



Toward Accurate and Robust Environmental Surveillance Using Metagenomics

Jiaxian Shen¹, Alexander G. McFarland¹, Vincent B. Young², Mary K. Hayden³ and Erica M. Hartmann^{1*}

¹ Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL, United States, ² Division of Infectious Diseases, Department of Internal Medicine, The University of Michigan Medical School, Ann Arbor, MI, United States, ³ Division of Infectious Diseases, Department of Internal Medicine, Rush University Medical Center, Chicago, IL, United States

OPEN ACCESS

Edited by:

Nicole Hansmeier,
Luther College, Canada

Reviewed by:

Kyle B. Gustafson,
Naval Surface Warfare Center
Carderock Division, United States
Alexander Mahner,
Medical University of Graz, Austria

*Correspondence:

Erica M. Hartmann
erica.hartmann@northwestern.edu

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 28 August 2020

Accepted: 21 January 2021

Published: 05 March 2021

Citation:

Shen J, McFarland AG,
Young VB, Hayden MK and
Hartmann EM (2021) Toward
Accurate and Robust Environmental
Surveillance Using Metagenomics.
Front. Genet. 12:600111.
doi: 10.3389/fgene.2021.600111

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.

Keywords: viability, limit of detection, metagenomics, taxonomic resolution, environmental surveillance, quantitative metagenomics

INTRODUCTION

Approximately 56% of the world's population lives in urban areas (United Nations, 2018) and people in developed nations spend at least 90% of the time indoors (Chau et al., 2002; Smith et al., 2016; Cincinelli and Martellini, 2017), making built environments hotspots with which humans frequently interact. Understanding and monitoring fomite transmission is critical in infection prevention (Stephens et al., 2019). 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 (Popovich et al., 2012). 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 (Deurenberg et al., 2017). 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 (Valdez et al., 2015; Xiao et al., 2017). 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 (Zhong et al., 2018; Selway et al., 2020). 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 (Eisenhofer et al., 2019).

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 (Figure 1A).

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 (Gomez-Silvan et al., 2018). 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 (Codony et al., 2020), 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 (Figure 1B).

Technical challenges and optimization efforts have accompanied PMA treatment throughout its development (Fittipaldi et al., 2012; Elizaquivel et al., 2014; Emerson et al., 2017; Codony et al., 2020). 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.

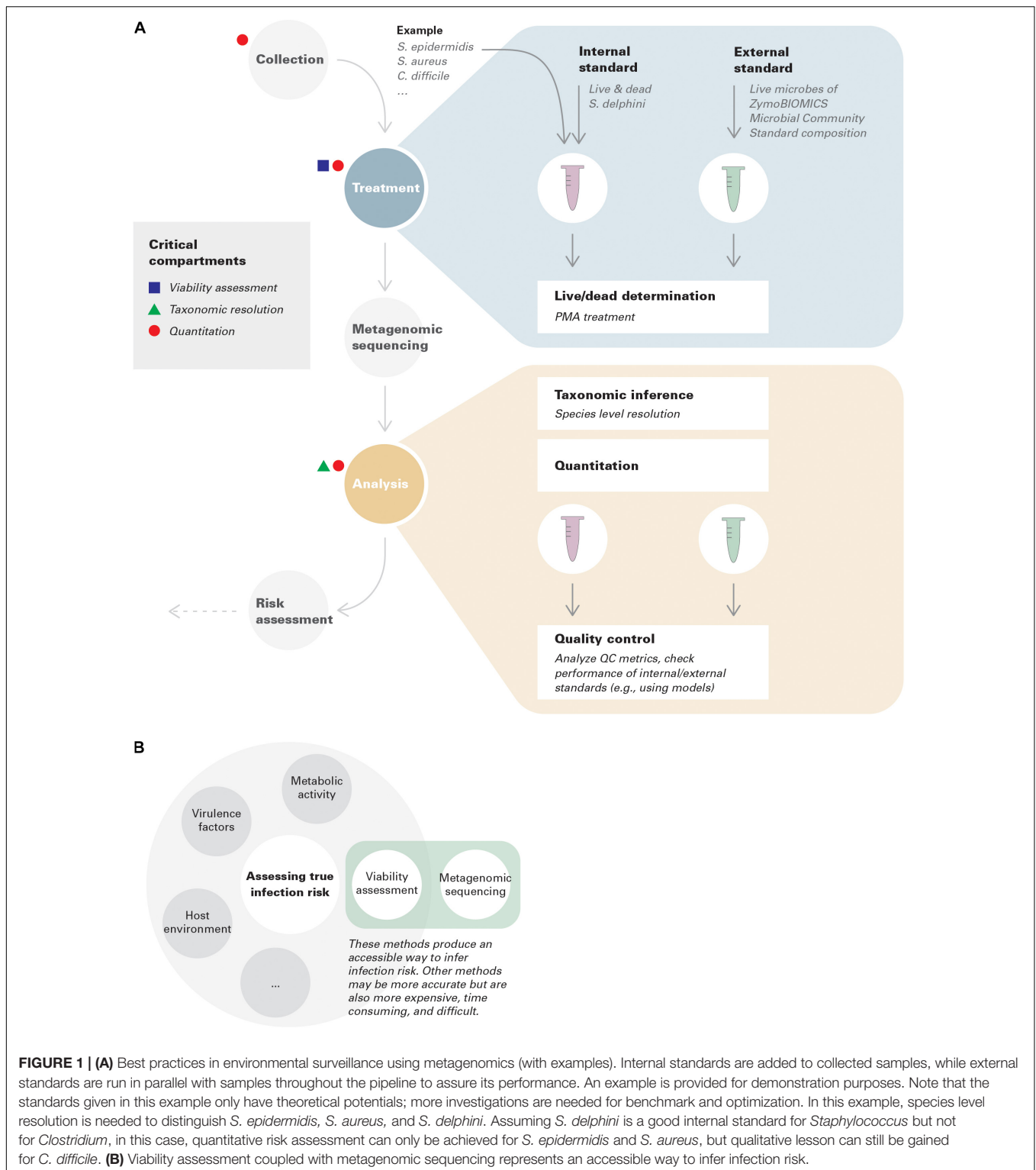
Application of PMA treatment to environmental samples has been hindered partly because these samples contain a diverse microbial community in a complex matrix. As pointed out by Wang et al., PMA-seq with a universal protocol is not feasible to quantify viability of realistic communities, even with *E. coli* controls spiked in Wang et al. (2020). 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 (Yang et al., 2012). Analogous principles have also been applied to the development of reagent enhancers by focusing on Gram-negative bacteria (Codony et al., 2020).

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* (Pacholewicz et al., 2019) 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 Fittipaldi et al. (2011) 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 toward implementation of PMA treatment in environmental surveillance.

INFERRING TAXONOMY IN LOW-BIOMASS METAGENOMES

Short-read shotgun metagenomic analysis reveals taxonomy without the limitations of amplicon sequencing or culture-based methods (Quince et al., 2017). 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. Afterward, 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 1; Jovel et al., 2016; Zaheer et al., 2018). Smaller read libraries are particularly challenged by a diminished capacity to detect rare taxa and accurately estimate overall taxonomic abundance of samples (Hillmann et al., 2018) because more abundant members in the metagenome have a higher likelihood of detection (Nayfach and Pollard, 2016). Additionally, the lowered overall information content of low coverage and depth read libraries impact the ability to identify low level taxonomic ranks.

Low-biomass samples are especially sensitive to the presence of contaminants, as the true signal can be easily overwhelmed (Eisenhofer et al., 2019). 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 (Tamariz et al., 2006; Silkie et al., 2008; Seitz et al., 2015; Minich et al., 2018). Metagenomic samples should be accompanied with kit extraction negative controls and DNA-spiked positive/internal controls during sequencing runs to identify sources of contamination (Minich et al., 2018; Eisenhofer et al., 2019; McLaren et al., 2019).

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 (Lusk, 2014; Lauder et al., 2016). The simplest approach is to remove sample reads that align to taxa found in the negative controls (Breitwieser et al., 2017). 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 (Davis et al., 2018). 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 (Schmieder and Edwards, 2011; Davis et al., 2018; Martí, 2019).

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 (Breitwieser et al., 2017; Ye et al., 2019). 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 (Tamames et al., 2019) 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 (Quince et al., 2017; Tamames et al., 2019). 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) (Truong et al., 2015; Brito and Alm, 2016; Roe et al., 2016; Hillmann et al., 2018). If strain-level variants are desired, merging paired-end reads or using sequencing technologies that generate longer reads may be necessary (Bruto and Alm, 2016).

Choosing an appropriate taxonomic reference database can greatly impact the breadth of taxa identified (Pereira-Marques et al., 2019; McArdle and Kafrou, 2020). 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 (Breitwieser et al., 2017). CAMISIM, a tool that simulates microbial metagenomic datasets, can be used by researchers to test different approaches (Fritz et al., 2019).

QUANTITATIVE METAGENOMICS IN ENVIRONMENTAL SURVEILLANCE

Conceptually, quantitative metagenomics has appeared in many ways in microbiome research, ranging from performing basic calculations of abundance, to normalizing metrics to these calculations, to the ultimate goal of absolute quantification as in qPCR (Pons et al., 2010).

TABLE 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	Huson et al., 2007; Rodriguez and Konstantinidis, 2014b
Read depth	The number of times a particular base is captured by a read	$base \times reads \text{ that map to base}$	Li et al., 2009; Quinlan and Hall, 2010
Mean read depth	The read depth averaged across the metagenome assembly	$\frac{reads \text{ mapped to assembly} \times average \text{ read length}}{assembly \text{ size}}$	Vezzulli et al., 2017

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 (Nayfach and Pollard, 2016; **Table 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 (Le Chatelier et al., 2013; Wen et al., 2017), to compare disease and health states (Qin et al., 2014), and to predict resistome (Ruppe et al., 2019), 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 catalog).

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 (Satinsky et al., 2013) and across taxa/genes (Frank and Sørensen, 2011; Nayfach and Pollard, 2016).

Technically, several challenges remain to be overcome toward 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 μ L 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 (Kerepesi and Grolmusz, 2016). 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 (Nayfach and Pollard, 2016) 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 (McLaren et al., 2019), adding genomic DNA just prior to cell lysis in DNA isolation (Satinsky et al., 2013, 2014), and including a set of synthetic DNA before library preparation (Blackburn et al., 2019). 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 (Danzer, 2007). 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 (McLaren et al., 2019).

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

TABLE 2 | Parameters to profile microbial communities summarized by Nayfach and Pollard.

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

Theoretical calculation equations are created based on our interpretation of the authors' descriptions, which are also adapted to the context of environmental surveillance.

clinical standards for antimicrobial susceptibility testing (CLSI, 2018), 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 (Ballenghien et al., 2017; Pereira-Marques et al., 2019). 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 (Rodriguez and Konstantinidis, 2014a). 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.

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*.

REFERENCES

- Ballenghien, M., Faivre, N., and Galtier, N. (2017). Patterns of cross-contamination in a multispecies population genomic project: detection, quantification, impact, and solutions. *BMC Biol.* 15:25. doi: 10.1186/s12915-017-0366-6
- Blackburn, J., Wong, T., Madala, B. S., Barker, C., Hardwick, S. A., Reis, A. L. M., et al. (2019). Use of synthetic DNA spike-in controls (sequins) for human genome sequencing. *Nat. Protoc.* 14, 2119–2151. doi: 10.1038/s41596-019-0175-1

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, cannot only help produce internally consistent results but also increase external compatibility with data generated in different studies (Nayfach and Pollard, 2016) or with different protocols (Sinha et al., 2017).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

JS managed the project. JS and AM co-wrote the manuscript. EH, VY, and MH provided editorial input. All authors have read, edited, and approved of the final manuscript.

FUNDING

This research was supported by the Searle Leadership Fund.

ACKNOWLEDGMENTS

We would like to thank Jason Romano for his help in making the figure.

- Breitwieser, F. P., Lu, J., and Salzberg, S. L. (2017). A review of methods and databases for metagenomic classification and assembly. *Brief. Bioinform.* 20, 1125–1136. doi: 10.1093/bib/bbx120
- Brito, I. L., and Alm, E. J. (2016). Tracking strains in the microbiome: insights from metagenomics and models. *Front. Microbiol.* 7:712. doi: 10.3389/fmicb.2016.00712
- Chau, C. K., Tu, E. Y., Chan, D. W., and Burnett, J. (2002). Estimating the total exposure to air pollutants for different population age groups in Hong Kong. *Environ. Int.* 27, 617–630. doi: 10.1016/s0160-4120(01)00120-9

- Cincinelli, A., and Martellini, T. (2017). Indoor air quality and health. *Int. J. Environ. Res. Public Health* 14:1286. doi: 10.3390/ijerph14111286
- CLSI (2018). *CLSI Supplement M100- Performance Standards for Antimicrobial Susceptibility Testing*, 28th Edn. Wayne, PA: Clinical and Laboratory Standards Institute.
- Codony, F., Dinh-Thanh, M., and Agustí, G. (2020). Key factors for removing bias in viability PCR-based methods: a review. *Curr. Microbiol.* 77, 682–687. doi: 10.1007/s00284-019-01829-y
- Danzer, K. (2007). “Alternative calibration procedures”, in *Analytical Chemistry: Theoretical and Metrological Fundamentals*, (Heidelberg: Springer-Verlag).
- Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., and Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 6:226. doi: 10.1186/s40168-018-0605-2
- Deurenberg, R. H., Bathoorn, E., Chlebowicz, M. A., Couto, N., Ferdous, M., Garcia-Cobos, S., et al. (2017). Application of next generation sequencing in clinical microbiology and infection prevention. *J. Biotechnol.* 243, 16–24. doi: 10.1016/j.jbiotec.2016.12.022
- Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., and Weyrich, L. S. (2019). Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends Microbiol.* 27, 105–117. doi: 10.1016/j.tim.2018.11.003
- Elizaquivel, P., Aznar, R., and Sanchez, G. (2014). Recent developments in the use of viability dyes and quantitative PCR in the food microbiology field. *J. Appl. Microbiol.* 116, 1–13. doi: 10.1111/jam.12365
- Emerson, J. B., Adams, R. I., Román, C. M. B., Brooks, B., Coil, D. A., Dahlhausen, K., et al. (2017). Schrödinger’s microbes: tools for distinguishing the living from the dead in microbial ecosystems. *Microbiome* 5:86. doi: 10.1186/s40168-017-0285-3
- Fittipaldi, M., Codony, F., Adrados, B., Camper, A. K., and Morató, J. (2011). Viable real-time PCR in environmental samples: can all data be interpreted directly? *Microb. Ecol.* 61, 7–12. doi: 10.1007/s00248-010-9719-1
- Fittipaldi, M., Nocker, A., and Codony, F. (2012). Progress in understanding preferential detection of live cells using viability dyes in combination with DNA amplification. *J. Microbiol. Methods* 91, 276–289. doi: 10.1016/j.mimet.2012.08.007
- Frank, J. A., and Sørensen, S. J. (2011). Quantitative metagenomic analyses based on average genome size normalization. *Appl. Environ. Microbiol.* 77, 2513–2521. doi: 10.1128/aem.02167-10
- Fritz, A., Hofmann, P., Majda, S., Dahms, E., Dröge, J., Fiedler, J., et al. (2019). CAMISIM: simulating metagenomes and microbial communities. *Microbiome* 7:17. doi: 10.1186/s40168-019-0633-6
- Gomez-Silvan, C., Leung, M. H. Y., Grue, K. A., Kaur, R., Tong, X., Lee, P. K. H., et al. (2018). A comparison of methods used to unveil the genetic and metabolic pool in the built environment. *Microbiome* 6:71. doi: 10.1186/s40168-018-0453-0
- Hillmann, B., Al-Ghalith, G. A., Shields-Cutler, R. R., Zhu, Q., Gohl, D. M., Beckman, K. B., et al. (2018). Evaluating the information content of shallow shotgun metagenomics. *mSystems* 3:e00069-18. doi: 10.1128/mSystems.00069-18
- Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386. doi: 10.1101/gr.5969107
- Jovel, J., Patterson, J., Wang, W., Hotte, N., O’Keefe, S., Mitchel, T., et al. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. *Front. Microbiol.* 7:459. doi: 10.3389/fmicb.2016.00459
- Kerepesi, C., and Grolmsuz, V. (2016). Evaluating the quantitative capabilities of metagenomic analysis software. *Curr. Microbiol.* 72, 612–616. doi: 10.1007/s00284-016-0991-2
- Lauder, A. P., Roche, A. M., Sherrill-Mix, S., Bailey, A., Laughlin, A. L., Bittinger, K., et al. (2016). Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome* 4:29. doi: 10.1186/s40168-016-0172-3
- Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., et al. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature* 500, 541–546. doi: 10.1038/nature12506
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Lusk, R. W. (2014). Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLoS One* 9:e110808. doi: 10.1371/journal.pone.0110808
- Martí, J. M. (2019). Recentrifuge: robust comparative analysis and contamination removal for metagenomics. *PLoS Comput. Biol.* 15:e1006967. doi: 10.1371/journal.pcbi.1006967
- McArdle, A. J., and Kaforou, M. (2020). Sensitivity of shotgun metagenomics to host DNA: abundance estimates depend on bioinformatic tools and contamination is the main issue. *Access Microbiol.* 2:acmi000104. doi: 10.1099/acmi.0.000104
- McLaren, M. R., Willis, A. D., and Callahan, B. J. (2019). Consistent and correctable bias in metagenomic sequencing experiments. *eLife* 8:e46923. doi: 10.7554/eLife.46923
- Minich, J. J., Zhu, Q., Janssen, S., Hendrickson, R., Amir, A., Vetter, R., et al. (2018). KatharoSeq enables high-throughput microbiome analysis from low-biomass samples. *mSystems* 3:e00218-17. doi: 10.1128/mSystems.00218-17
- Nayfach, S., and Pollard, K. S. (2016). Toward accurate and quantitative comparative metagenomics. *Cell* 166, 1103–1116. doi: 10.1016/j.cell.2016.08.007
- Pacholewicz, E., Buhler, C., Wulsten, I. F., Kraushaar, B., Luu, H. Q., Iwobi, A. N., et al. (2019). Internal sample process control improves cultivation-independent quantification of thermotolerant *Campylobacter*. *Food Microbiol.* 78, 53–61. doi: 10.1016/j.fm.2018.09.017
- Pereira-Marques, J., Hout, A., Ferreira, R. M., Weber, M., Pinto-Ribeiro, I., van Doorn, L.-J., et al. (2019). Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis. *Front. Microbiol.* 10:1277. doi: 10.3389/fmicb.2019.01277
- Pons, N., Batto, J. M., Kennedy, S., Almeida, M., Boumezbeur, F., Moumen, B., et al. (2010). METEOR, a platform for quantitative metagenomic profiling of complex ecosystems. *Journées Ouvertes Biol. Inform. Math.*
- Popovich, K. J., Lyles, R., Hayes, R., Hota, B., Trick, W., Weinstein, R. A., et al. (2012). 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.* 33, 889–896. doi: 10.1086/667371
- Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., et al. (2014). Alterations of the human gut microbiome in liver cirrhosis. *Nature* 513, 59–64. doi: 10.1038/nature13568
- Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35, 833–844. doi: 10.1038/nbt.3935
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Rodriguez, R. L., and Konstantinidis, K. T. (2014a). Estimating coverage in metagenomic data sets and why it matters. *ISME J.* 8, 2349–2351. doi: 10.1038/ismej.2014.76
- Rodriguez, R. L., and Konstantinidis, K. T. (2014b). Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. *Bioinformatics* 30, 629–635. doi: 10.1093/bioinformatics/btt584
- Roe, C. C., Horn, K. S., Driebe, E. M., Bowers, J., Terriquez, J. A., Keim, P., et al. (2016). 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* 153:11. doi: 10.1186/s41065-016-0017-x
- Ruppe, E., Ghoulane, A., Tap, J., Pons, N., Alvarez, A. S., Maziers, N., et al. (2019). Prediction of the intestinal resistome by a three-dimensional structure-based method. *Nat. Microbiol.* 4, 112–123. doi: 10.1038/s41564-018-0292-6
- Satinsky, B. M., Gifford, S. M., Crump, B. C., and Moran, M. A. (2013). “Chapter twelve - use of internal standards for quantitative metatranscriptome and metagenome analysis,” in *Methods in Enzymology*, ed. E. F. DeLong (Burlington, MA: Academic Press), 237–250.
- Satinsky, B. M., Zielinski, B. L., Doherty, M., Smith, C. B., Sharma, S., Paul, J. H., et al. (2014). The Amazon continuum dataset: quantitative metagenomic and metatranscriptomic inventories of the Amazon River plume. *Microbiome* 2:17. doi: 10.1186/2049-2618-2-17

- Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 6:e17288. doi: 10.1371/journal.pone.0017288
- Seitz, V., Schaper, S., Dröge, A., Lenze, D., Hummel, M., and Hennig, S. (2015). A new method to prevent carry-over contaminations in two-step PCR NGS library preparations. *Nucleic Acids Res.* 43:e135. doi: 10.1093/nar/gkv694
- Selway, C. A., Eisenhofer, R., and Weyrich, L. S. (2020). Microbiome applications for pathology: challenges of low microbial biomass samples during diagnostic testing. *J. Pathol. Clin. Res.* 6, 97–106. doi: 10.1002/cjp2.151
- Silkie, S. S., Tolcher, M. P., and Nelson, K. L. (2008). Reagent decontamination to eliminate false-positives in *Escherichia coli* qPCR. *J. Microbiol. Methods* 72, 275–282. doi: 10.1016/j.mimet.2007.12.011
- Sinha, R., Abu-Ali, G., Vogtmann, E., Fodor, A. A., Ren, B., Amir, A., et al. (2017). Assessment of variation in microbial community amplicon sequencing by the microbiome quality control (MBQC) project consortium. *Nat. Biotechnol.* 35, 1077–1086. doi: 10.1038/nbt.3981
- Smith, J. D., Mitsakou, C., Kitwiroon, N., Barratt, B. M., Walton, H. A., Taylor, J. G., et al. (2016). London hybrid exposure model: improving human exposure estimates to NO₂ and PM_{2.5} in an Urban setting. *Environ. Sci. Technol.* 50, 11760–11768. doi: 10.1021/acs.est.6b01817
- Stephens, B., Azimi, P., Thoemmes, M. S., Heidarinejad, M., Allen, J. G., and Gilbert, J. A. (2019). Microbial exchange via fomites and implications for human health. *Curr. Pollut. Rep.* 5, 198–213.
- Tamames, J., Cobo-Simón, M., and Puente-Sánchez, F. (2019). Assessing the performance of different approaches for functional and taxonomic annotation of metagenomes. *BMC Genomics* 20:960. doi: 10.1186/s12864-019-6289-6
- Tamariz, J., Voynarovska, K., Prinz, M., and Caragine, T. (2006). The application of ultraviolet irradiation to exogenous sources of DNA in plasticware and water for the amplification of low copy number DNA. *J. Forensic Sci.* 51, 790–794. doi: 10.1111/j.1556-4029.2006.00172.x
- Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., et al. (2015). MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* 12, 902–903. doi: 10.1038/nmeth.3589
- United Nations (2018). Available online at: <https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html> (accessed June 8, 2020).
- Valdez, M. K., Sexton, J. D., Lutz, E. A., and Reynolds, K. A. (2015). Spread of infectious microbes during emergency medical response. *Am. J. Infect. Control* 43, 606–611. doi: 10.1016/j.ajic.2015.02.025
- Vezzulli, L., Grande, C., Tassistro, G., Brettar, I., Höfle, M. G., Pereira, R. P. A., et al. (2017). Whole-genome enrichment provides deep insights into *Vibrio cholerae* metagenome from an African River. *Microb. Ecol.* 73, 734–738. doi: 10.1007/s00248-016-0902-x
- Wang, Y., Yan, Y., Thompson, K. N., Bae, S., Accorsi, E. K., Zhang, Y., et al. (2020). Whole microbial community viability is not quantitatively reflected by propidium monoazide sequencing approach. *Microbiome* 9:17. doi: 10.21203/rs.3.rs-59563/v1
- Wen, C., Zheng, Z., Shao, T., Liu, L., Xie, Z., Le Chatelier, E., et al. (2017). Quantitative metagenomics reveals unique gut microbiome biomarkers in ankylosing spondylitis. *Genome Biol.* 18:142. doi: 10.1186/s13059-017-1271-6
- Xiao, S., Li, Y., Wong, T. W., and Hui, D. S. C. (2017). Role of fomites in SARS transmission during the largest hospital outbreak in Hong Kong. *PLoS One* 12:e0181558. doi: 10.1371/journal.pone.0181558
- Yang, Y., Wan, C., Xu, H., Lai, W., Xiong, Y., Xu, F., et al. (2012). 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* 28, 456–462. doi: 10.1016/j.foodcont.2012.05.061
- Ye, S. H., Siddle, K. J., Park, D. J., and Sabeti, P. C. (2019). Benchmarking metagenomics tools for taxonomic classification. *Cell* 178, 779–794. doi: 10.1016/j.cell.2019.07.010
- Zaheer, R., Noyes, N., Ortega Polo, R., Cook, S. R., Marinier, E., Van Domselaar, G., et al. (2018). Impact of sequencing depth on the characterization of the microbiome and resistome. *Sci. Rep.* 8:5890. doi: 10.1038/s41598-018-24280-8
- Zhong, Z.-P., Solonenko, N. E., Gazitúa, M. C., Kenny, D. V., Mosley-Thompson, E., Rich, V. I., et al. (2018). Clean low-biomass procedures and their application to ancient ice core microorganisms. *Front. Microbiol.* 9:1094. doi: 10.3389/fmicb.2018.01094

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Shen, McFarland, Young, Hayden and Hartmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.