics.** *PLOS Computational Biology* 2019, **15:**e1006967.
66. Schmieder R, Edwards R: **Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets.** *PLoS ONE* 2011, **6**.
67. Ye SH, Siddle KJ, Park DJ, Sabeti PC: **Benchmarking Metagenomics Tools for Taxonomic Classification.** *Cell* 2019, **178:**779-794.
68. Tamames J, Cobo-Simón M, Puente-Sánchez F: **Assessing the performance of different approaches for functional and taxonomic annotation of metagenomes.** *BMC Genomics* 2019, **20:**960.
69. Brito IL, Alm EJ: **Tracking Strains in the Microbiome: Insights from Metagenomics and Models.** *Frontiers in Microbiology* 2016, **7**.
70. Roe CC, Horn KS, Driebe EM, Bowers J, Terriquez JA, Keim P, Engelthaler DM: **Whole genome SNP typing to investigate methicillin-resistant Staphylococcus aureus carriage in a health-care provider as the source of multiple surgical site infections.** *Hereditas* 2016, **153**.
71. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N: **MetaPhlAn2 for enhanced metagenomic taxonomic profiling.** *Nature Methods* 2015, **12:**902-903.
72. McArdle AJ, Kaforou M: **Sensitivity of shotgun metagenomics to host DNA: abundance estimates depend on bioinformatic tools and contamination is the main issue.** *Access Microbiology* 2020, **2**.
73. Pereira-Marques J, Hout A, Ferreira RM, Weber M, Pinto-Ribeiro I, van Doorn L-J, Knetsch CW, Figueiredo C: **Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis.** *Frontiers in Microbiology* 2019, **10**.
74. Fritz A, Hofmann P, Majda S, Dahms E, Dröge J, Fiedler J, Lesker TR, Belmann P, DeMaere MZ, Darling AE, et al: **CAMISIM: simulating metagenomes and microbial communities.** *Microbiome* 2019, **7:**17.
75. Pons N, Batto JM, Kennedy S, Almeida M, Boumezbeur F, Moumen B, Leonard P, Chatelier EL, Ehrlich SD, Renault P: **METEOR, a platform for quantitative metagenomic profiling of complex ecosystems.** *Journées Ouvertes en Biologie, Informatique et Mathématiques* 2010.

76. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto JM, Kennedy S, et al: **Richness of human gut microbiome correlates with metabolic markers.** *Nature* 2013, **500:**541-546.
77. Wen C, Zheng Z, Shao T, Liu L, Xie Z, Le Chatelier E, He Z, Zhong W, Fan Y, Zhang L, et al: **Quantitative metagenomics reveals unique gut microbiome biomarkers in ankylosing spondylitis.** *Genome Biology* 2017, **18:**142.
78. Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, Guo J, Le Chatelier E, Yao J, Wu L, et al: **Alterations of the human gut microbiome in liver cirrhosis.** *Nature* 2014, **513:**59-64.
79. Ruppe E, Ghozlane A, Tap J, Pons N, Alvarez AS, Maziers N, Cuesta T, Hernando-Amado S, Clares I, Martinez JL, et al: **Prediction of the intestinal resistome by a three-dimensional structure-based method.** *Nat Microbiol* 2019, **4:**112-123.
80. Satinsky BM, Gifford SM, Crump BC, Moran MA: **Chapter Twelve - Use of Internal Standards for Quantitative Metatranscriptome and Metagenome Analysis.** In *Methods in Enzymology. Volume* 531. Edited by DeLong EF: Academic Press; 2013: 237-250
81. Frank JA, Sørensen SJ: **Quantitative Metagenomic Analyses Based on Average Genome Size Normalization.** *Applied and Environmental Microbiology* 2011, **77:**2513-2521.
82. Kerepesi C, Grolmusz V: **Evaluating the Quantitative Capabilities of Metagenomic Analysis Software.** *Current Microbiology* 2016, **72:**612-616.
83. Satinsky BM, Zielinski BL, Doherty M, Smith CB, Sharma S, Paul JH, Crump BC, Moran MA: **The Amazon continuum dataset: quantitative metagenomic and metatranscriptomic inventories of the Amazon River plume, June 2010.** *Microbiome* 2014, **2:**17.
84. Blackburn J, Wong T, Madala BS, Barker C, Hardwick SA, Reis ALM, Deveson IW, Mercer TR: **Use of synthetic DNA spike-in controls (sequins) for human genome sequencing.** *Nature Protocols* 2019, **14:**2119-2151.
85. Danzer K: **6.2.6 Alternative Calibration Procedures.** In *Analytical Chemistry: Theoretical and Metrological Fundamentals.* Heidelberg: Springer-Verlag: Springer; 2007
86. Clinical, Laboratory Standards I: **Performance standards for antimicrobial susceptibility testing.** Clinical and Laboratory Standards Institute Wayne, PA; 2017.
87. Ballenghien M, Faivre N, Galtier N: **Patterns of cross-contamination in a multispecies population genomic project: detection, quantification, impact, and solutions.** *BMC Biology* 2017, **15:**25.
88. Rodriguez RL, Konstantinidis KT: **Estimating coverage in metagenomic data sets and why it matters.** *ISME J* 2014, **8:**2349-2351.
89. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S, Abnet CC, et al: **Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium.** *Nature Biotechnology* 2017, **35:**1077-1086.
90. Brooks B, Olm MR, Firek BA, Baker R, Thomas BC, Morowitz MJ, Banfield JF: **Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome.** *Nature Communications* 2017, **8:**1814.

91. Raveh-Sadka T, Firek B, Sharon I, Baker R, Brown CT, Thomas BC, Morowitz MJ, Banfield JF: **Evidence for persistent and shared bacterial strains against a background of largely unique gut colonization in hospitalized premature infants.** *Isme j* 2016, **10:**2817-2830.
92. Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, Gibbons SM, Larsen P, Shogan BD, Weiss S, et al: **Longitudinal analysis of microbial interaction between humans and the indoor environment.** *Science* 2014, **345:**1048.
93. Vickery K, Deva A, Jacombs A, Allan J, Valente P, Gosbell IB: **Presence of biofilm containing viable multiresistant organisms despite terminal cleaning on clinical surfaces in an intensive care unit.** *Journal of Hospital Infection* 2012, **80:**52-55.
94. Constantinides B, Chau KK, Quan TP, Rodger G, Andersson MI, Jeffery K, Lipworth S, Gweon HS, Peniket A, Pike G, et al: **Genomic surveillance of Escherichia coli and Klebsiella spp. in hospital sink drains and patients.** *Microb Genom* 2020, **6**.
95. Martineau C, Li X, Lalancette C, Perreault T, Fournier E, Tremblay J, Gonzales M, Yergeau É, Quach C, Mellmann A: **Serratia marcescens Outbreak in a Neonatal Intensive Care Unit: New Insights from Next-Generation Sequencing Applications.** *Journal of Clinical Microbiology* 2018, **56:**e00235-00218.
96. Hu H, Johani K, Gosbell IB, Jacombs AS, Almatroudi A, Whiteley GS, Deva AK, Jensen S, Vickery K: **Intensive care unit environmental surfaces are contaminated by multidrug-resistant bacteria in biofilms: combined results of conventional culture, pyrosequencing, scanning electron microscopy, and confocal laser microscopy.** *J Hosp Infect* 2015, **91:**35-44.
97. Burgos-Garay M, Ganim C, de Man TJB, Davy T, Mathers AJ, Kotay S, Daniels J, Perry KA, Breaker E, Donlan RM: **Colonization of carbapenem-resistant Klebsiella pneumoniae in a sink-drain model biofilm system.** *Infect Control Hosp Epidemiol* 2021, **42:**722-730.
98. Kotay S, Chai W, Guilford W, Barry K, Mathers AJ: **Spread from the Sink to the Patient: In Situ Study Using Green Fluorescent Protein (GFP)-Expressing Escherichia coli To Model Bacterial Dispersion from Hand-Washing Sink-Trap Reservoirs.** *Applied and environmental microbiology* 2017, **83:**e03327-03316.
99. Shen J, McFarland AG, Young VB, Hayden MK, Hartmann EM: **Toward Accurate and Robust Environmental Surveillance Using Metagenomics.** *Frontiers in Genetics* 2021, **12**.
100. Bowers RM, Clum A, Tice H, Lim J, Singh K, Ciobanu D, Ngan CY, Cheng JF, Tringe SG, Woyke T: **Impact of library preparation protocols and template quantity on the metagenomic reconstruction of a mock microbial community.** *BMC Genomics* 2015, **16:**856.
101. Mahnert A, Moissl-Eichinger C, Zojer M, Bogumil D, Mizrahi I, Rattei T, Martinez JL, Berg G: **Man-made microbial resistances in built environments.** *Nature Communications* 2019, **10:**968.
102. Nocker A, Cheung CY, Camper AK: **Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells.** *J Microbiol Methods* 2006, **67:**310-320.

103. Wang Y, Yan Y, Thompson KN, Bae S, Accorsi EK, Zhang Y, Shen J, Vlamakis H, Hartmann EM, Huttenhower C: **Whole microbial community viability is not quantitatively reflected by propidium monoazide sequencing approach.** *Microbiome* 2021, **9:**17.
104. Morton JT, Marotz C, Washburne A, Silverman J, Zaramela LS, Edlund A, Zengler K, Knight R: **Establishing microbial composition measurement standards with reference frames.** *Nat Commun* 2019, **10:**2719.
105. Liwinski T, Leshem A, Elinav E: **Breakthroughs and Bottlenecks in Microbiome Research.** *Trends Mol Med* 2021, **27:**298-301.
106. Vandeputte D, Kathagen G, D'Hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, Wang J, Tito RY, De Commer L, Darzi Y, et al: **Quantitative microbiome profiling links gut community variation to microbial load.** *Nature* 2017, **551:**507-511.
107. Hingamp P, Grimsley N, Acinas SG, Clerissi C, Subirana L, Poulain J, Ferrera I, Sarmento H, Villar E, Lima-Mendez G, et al: **Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes.** *ISME J* 2013, **7:**1678-1695.
108. Yamamoto N, Hospodsky D, Dannemiller KC, Nazaroff WW, Peccia J: **Indoor emissions as a primary source of airborne allergenic fungal particles in classrooms.** *Environ Sci Technol* 2015, **49:**5098-5106.
109. Tkacz A, Hortala M, Poole PS: **Absolute quantitation of microbiota abundance in environmental samples.** *Microbiome* 2018, **6:**110.
110. Stammler F, Glasner J, Hiergeist A, Holler E, Weber D, Oefner PJ, Gessner A, Spang R: **Adjusting microbiome profiles for differences in microbial load by spike-in bacteria.** *Microbiome* 2016, **4:**28.
111. Rodriguez-R LM, Gunturu S, Tiedje JM, Cole JR, Konstantinidis KT: **Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity.** *mSystems* 2018, **3:**e00039-00018.
112. O'Hara NB, Reed HJ, Afshinnekoo E, Harvin D, Caplan N, Rosen G, Frye B, Woloszynek S, Ounit R, Levy S, et al: **Metagenomic characterization of ambulances across the USA.** *Microbiome* 2017, **5:**125.
113. Shen J, Hartmann EM: **Collect, extract, pool, and cultivate surface swab samples from built environments.** *Protocol Exchange* 2022.
114. Blaustein RA, Michelitsch LM, Glawe AJ, Lee H, Huttelmaier S, Hellgeth N, Ben Maamar S, Hartmann EM: **Toothbrush microbiomes feature a meeting ground for human oral and environmental microbiota.** *Microbiome* 2021, **9:**32.
115. Shen J, Hartmann EM: **Snap freezing of environmental microbial samples in liquid.** *Protocol Exchange* 2022.
116. Wang X, Howe S, Deng F, Zhao J: **Current Applications of Absolute Bacterial Quantification in Microbiome Studies and Decision-Making Regarding Different Biological Questions.** *Microorganisms* 2021, **9:**1797.
117. Fahimipour AK, Hartmann EM, Siemens A, Kline J, Levin DA, Wilson H, Betancourt-Román CM, Brown G, Fretz M, Northcutt D, et al: **Daylight exposure modulates bacterial communities associated with household dust.** *Microbiome* 2018, **6:**175.
118. Shen J, Rose LJ, Perry-Dow KA, Hartmann EM: **An optimized protocol for propidium monoazide treatment.** *Protocol Exchange* 2021.

119. Adams RI, Bateman AC, Bik HM, Meadow JF: **Microbiota of the indoor environment: a meta-analysis.** *Microbiome* 2015, **3:**49.
120. Leung MHY, Tong X, Boifot KO, Bezdan D, Butler DJ, Danko DC, Gohli J, Green DC, Hernandez MT, Kelly FJ, et al: **Characterization of the public transit air microbiome and resistome reveals geographical specificity.** *Microbiome* 2021, **9:**112.
121. Shen J, Hartmann EM: **DNA extraction protocol for low-biomass environmental samples: adapted from the Lucigen MasterPure Complete DNA and RNA Purification Kit manual.** *Protocol Exchange* 2021.
122. Singer VL, Jones LJ, Yue ST, Haugland RP: **Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation.** *Anal Biochem* 1997, **249:**228-238.
123. Hu J, Ben Maamar S, Glawe AJ, Gottel N, Gilbert JA, Hartmann EM: **Impacts of indoor surface finishes on bacterial viability.** *Indoor Air* 2019, **29:**551-562.
124. Bengtsson-Palme J, Richardson RT, Meola M, Wurzbacher C, Tremblay ED, Thorell K, Kanger K, Eriksson KM, Bilodeau GJ, Johnson RM, et al: **Metaxa2 Database Builder: enabling taxonomic identification from metagenomic or metabarcoding data using any genetic marker.** *Bioinformatics* 2018, **34:**4027-4033.
125. Bengtsson-Palme J, Thorell K, Wurzbacher C, Sjöling Å, Nilsson RH: **Metaxa2 Diversity Tools: Easing microbial community analysis with Metaxa2.** *Ecological Informatics* 2016, **33:**45-50.
126. Bengtsson-Palme J, Hartmann M, Eriksson KM, Pal C, Thorell K, Larsson DG, Nilsson RH: **METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data.** *Mol Ecol Resour* 2015, **15:**1403-1414.
127. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glockner FO: **Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies.** *Nucleic Acids Res* 2013, **41:**e1.
128. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO: **The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks.** *Nucleic acids research* 2014, **42:**D643-D648.
129. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C: **Metagenomic microbial community profiling using unique clade-specific marker genes.** *Nature Methods* 2012, **9:**811-814.
130. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, Mailyan A, Manghi P, Scholz M, Thomas AM, et al: **Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3.** *eLife* 2021, **10:**e65088.
131. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N: **Microbial strain-level population structure and genetic diversity from metagenomes.** *Genome Res* 2017, **27:**626-638.
132. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W: **MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph.** *Bioinformatics* 2015, **31:**1674-1676.
133. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA: **metaSPAdes: a new versatile metagenomic assembler.** *Genome Res* 2017, **27:**824-834.

134. Gurevich A, Saveliev V, Vyahhi N, Tesler G: **QUAST: quality assessment tool for genome assemblies.** *Bioinformatics* 2013, **29:**1072-1075.
135. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z: **MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies.** *PeerJ* 2019, **7:**e7359.
136. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW: **CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.** *Genome Res* 2015, **25:**1043-1055.
137. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH: **GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database.** *Bioinformatics* 2019, **36:**1925-1927.
138. Team RC: **R: A language and environment for statistical computing.** 2013.
139. Wickham H: *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York; 2016.
140. Leung MH, Wilkins D, Li EK, Kong FK, Lee PK: **Indoor-air microbiome in an urban subway network: diversity and dynamics.** *Appl Environ Microbiol* 2014, **80:**6760-6770.
141. Davis A, Kohler C, Alsallaq R, Hayden R, Maron G, Margolis E: **Improved yield and accuracy for DNA extraction in microbiome studies with variation in microbial biomass.** *BioTechniques* 2019, **66:**285-289.
142. Mo Y, Zhang W, Yang J, Lin Y, Yu Z, Lin S: **Biogeographic patterns of abundant and rare bacterioplankton in three subtropical bays resulting from selective and neutral processes.** *ISME J* 2018, **12:**2198-2210.
143. Nyirabuhoro P, Liu M, Xiao P, Liu L, Yu Z, Wang L, Yang J: **Seasonal Variability of Conditionally Rare Taxa in the Water Column Bacterioplankton Community of Subtropical Reservoirs in China.** *Microbial Ecology* 2020, **80:**14-26.
144. Ji BW, Sheth RU, Dixit PD, Huang Y, Kaufman A, Wang HH, Vitkup D: **Quantifying spatiotemporal variability and noise in absolute microbiota abundances using replicate sampling.** *Nature methods* 2019, **16:**731-736.
145. Barbau-Piednoir E, Mahillon J, Pillyser J, Coucke W, Roosens NH, Botteldoorn N: **Evaluation of viability-qPCR detection system on viable and dead Salmonella serovar Enteritidis.** *J Microbiol Methods* 2014, **103:**131-137.
146. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO: **The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.** *Nucleic Acids Res* 2013, **41:**D590-596.
147. Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A: **Descriptive statistics and normality tests for statistical data.** *Annals of cardiac anaesthesia* 2019, **22:**67-72.
148. Glass EM, Dribinsky Y, Yilmaz P, Levin H, Van Pelt R, Wendel D, Wilke A, Eisen JA, Huse S, Shipanova A, et al: **MIxS-BE: a MIxS extension defining a minimum information standard for sequence data from the built environment.** *The ISME Journal* 2014, **8:**1-3.
149. El Khouli RH, Macura KJ, Barker PB, Habba MR, Jacobs MA, Bluemke DA: **Relationship of temporal resolution to diagnostic performance for dynamic contrast enhanced MRI of the breast.** *J Magn Reson Imaging* 2009, **30:**999-1004.

150. Escobar-Zepeda A, Godoy-Lozano EE, Raggi L, Segovia L, Merino E, Gutierrez-Rios RM, Juarez K, Licea-Navarro AF, Pardo-Lopez L, Sanchez-Flores A: **Analysis of sequencing strategies and tools for taxonomic annotation: Defining standards for progressive metagenomics.** *Sci Rep* 2018, **8:**12034.
151. Wang Y, Thompson KN, Yan Y, Short MI, Zhang Y, Franzosa EA, Shen J, Hartmann EM, Huttenhower C: **Characterizing microbial community viability with RNA-based high-throughput sequencing.** *Research Square* 2022.
152. Joseph TA, Chlenski P, Litman A, Korem T, Pe'er I: **Accurate and robust inference of microbial growth dynamics from metagenomic sequencing reveals personalized growth rates.** *Genome Res* 2022, **32:**558-568.
153. Emiola A, Oh J: **High throughput in situ metagenomic measurement of bacterial replication at ultra-low sequencing coverage.** *Nature Communications* 2018, **9:**4956.
154. Gao Y, Li H: **Quantifying and comparing bacterial growth dynamics in multiple metagenomic samples.** *Nature Methods* 2018, **15:**1041-1044.
155. Korem T, Zeevi D, Suez J, Weinberger A, Avnit-Sagi T, Pompan-Lotan M, Matot E, Jona G, Harmelin A, Cohen N, et al: **Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples.** *Science* 2015, **349:**1101-1106.
156. Long AM, Hou S, Ignacio-Espinoza JC, Fuhrman JA: **Benchmarking microbial growth rate predictions from metagenomes.** *The ISME Journal* 2021, **15:**183-195.
157. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K: **Improving saliva shotgun metagenomics by chemical host DNA depletion.** *Microbiome* 2018, **6:**42.
158. Nelson MT, Pope CE, Marsh RL, Wolter DJ, Weiss EJ, Hager KR, Vo AT, Brittnacher MJ, Radey MC, Hayden HS, et al: **Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles.** *Cell Rep* 2019, **26:**2227-2240 e2225.
159. Amar Y, Lagkouvardos I, Silva RL, Ishola OA, Foesel BU, Kublik S, Schöler A, Niedermeier S, Bleuel R, Zink A, et al: **Pre-digest of unprotected DNA by Benzonase improves the representation of living skin bacteria and efficiently depletes host DNA.** *Microbiome* 2021, **9:**123.
160. !!! INVALID CITATION !!! [23].
161. Barlow JT, Bogatyrev SR, Ismagilov RF: **A quantitative sequencing framework for absolute abundance measurements of mucosal and lumenal microbial communities.** *Nat Commun* 2020, **11:**2590.
162. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, Tramontano M, Driessen M, Hercog R, Jung FE, et al: **Towards standards for human fecal sample processing in metagenomic studies.** *Nat Biotechnol* 2017, **35:**1069-1076.
163. Hardwick SA, Chen WY, Wong T, Kanakamedala BS, Deveson IW, Ongley SE, Santini NS, Marcellin E, Smith MA, Nielsen LK, et al: **Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis.** *Nat Commun* 2018, **9:**3096.
164. Tourlousse DM, Yoshiike S, Ohashi A, Matsukura S, Noda N, Sekiguchi Y: **Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing.** *Nucleic Acids Res* 2017, **45:**e23.

165. Cheung HY, Wong MM, Cheung SH, Liang LY, Lam YW, Chiu SK: **Differential actions of chlorhexidine on the cell wall of Bacillus subtilis and Escherichia coli.** *PLoS One* 2012, **7:**e36659.
166. Malakova IHT: **Effect on Skin Microflora in Hygienic Bathing of Patients with Chlorhexidine Gluconate Disinfectant and Non-Rinse Product.** *Health Sci J* 2020.
167. Rupp ME, Cavalieri RJ, Lyden E, Kucera J, Martin M, Fitzgerald T, Tyner K, Anderson JR, VanSchooneveld TC: **Effect of hospital-wide chlorhexidine patient bathing on healthcare-associated infections.** *Infect Control Hosp Epidemiol* 2012, **33:**1094-1100.
168. Das R, Sarkar S, Bhattacharjee C: **Photocatalytic degradation of chlorhexidine—A chemical assessment and prediction of optimal condition by response surface methodology.** *Journal of Water Process Engineering* 2014, **2:**79-86.
169. Sarkar S, Bhattacharjee C, Curcio S: **Studies on adsorption, reaction mechanisms and kinetics for photocatalytic degradation of CHD, a pharmaceutical waste.** *Ecotoxicol Environ Saf* 2015, **121:**154-163.
170. Zong Z, Kirsch LE: **Studies on the Instability of Chlorhexidine, Part I: Kinetics and Mechanisms.** *Journal of Pharmaceutical Sciences* 2012, **101:**2417-2427.
171. Kampf G: **Acquired resistance to chlorhexidine - is it time to establish an 'antiseptic stewardship' initiative?** *J Hosp Infect* 2016, **94:**213-227.
172. Kõljalg S, Naaber P, Mikelsaar M: **Antibiotic resistance as an indicator of bacterial chlorhexidine susceptibility.** *Journal of Hospital Infection* 2002, **51:**106-113.
173. Keck N, Dunie-Merigot A, Dazas M, Hirchaud E, Laurence S, Gervais B, Madec JY, Haenni M: **Long-lasting nosocomial persistence of chlorhexidine-resistant Serratia marcescens in a veterinary hospital.** *Vet Microbiol* 2020, **245:**108686.
174. Vigeant P, Loo VG, Bertrand C, Dixon C, Hollis R, Pfaller MA, McLean AP, Briedis DJ, Perl TM, Robson HG: **An outbreak of Serratia marcescens infections related to contaminated chlorhexidine.** *Infect Control Hosp Epidemiol* 1998, **19:**791-794.
175. Okuda T, Endo N, Osada Y, Zen-Yoji H: **Outbreak of nosocomial urinary tract infections caused by Serratia marcescens.** *Journal of Clinical Microbiology* 1984, **20:**691-695.
176. Carey DE, McNamara PJ: **The impact of triclosan on the spread of antibiotic resistance in the environment.** *Front Microbiol* 2014, **5:**780.
177. Jutkina J, Marathe NP, Flach CF, Larsson DGJ: **Antibiotics and common antibacterial biocides stimulate horizontal transfer of resistance at low concentrations.** *Sci Total Environ* 2018, **616-617:**172-178.
178. Hartmann EM, Hickey R, Hsu T, Betancourt Roman CM, Chen J, Schwager R, Kline J, Brown GZ, Halden RU, Huttenhower C, Green JL: **Antimicrobial Chemicals Are Associated with Elevated Antibiotic Resistance Genes in the Indoor Dust Microbiome.** *Environ Sci Technol* 2016, **50:**9807-9815.
179. Vásquez-Giraldo DF, Libreros-Zúñiga GA, Crespo-Ortiz MdP: **Effects of biocide exposure on P. Aeruginosa, E. coli and A. Baumannii complex isolates from hospital and household environments.** *Infectio* 2017, **21:**243-250.
180. Braga TM, Pomba C, Lopes MFS: **High-level vancomycin resistant Enterococcus faecium related to humans and pigs found in dust from pig breeding facilities.** *Veterinary Microbiology* 2013, **161:**344-349.

181. Namaki Kheljan M, Teymorpour R, Peeri Doghaheh H, Arzanlou M: **Antimicrobial Biocides Susceptibility and Tolerance-Associated Genes in Enterococcus faecalis and Enterococcus faecium Isolates Collected from Human and Environmental Sources.** *Current Microbiology* 2022, **79:**170.
182. Evans HL, Dellit TH, Chan J, Nathens AB, Maier RV, Cuschieri J: **Effect of chlorhexidine whole-body bathing on hospital-acquired infections among trauma patients.** *Arch Surg* 2010, **145:**240-246.
183. Climo MW, Yokoe DS, Warren DK, Perl TM, Bolon M, Herwaldt LA, Weinstein RA, Sepkowitz KA, Jernigan JA, Sanogo K, Wong ES: **Effect of daily chlorhexidine bathing on hospital-acquired infection.** *N Engl J Med* 2013, **368:**533-542.
184. Arnold B, Blum A, Branyan J, Bruton T, Carignan C, Cortopassi G, Datta S, DeWitt J, Doherty A-C, Halden RU, et al: **Review of Quaternary Ammonium Compounds: A Chemical Class of Emerging Concern.** *Environmental Science & Technology*.
185. Shen J, McFarland AG, Blaustein RA, Rose LJ, Perry-Dow KA, Moghadam AA, Hayden MK, Young VB, Hartmann EM: **An improved workflow for accurate and robust healthcare environmental surveillance using metagenomics.** *Microbiome* 2022, **10:**206.
186. Edmiston CE, Jr., Krepel CJ, Seabrook GR, Lewis BD, Brown KR, Towne JB: **Preoperative shower revisited: can high topical antiseptic levels be achieved on the skin surface before surgical admission?** *J Am Coll Surg* 2008, **207:**233-239.
187. Gefter Shenderovich J, Zaks B, Kirmayer D, Lavy E, Steinberg D, Friedman M: **Chlorhexidine sustained-release varnishes for catheter coating - Dissolution kinetics and antibiofilm properties.** *Eur J Pharm Sci* 2018, **112:**1-7.
188. Rintala H, Pitkaranta M, Taubel M: **Microbial communities associated with house dust.** *Adv Appl Microbiol* 2012, **78:**75-120.
189. Lax S, Sangwan N, Smith D, Larsen P, Handley KM, Richardson M, Guyton K, Krezalek M, Shogan BD, Defazio J, et al: **Bacterial colonization and succession in a newly opened hospital.** *Sci Transl Med* 2017, **9**.
190. Huslage K, Rutala WA, Sickbert-Bennett E, Weber DJ: **A quantitative approach to defining "high-touch" surfaces in hospitals.** *Infect Control Hosp Epidemiol* 2010, **31:**850-853.
191. Sehulster LM, Chinn RYW, Arduino MJ, Carpenter J, Donlan R, Ashford D, Besser R, Fields B, McNeil MM, Whitney C, et al: **Guidelines for environmental infection control in health-care facilities. Recommendations from CDC and the Healthcare Infection Control Practices Advisory Committee (HICPAC).** Chicago IL; 2004.
192. Jinadatha C, Villamaria FC, Coppin JD, Dale CR, Williams MD, Whitworth R, Stibich M: **Interaction of healthcare worker hands and portable medical equipment: a sequence analysis to show potential transmission opportunities.** *BMC Infect Dis* 2017, **17:**800.
193. Cheng VC, Chau PH, Lee WM, Ho SK, Lee DW, So SY, Wong SC, Tai JW, Yuen KY: **Hand-touch contact assessment of high-touch and mutual-touch surfaces among healthcare workers, patients, and visitors.** *J Hosp Infect* 2015, **90:**220-225.

194. Xu S, An C, Kim S, Lee S, Lee K, Yamamoto N: **Effects of the biocides on the culturable house dust-borne bacterial compositions and diversities.** *Human and Ecological Risk Assessment: An International Journal* 2016, **22:**1133-1146.
195. Lederberg J, Lederberg EM: **Replica plating and indirect selection of bacterial mutants.** *Journal of bacteriology* 1952, **63:**399-406.
196. McFarland AG, Bertucci HK, Littman E, Shen J, Huttenhower C, Hartmann EM: **Triclosan Tolerance Is Driven by a Conserved Mechanism in Diverse Pseudomonas Species.** *Applied and Environmental Microbiology* 2021, **87**.
197. CLSI: **Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically. 11th ed. CLSI standard M07.** Wayne, PA: Clinical and Laboratory Standards Institute; 2018.
198. Denkel LA, Kramer TS, Schwab F, Golembus J, Wolke S, Gastmeier P, Geffers C: **Chlorhexidine and octenidine susceptibility of bacterial isolates from clinical samples in a three-armed cluster randomised decolonisation trial.** *PLoS One* 2022, **17:**e0278569.
199. Horner C, Mawer D, Wilcox M: **Reduced susceptibility to chlorhexidine in staphylococci: is it increasing and does it matter?** *J Antimicrob Chemother* 2012, **67:**2547-2559.
200. Morrissey I, Oggioni MR, Knight D, Curiao T, Coque T, Kalkanci A, Martinez JL, the BC: **Evaluation of Epidemiological Cut-Off Values Indicates that Biocide Resistant Subpopulations Are Uncommon in Natural Isolates of Clinically-Relevant Microorganisms.** *PLOS ONE* 2014, **9:**e86669.
201. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al: **SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.** *J Comput Biol* 2012, **19:**455-477.
202. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen AV, Cheng AA, Liu S, et al: **CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database.** *Nucleic Acids Res* 2020, **48:**D517-D525.
203. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, et al: **The Comprehensive Antibiotic Resistance Database.** *Antimicrobial Agents and Chemotherapy* 2013, **57:**3348.
204. Wickham H, François R, Henry L, Müller K: **dplyr: A Grammar of Data Manipulation.** 2022.
205. West AM, Nkemngong CA, Voorn MG, Wu T, Li X, Teska PJ, Oliver HF: **Surface area wiped, product type, and target strain impact bactericidal efficacy of ready-to-use disinfectant Towelettes.** *Antimicrob Resist Infect Control* 2018, **7:**122.
206. Tran G, Huynh T, Bruins FM, Ahmad N, Budris WA, Posligua AL, Hammel JA, Nardone B, West DP: **Evidence of incompatibility for topical anionic agents used in conjunction with chlorhexidine gluconate: A systematic review.** *Journal of Surgical Dermatology* 2016.

207. Kaiser N, Klein D, Karanja P, Greten Z, Newman J: **Inactivation of chlorhexidine gluconate on skin by incompatible alcohol hand sanitizing gels.** *Am J Infect Control* 2009, **37:**569-573.
208. Zong Z, Kirsch LE: **Studies on the instability of chlorhexidine, part I: kinetics and mechanisms.** *J Pharm Sci* 2012, **101:**2417-2427.
209. Abu-Obaid E, Salama F, Abu-Obaid A, Alanazi F, Salem M, Auda S: **Comparative Evaluation of the Antimicrobial Effects of Different Mouthrinses against Streptococcus Mutans: An in Vitro Study.** *J Clin Pediatr Dent* 2019, **43:**398-407.
210. Choi JS: **[Evaluation of a waterless, scrubless chlorhexidine gluconate/ethanol surgical scrub and povidone-iodine for antimicrobial efficacy].** *Taehan Kanho Hakhoe Chi* 2008, **38:**39-44.
211. Pons JL, Bonnaveiro N, Chevalier J, Crémieux A: **Evaluation of antimicrobial interactions between chlorhexidine, quaternary ammonium compounds, preservatives and excipients.** *J Appl Bacteriol* 1992, **73:**395-400.
212. Gkatzonis AM, Vassilopoulos SI, Karoussis IK, Kaminari A, Madianos PN, Vrotsos IA: **A randomized controlled clinical trial on the effectiveness of three different mouthrinses (chlorhexidine with or without alcohol and C31G), adjunct to periodontal surgery, in early wound healing.** *Clin Oral Investig* 2018, **22:**2581-2591.
213. Marty Cooney R, Manickam N, Becherer P, Harmon LS, Gregg L, Farkas Z, Shea LM, Parekh P, Murphy J, Shade W: **The use of 3.15% chlorhexidine gluconate/70% alcohol hub disinfection to prevent central line-associated bloodstream infections in dialysis patients.** *Br J Nurs* 2020, **29:**S24-s26.
214. Purdy KR: **Aspects of chlorhexidine degradation.** University of Bath, Department of Life Sciences; 1987.
215. Adams RI, Tian Y, Taylor JW, Bruns TD, Hyvarinen A, Taubel M: **Passive dust collectors for assessing airborne microbial material.** *Microbiome* 2015, **3:**46.
216. Barberan A, Dunn RR, Reich BJ, Pacifici K, Laber EB, Menninger HL, Morton JM, Henley JB, Leff JW, Miller SL, Fierer N: **The ecology of microscopic life in household dust.** *Proc Biol Sci* 2015, **282**.
217. Emerson JB, Keady PB, Clements N, Morgan EE, Awerbuch J, Miller SL, Fierer N: **High temporal variability in airborne bacterial diversity and abundance inside single-family residences.** *Indoor Air* 2017, **27:**576-586.
218. Sum R, Swaminathan M, Rastogi SK, Piloto O, Cheong I: **Beta-Hemolytic Bacteria Selectively Trigger Liposome Lysis, Enabling Rapid and Accurate Pathogen Detection.** *ACS Sens* 2017, **2:**1441-1451.
219. FA P, DJ F: **Prevalence of gentamicin- and amikacin-resistant bacteria in sink drains.** *Journal of clinical microbiology* 1980, **12**.
220. MH L, B O, C N, SA K, RA W: **Pseudomonas in the sinks in an intensive care unit: relation to patients.** *Journal of clinical pathology* 1984, **37**.
221. Rutala WA, Weber DJ: **Water as a reservoir of nosocomial pathogens.** *Infect Control Hosp Epidemiol* 1997, **18:**609-616.
222. Aftab R, Dodhia VH, Jeanes C, Wade RG: **Bacterial sensitivity to chlorhexidine and povidone-iodine antiseptics over time: a systematic review and meta-analysis of human-derived data.** *Scientific Reports* 2023, **13:**347.

223. Buxser S: **Has resistance to chlorhexidine increased among clinically-relevant bacteria? A systematic review of time course and subpopulation data.** *PLOS ONE* 2021, **16:**e0256336.
224. Kawamura-Sato K, Wachino JI, Kondo T, Ito H, Arakawa Y: **Correlation between reduced susceptibility to disinfectants and multidrug resistance among clinical isolates of Acinetobacter species.** *Journal of Antimicrobial Chemotherapy* 2010, **65:**1975-1983.
225. Lambert RJW, Joynson J, Forbes B: **The relationships and susceptibilities of some industrial, laboratory and clinical isolates of pseudomonas aeruginosa to some antibiotics and biocides.** *Journal of Applied Microbiology* 2001, **91:**972-984.
226. Lepainteur M, Royer G, Bourrel AS, Romain O, Duport C, Doucet-Populaire F, Decousser JW: **Prevalence of resistance to antiseptics and mupirocin among invasive coagulase-negative staphylococci from very preterm neonates in NICU: The creeping threat?** *Journal of Hospital Infection* 2013, **83:**333-336.
227. Long M, Lai H, Deng W, Zhou K, Li B, Liu S, Fan L, Wang H, Zou L: **Disinfectant susceptibility of different Salmonella serotypes isolated from chicken and egg production chains.** *Journal of Applied Microbiology* 2016, **121:**672-681.
228. Cowley NL, Forbes S, Amezquita A, McClure P, Humphreys GJ, McBain AJ: **Effects of Formulation on Microbicide Potency and Mitigation of the Development of Bacterial Insusceptibility.** *Appl Environ Microbiol* 2015, **81:**7330-7338.
229. de Frutos M, Lopez-Urrutia L, Dominguez-Gil M, Arias M, Munoz-Bellido JL, Eiros JM, Ramos C: **Serratia marcescens outbreak due to contaminated 2% aqueous chlorhexidine.** *Enferm Infecc Microbiol Clin* 2017, **35:**624-629.
230. Brooke JS: **Stenotrophomonas maltophilia: an emerging global opportunistic pathogen.** *Clin Microbiol Rev* 2012, **25:**2-41.
231. Steinberg JP, Burd EM: **238 - Other Gram-Negative and Gram-Variable Bacilli.** In *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (Eighth Edition).* Edited by Bennett JE, Dolin R, Blaser MJ. Philadelphia: W.B. Saunders; 2015: 2667-2683.e2664
232. Zhu M, Zhu Q, Yang Z, Liang Z: **Clinical Characteristics of Patients with Micrococcus luteus Bloodstream Infection in a Chinese Tertiary-Care Hospital.** *Pol J Microbiol* 2021, **70:**321-326.
233. Crofts TS, McFarland AG, Hartmann EM: **Mosaic Ends Tagmentation (METa) Assembly for Highly Efficient Construction of Functional Metagenomic Libraries.** *mSystems* 2021, **6:**e0052421.
234. Bramer WM, De Jonge GB, Rethlefsen ML, Mast F, Kleijnen J: **A systematic approach to searching: an efficient and complete method to develop literature searches.** *Journal of the Medical Library Association* 2018, **106**.
235. Jiang Y, Lin C, Meng W, Yu C, Cohen AM, Smalheiser NR: **Rule-based deduplication of article records from bibliographic databases.** *Database (Oxford)* 2014, **2014:**bat086.
236. Rathbone J, Carter M, Hoffmann T, Glasziou P: **Better duplicate detection for systematic reviewers: evaluation of Systematic Review Assistant-Deduplication Module.** *Syst Rev* 2015, **4:**6.

237. Westgate MJ: **revtools: An R package to support article screening for evidence synthesis.** *Res Synth Methods* 2019, **10:**606-614.
238. van der Loo MPJ: **The stringdist package for approximate string matching.** *The R Journal* 2014, **6:**111-122.
239. Wickham H: **stringr: Simple, consistent wrappers for common string operations.** 2022.
240. Wickham H, François R, Henry L, Müller K: **dplyr: A grammar of data manipulation.** 2022.
241. Wickham H, Girlich M: **tidyr: Tidy messy data.** 2022.
242. Grames EM, Stillman AN, Tingley MW, Elphick CS: **An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks.** *Methods in Ecology and Evolution* 2019, **10:**1645-1654.
243. McKeown S, Mir ZM: **Considerations for conducting systematic reviews: evaluating the performance of different methods for de-duplicating references.** *Syst Rev* 2021, **10:**38.