Journal of Biomolecular Techniques • Volume 34(3); 2023 Sep

Analysis of Microbiomes from Ultra-Low Biomass Surfaces Using Novel Surface Sampling and Nanopore Sequencing

Anna C. Simpson[1] Scott Tighe[2] Season Wong[3] Patrick Leo[1] Ceth Parker[1] Atul M. Chander[1] Michael Williams[1] Hao-Wei Wu[3] Kasthuri Venkateswaran[1] Nitin K. Singh[1]

[1]Jet Propulsion Laboratory, California Institute of Technology, Biotechnology and Planetary Protection Group, Pasadena, California 91109, USA,
[2]Vermont Integrative Genomics, University of Vermont, Burlington, Vermont,
[3]AI Biosciences, Inc., College Station, Texas

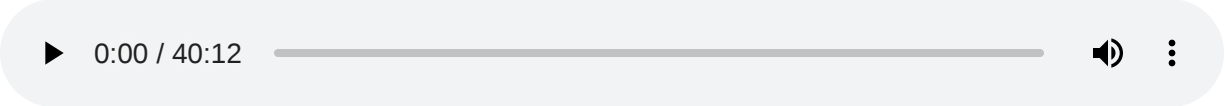
Association of Biomolecular Resource Facilities

Published on: Aug 10, 2023

DOI: https://doi.org/10.7171/3fc1f5fe.bac4a5b3

ABSTRACT

The rapid assessment of microbiomes from ultra-low biomass environments such as cleanrooms or hospital operating rooms has a number of applications for human health and spacecraft manufacturing. Current techniques often employ lengthy protocols using short-read DNA sequencing technology to analyze amplified DNA and have the disadvantage of a longer analysis time and lack of portability. Here, we demonstrate a rapid (~24 hours) on-site nanopore-based sequencing approach to characterize the microbiome of a NASA Class 100K cleanroom where spacecraft components are assembled. This approach employs a modified protocol of Oxford Nanopore's Rapid PCR Barcoding Kit in combination with the recently developed Squeegee-Aspirator for Large Sampling Area (SALSA) surface sampling device. Results for these ultra-low biomass samples revealed DNA amplification ~1 to 2 orders of magnitude above process control samples and were dominated primarily by *Paracoccus* and *Acinetobacter* species. Negative control samples were collected to provide critical data on background contamination, including *Cutibacterium acnes*, which most likely originated from the sampling reagents-associated microbiome (kitome). Overall, these results provide data on a novel approach for rapid low-biomass DNA profiling using the SALSA sampler combined with modified nanopore sequencing. These data highlight the critical need for employing multiple negative controls, along with using DNA-free reagents and techniques, to enable a proper assessment of ultra-low biomass samples.



Listen to this article

ADDRESS CORRESPONDENCE TO: Anna C. Simpson, 4800 Oak Grove Drive, Jet Propulsion Laboratory, California Institute of Technology, Biotechnology and Planetary Protection Group, Pasadena, California 91109, USA (Phone: (818) 354-5799; E-mail: anna.c.simpson@jpl.nasa.gov).

ADDRESS CORRESPONDENCE TO: Nitin K. Singh, 4800 Oak Grove Drive, Jet Propulsion Laboratory, California Institute of Technology, Biotechnology and Planetary Protection Group, Pasadena, California 91109, USA (Phone: (818) 354-5799; E-mail: nitin.k.singh@jpl.nasa.gov).

Keywords: low-biomass DNA, microbiome surface sampling, nanopore sequencing, spacecraft assembly facility, rapid PCR barcoding

INTRODUCTION

Long-read nanopore sequencing has become a well-established technology for routine laboratory and field-based DNA sequencing since its commercial release in 2014.[1],[2],[3] Significant improvements to this technology, including refined portability and size, improved flow cells, reagents, and base-calling algorithms,

have resulted in vastly improved sequence quality, making nanopore a choice tool for shotgun metagenomic sequencing. However, despite these improvements, many of the methods for nanopore library synthesis require DNA concentrations that often disqualify the analysis of low-biomass samples.[4] While some protocols such as rapid PCR barcoding allow for lower inputs (1-5 ng), successful PCR is not always realized. Inhibitors, low molecular weight DNA, sequence context, and PCR reagents all play a role in generating good quality reads.[5]

Low-biomass sampling for metagenome sequencing requires a high-efficiency collection method followed by a concentration protocol in order to achieve a higher concentration of the analyte to detect the target microbiome. Due to the lack of availability of DNA-free reagents, the steps of sample collection, amplification, and next-generation sequencing library preparation must be combined with multiple negative controls and reagent blanks at various processing points to discern signal from noise. Sampling intact microbes with swabs, wipes, tape strips, and other sampling devices such as the Biological Sampling Kit has a maximum recovery efficiency of ~10 to 50% and is usually far lower, with the recovery of environmental DNA (eDNA) even less efficient.[6],[7],[8],[9],[10],[11] The elution volume used for releasing bound cells and DNA from swabs and wipes is a trade-off between recovery efficiency (supported by higher volumes) and the concentration of the sample for downstream applications. Methods for concentrating and collecting low-biomass samples are well established, including liquid filtering of a diluted sample (filter paper, high-volume hollow fiber such as InnovaPrep CP, etc), SpeedVac concentration, or using novel or disruptive technologies such as gradient flotation or magnetic capture techniques,[12],[13],[14] and additional PCR cycles can be used to increase DNA concentrations after extraction. Nonetheless, using any of these methods remains challenging, as increased sample manipulation will result in subsequent unavoidable secondary contamination of a sample.[15] This is especially exaggerated when sampling very clean surfaces such as those in cleanrooms, hospital operating rooms, and spacecraft-associated components. DNA recovered from these samples are difficult to prescribe to the sample itself due to the unavoidable presence of microbial contamination associated with DNA extraction and library prep kits (“kitomes”)[16],[17],[18] and require multiple control samples from each stage of sample preparation to be sequenced along with true samples.

Oxford Nanopore has developed several PCR amplification–based library preparation methods including 16s ribosomal RNA (rRNA) and rapid PCR barcoding for low DNA input (minimum 1 ng); however, these methods still require modification for ultra-low input (<10 pg/sample). Specialized computational processing is also critical; nanopore sequencing produces a certain number of high-quality “noise reads” that pass quality and length filters but do not map to controlled input genomes or cellular life in general and may be artifacts from a subset of flow cell pores.[19] These anomalies are exacerbated by the incomplete or contradictory classification databases used for taxonomic assignment of shotgun metagenomic reads, as highlighted in a recent study at the Jet Propulsion Laboratory Spacecraft Assembly Facility (JPL-SAF), indicating that only ~15% of reads could be assigned to the species level.[20]

Low-level microbiome detection is an emerging technique for many technology groups including NASA. NASA's Planetary Protection group has pioneered a range of microbiome methodologies,[21],[22] including a recent effort towards rapid shotgun metagenomic sequencing of spacecraft and associated surfaces with accurate functional and species-level annotation, which can be used for probabilistic risk assessment–based models for different missions.[23],[24] Long-read sequencing platforms such as Nanopore have the potential to provide this data if modifications to sample collection, DNA extraction, and library preparation are successful. Multiple studies have shown success in sequencing low-biomass samples of mock communities or single genomes[19],[25],[26] using kits with long prep time. Thus far, there are no published attempts to provide rapid nanopore-based shotgun sequencing of an ultra-low biomass-built environment, although several in-house JPL studies by Tighe and Venkateswaran (unpublished) have confirmed successful sequencing of as little as 200 pg of input DNA using a Nanopore Rapid DNA kit (RAD004) by adding nonspecific carrier DNA.

In this study, we used a new handheld collection instrument called Squeegee-Aspirator for Large Sampling Area (SALSA) in combination with a concentration step and long-read amplification using a modification of Oxford Nanopore's Rapid PCR Barcoding kit followed by nanopore sequencing (~9 hours of sample to sequencing time) to provide species-level data within 24 hours of collection. The SALSA device combines squeegee action and aspiration of liquid from surfaces into a collection tube, completely bypassing the problem of cell and DNA adsorption to swab fibers. Current recovery efficiency of the SALSA device (60% or higher depending on surface) is an increased efficiency for most sample types compared to swabs (10%) (Centers for Disease Control and Prevention report, S. Wang communication).[27] Here, we demonstrate that rapid turnaround time shotgun metagenomic sequencing is possible for ultra-low biomass environments using current nanopore technology and discuss future modifications to increase accuracy and lower contamination levels.

METHODS

Description of sampling device

The SALSA microbiome sample collector is a patent-pending, handheld, battery-operated device that uses a vacuum and squeegee function to sample a large surface area (~1 m^2) that is prewet with DNA-free water or buffer commonly used in surface sampling. Figure 1 shows the SALSA in action sampling a stainless steel 4 x 4 in area. The battery aspirator is the only reusable unit, whereas the manifolds, 5-mL collection tubes, and the squeegee heads (made from off-the-shelf duck-bill valves) are disposable consumables. SALSA components for this study were 3D printed in a biohood and decontaminated with 70% ethanol and UV exposure; the wetting buffer currently used is sterile PCR-grade water, preloaded in a UV-treated spray bottle, that is deployed along with the SALSA unit. In the future, these components could be certified as DNA free and sterile. The advantage of the SALSA technique is its "scrub and sample" approach in which the sampling solution is transferred directly into a sterile 5-mL microcentrifuge collection tube, making it readily available

for culture or molecular methods without the need for an elution step from swabs, sponges, or wipes.[6],[7],[9], [28],[29] The battery-operated device is USB rechargeable and has 3 suction levels.

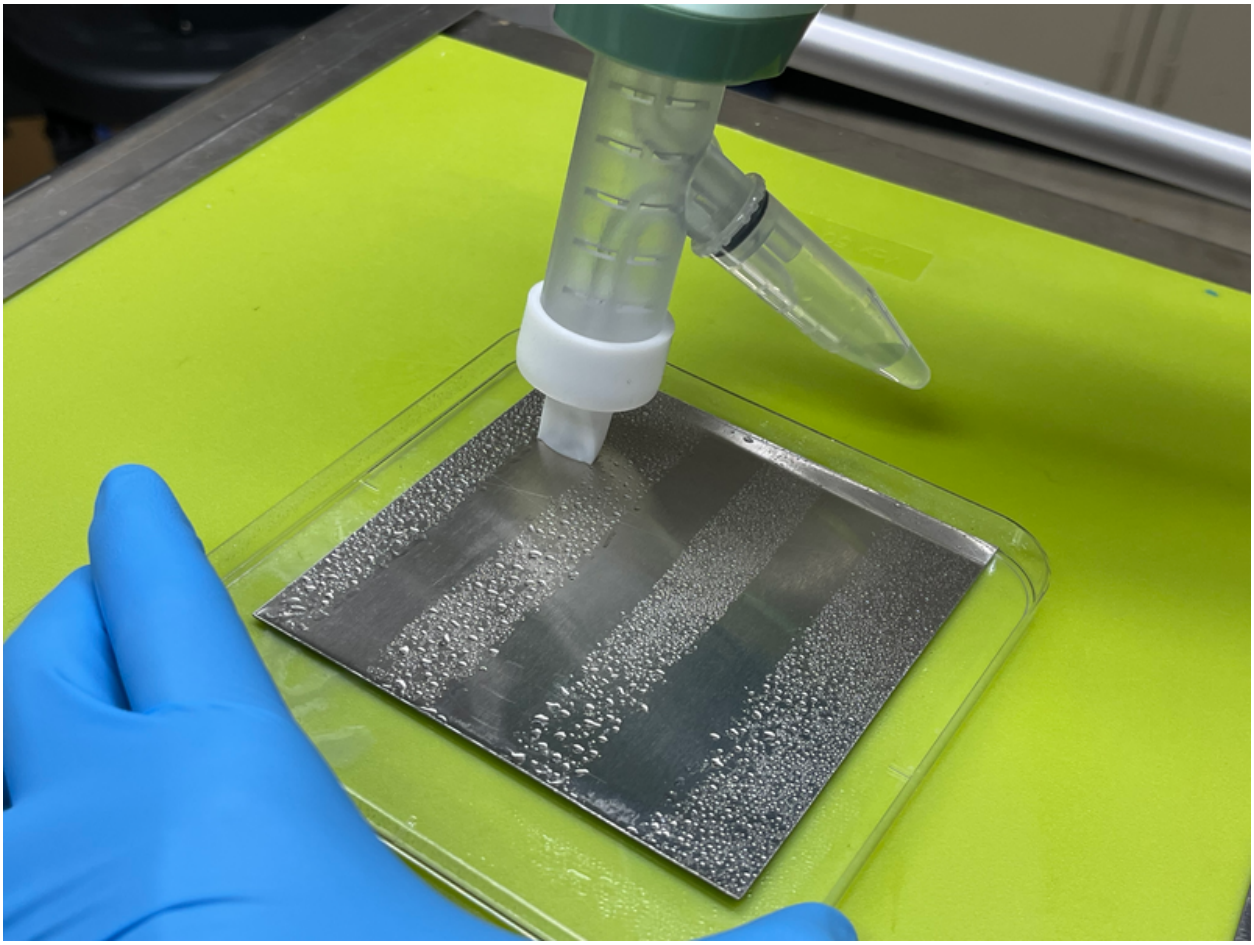


Figure 1
Demonstration of the SALSA recovering liquid in 1 pass as demonstrated by moving the device across a 4 x 4" stainless steel plate wetted by the spray vial.

SALSA collection efficiency

The collection efficiency of the SALSA device was compared to that of the JPL swab assay in April 2022. Ten adjacent 100-cm^2 areas on the JPL-SAF (a Class 100K cleanroom facility) floor and 10 similar areas on the floor of a JPL microbiology laboratory were sampled, 5 at each location with the SALSA instrument and the other 5 at each location using swabs. For swabbed samples, the NASA standard swab assay was performed as established previously[30] with one alteration: 1 swab instead of 4 was used to sample the 10 x 10 cm area in order to reach swab saturation. Suitable aliquots (100 µL) of sample from the SALSA device or from the swab elution were extracted for DNA using a Maxwell RSC device (Promega, Madison, Wisconsin) with a Maxwell RSC Cell kit and with elution performed using 50 µL of 10-mM Tris buffer.[13] Quantification via 16S rRNA

qPCR was performed using a Femto Bacterial DNA Quantification Kit (Zymo) on a QuantStudio 6 Pro Real-Time PCR System (Applied Biosystems).[13]

Sample collection and concentration

Cleanroom floor samples were collected on August 10, 2022, from the JPL-SAF using a SALSA device. Three locations were sampled: inside the cleanroom near the entrance (ECR), inside the cleanroom far from the entrance (FCR), and inside the anteroom between the cleanroom and the air shower (AR). Three replicate floor areas approximately 12 x 12” were selected at each location, and sterile PCR-grade water (4 pumps/2 mL) was sprayed on each target sampling area. Using a new sterile collection tip for each sample, the SALSA aspirator was used over the entire target area, ensuring complete sample collection and deposition into the attached microcentrifuge tube. Three “process control” samples were collected from the sprayer by aspirating 2-mL aliquots of the sprayer water using 3 separate sets of SALSA collection heads and collection tubes without active sampling (CON). A negative control sample consisting of sterile water from the same container as samples and process controls was also collected in the laboratory.

Samples and controls were immediately concentrated on an InnovaPrep CP-150 device (InnovaPrep LLC, Drexel, Missouri) with a 0.2-µm polysulfone hollow fiber concentrating pipette tip (InnovaPrep LLC, SKU CC08022-10) using a preset elution volume of 150 µL of the InnovaPrep CP phosphate-buffered saline (PBS). [13] Aliquots of 100 µL were transferred to a 0.2-mL PCR tube and incubated for 1 hour at 35 °C with 5 µL of DNA-free metapolyzyme (MAC4L-DF Sigma-Aldrich, St. Louis, Missouri) to degrade microbial cell walls and allow for a more effective lysis. Complete sample lysis and DNA extraction/purification was accomplished using a Maxwell RSC device (Promega, Madison, Wisconsin) with a Maxwell RSC Cell kit and eluted in 35 µL of 10-mM Tris buffer. DNA quantification was performed using a Qubit DNA HS kit (Cat Q32851, ThermoFisher, Waltham, Massachussetts) and by qPCR using the Femto Bacterial DNA Quantification kit (Cat E2006, Zymo Corp. Irvine, California) on a QuantStudio™ 6 Pro Real-Time PCR System (Applied Biosystems, Foster City, California).

Culturing

Non-quantitative culturing was performed on a subaliquot of the same samples by inoculation of R2A agar (BD Difco, Cat# 218263) using the standard spread plate technique and incubated for 7 days at 22 °C. The resulting colonies were sub-cultured for purification on R2A and taxonomically identified using either amplicon Sanger sequencing (16S rRNA or Internal Transcriber Space (ITS) based;[31] Azenta Life Sciences, Southfield, New Jersey; see ref. [32] for detailed methods) or whole genome Rapid Nanopore sequencing (RAD004) for failed amplicon Sanger results.

Nanopore library prep and sequencing

Libraries for nanopore shotgun metagenomic sequencing were prepared for all samples including negative controls using the Rapid PCR Barcoding kit (SQK-RPB004, Oxford Nanopore, Oxford United Kingdom), which employs transposase-mediated tagging of genomic DNA followed by simultaneous barcoding and PCR amplification. The following modifications were made to the original protocol to account for low-biomass samples: Modification 1: 7 µL of undiluted purified DNA and 2 µL of fragmentation mix (increased from protocol's 3 µL DNA and 1 µL of fragmentation mix) were adjusted to 20 µL with PCR-grade H_2O and used as inputs to the LongAmp Taq PCR reactions (NEB LongAmp Taq #M0323S, Ipswich, Massachussetts). Modification 2: The number of PCR cycles was increased from 14 to 18. Library cleanup was performed using Agencourt AMPure XP beads (Beckman Coulter Corp., Brea, California) at the standard 0.6X ratio as stated in the manufacturer's protocol, with a final elution of 11 µL of 10-mM Tris. Concentrations of PCR product were determined using the Qubit DNA HS reagent kit. Modification 3: Rather than combining 0.5 to 2 µL from each barcoded sample for a final input volume of 10 µL, all 12 barcoded sample volumes were pooled for a total volume of 120 µL and concentrated a second time using 72 µL of AMPure bead solution (0.6X ratio), cleaning with 80% ethanol, and eluting with 12 µL of 10-mM Tris (pH 8). This final 12-µL concentrate of pooled sample was prepared into a standard nanopore sequencing mix by combining it with 34 µl Sequencing Buffer, 25.5 µL loading beads, and 3.5 µL nuclease-free water (see SQK-RPB004, Oxford Nanopore). Sequencing was performed using a R9.4 flow cell for 24 hours using an Oxford Nanopore MinION MK1C sequencer equipped with MinKNOW ver. 5.1.8 software without adaptive sequencing, reserve pores, or barcode balancing option. Mux scans were set to 1-hour intervals. Although we allowed sequencing to proceed for 24 hours, the presence of any target genes or species can generally be detected within 1 to 6 hours.[33]

Sequence quality control, annotation, and statistics

Demultiplexed sequences that passed MinKNOW quality filtering were trimmed of barcodes/adapters using CLC Genomics Workbench (Qiagen) and annotated using multiple pipelines, including Kraken2/Braken,[34] GTDB-tk,[35] and DIAMOND/Megan6 (with minScore = 50, maxExpected = 0.01, topPercent = 10, and minSupportPercent = 0.01)[36],[37] using the National Center for Biotechnology Information (NCBI) nucleotide, nonredundant protein, and RefSeq databases.[38],[39] DIAMOND frameshift and range curling modes were used to adjust for long-read sequences. Only reads with a taxonomic assignment of 90% identity or greater to the NCBI nr database were considered to be annotated. Sequences were also annotated for functional traits using the EggNOG, KEGG, and SEED databases.[40],[41],[42] All statistical analysis and figure generation was carried out in R version 4.1.1 ("Kick Things") using the vegan and tidyverse packages.[43],[44]

RESULTS

Comparison of SALSA device versus swab method for sample collection

An evaluation of sample collection using the SALSA device versus modified single-swab NASA standard swab collection on 100-cm^2 replicate areas of the JPL-SAF floor and standard laboratory floor in April 2022 showed that the SALSA instrument collected significantly more biomass and debris compared to swab collection. Visual inspection of the samples showed considerably darker, higher-biomass samples collected using the SALSA device with the same volume of elution fluid compared to samples collected with swabs (Figure 2). Quantitative analysis using 16S qPCR revealed that higher amounts of genetic material were recovered using the SALSA device compared to swabs for both cleanroom and laboratory floors (Figure 3).

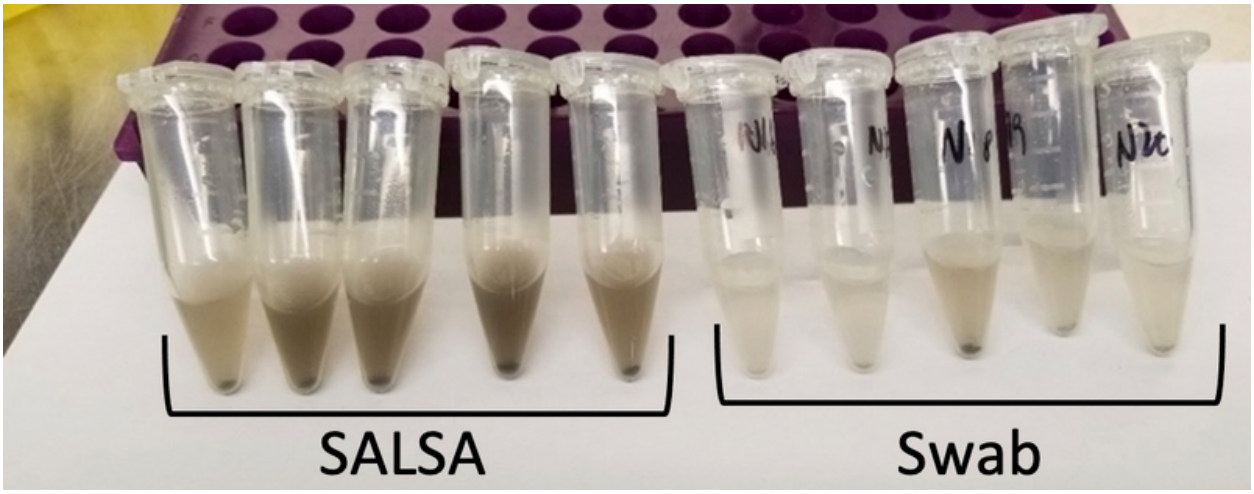


Figure 2
Visual demonstration of samples collected from the same area of a laboratory floor using SALSA or swabs.

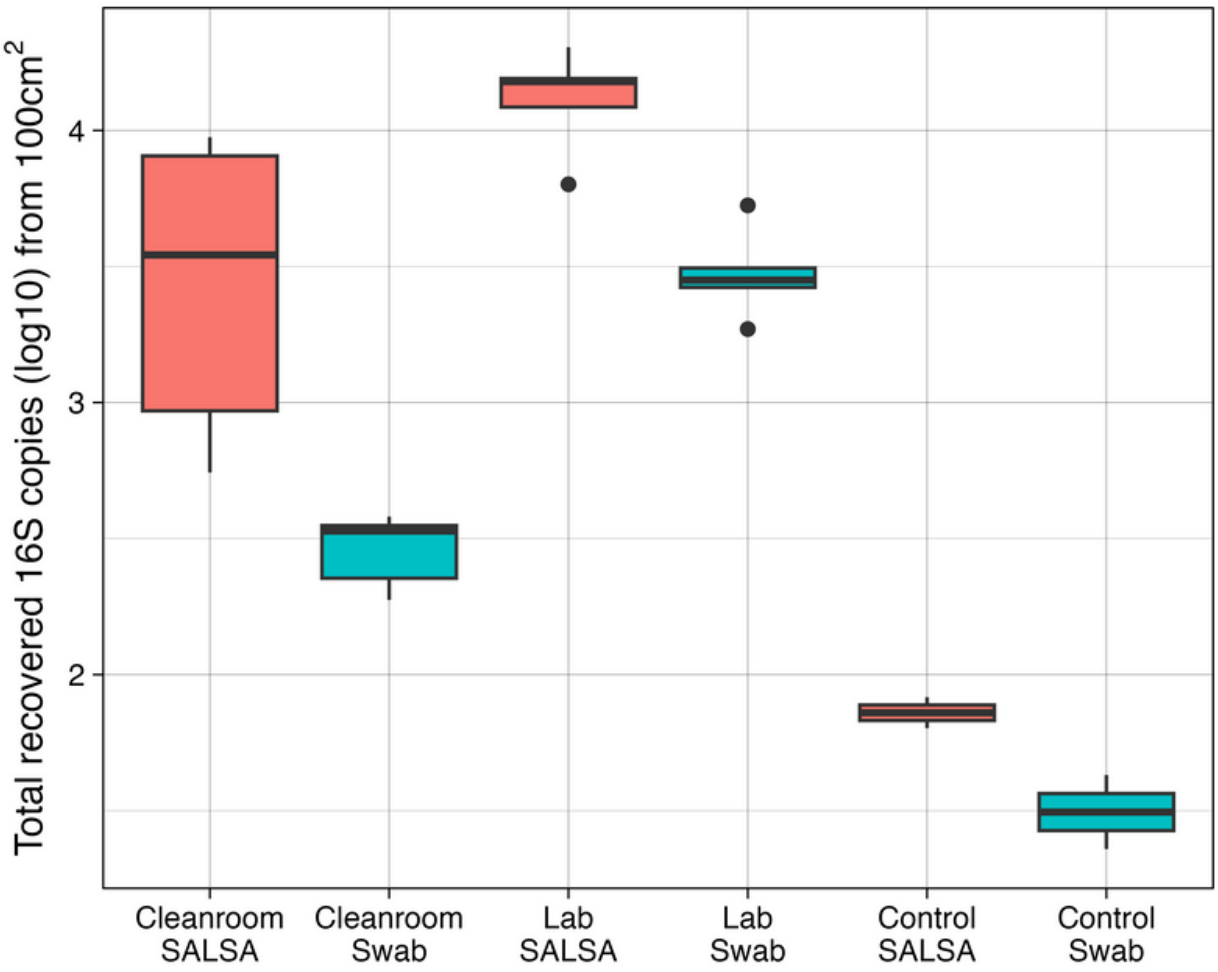


Figure 3
Comparison of total 16S rRNA gene yield from 100-cm2 areas of the JPL-SAF floor (*n* = 5), the BPPG R&D lab floor (*n* = 5), and control samples (*n* = 2) for samples collected from the SALSA device versus traditional swab assay in 1 mL of sterile water.

Quantification and quality check

Qubit quantification of extracted DNA of the JPL-SAF collected in August 2022 indicated very low concentrations of DNA (Below Detection Limit (BDL) to 1.05 ng; Table 1) and further required the use of 16S rRNA qPCR. While only 3 samples were within the detection limit using the Qubit HS DNA kit (0.005 ng/µL), all were successfully quantified by 16S rRNA qPCR. This is not unexpected given the extreme cleanliness of the sampling site and lower sensitivity of the Qubit. The qPCR data indicated that all sampling sites demonstrated 16s rRNA abundance 1 to 2 orders of magnitude higher than process control samples (Table 1).

Table 1

Quantification of biomass using different methods throughout the workflow, including unamplified original DNA sample via Qubit, 16S qPCR amplicons, quantification of long-amplification PCR product via Rapid PCR Barcoding kit; reads passing MinKnow's quality filter; and cultivable bacterial species from each sample.

Location	Sample	% Original sample extracted for DNA	Total DNA recovered (ng)[a]	Total 16S copies	Total LongAmp PCR product (ng)[b]	Nanopore reads past filter	Identification of cultivable microorganisms
Anteroom to cleanroom	AR-1	99%	BDL	1.8E+04	3.5	11883	*Bacillus haynesii*
	AR-2	96%	BDL	1.4E+04	BDL	1622	*Sphingomonas jeddahensis*, *Roseomonas vinacea*, *Paracoccus* sp. MC1862
	AR-3	92%	BDL	n.d.	BDL	4333	*Methylbacterium durans*, *Methylobacterium soli*, *Aspergillus pseudoreflectus*
Cleanroom entry area	ECR-1	87%	0.7	3.8E+04	BDL	367	*Paracoccus angustae*, *Bacillus megaterium*
	ECR-2	39%	0.7	1.1E+04	BDL	4880	*Bacillus* sp. PAMC26543
	ECR-3	79%	BDL	2.4E+04	BDL	1684	*Brevibacterium frigoritolerans*

Cleanroom far side	FCR-1	91%	BDL	5.3E+04	5	28732	n/a
	FCR-2	51%	BDL	1.8E+04	4.5	23432	*Sphingomonas jeddahensis*, *Paracoccus caeni*
	FCR-3	93%	1.05	3.5E+04	4.5	24487	*Sphingomonas jeddahensis*
Process control sample	CON-1	57%	BDL	1.2E+03	1.5	6242	n/a
	CON-2	27%	BDL	7.7E+02	1.5	6067	*Bacillus pumilis* SAFR-032
	CON-3	36%	BDL	6.7E+02	1.5	7709	*Paenibacillus lautus*
Negative control sample	NEG-1	60%	BDL	n.d.	BDL	523	n/a

[a]*BDL indicates <0.175 ng (*Limit of Detection (*LOD) 0.005 ng/µL x DNA volume of 35 µL); total DNA indicates concentration (ng/µL) x DNA volume of 35 µL.*

[b]*BDL indicates <0.25 ng (LOD 0.005 ng/µL x PCR reaction volume of 50 µL).*

Nanopore library synthesis and sequencing

Based on Qubit quantification results, the maximum DNA input used for nanopore library preparation (Rapid PCR Barcoding -RPB004) for any sample was 0.21 ng (7-µL input x maximum concentration of 0.03 ng/µL), with most samples below ~0.03 ng. After LongAmp PCR amplification of 18 cycles, quantification indicated that 4 samples amplified strongly, while the remaining samples were at or below the concentration of the process controls. The 3 process control samples also showed low-level amplification. These data were supported at the read level as well (Table 1).

Nanopore sequencing was performed for a 24-hour period and resulted in a total of 318.44 mb of data comprising 138,280 reads passing the MinKNOW quality filter with N50 values of ~2.9 kb (per kit protocol, the N50 value here reflects the Rapid PCR Barcoding kit's amplification step rather than original DNA quality). This yield is significantly below what might be expected for high-biomass samples (which are generally in the Gb, or tens of Gbs, range) but not unexpected for such low-biomass samples. Barcodes were

detected in 121,348 reads (~88%) which were passed on to data analysis. Of those, 87,405 reads were classified as matching to cellular life, and 97% of these reads were classified to the domain level. Using a conservative sequence similarity to NCBI's nr database of 90%, ~50% of cellular reads were annotated to the genus level.

Annotation of nanopore reads

The high-level taxonomic assignment of "passed" barcoded reads indicate that floor samples that successfully amplified via LongAmp PCR above the level of process controls had a much higher percentage of sequences classified in the Proteobacteria, as well as the presence of reads classified in the Bacteroidetes and Arthropoda, compared to process control samples (Figure S1). The top 10 microbial genera detected were *Acinetobacter*, *Brevundimonas*, *Corynebacterium*, *Cutibacterium*, *Paracoccus*, *Pseudomonas*, *Ralstonia*, *Sphingomonas*, *Staphylococcus*, and *Streptococcus* (Figure 4). Of these genera, samples amplifying an order of magnitude above process control samples mostly contained *Paracoccus* and *Acinetobacter* reads, while all other samples including process controls had higher levels of *Staphylococcus* and *Streptococcus* reads. *Sphingomonas* reads made up a much larger percentage of samples from the JPL-SAF air shower/anteroom floor compared to other locations or process control samples. Process control samples resulted in higher levels of *Ralstonia* reads, which made up a much smaller proportion in other samples. This is not unexpected since the process control samples are the background "kitome" for all samples but at a much lower contribution for actual samples.

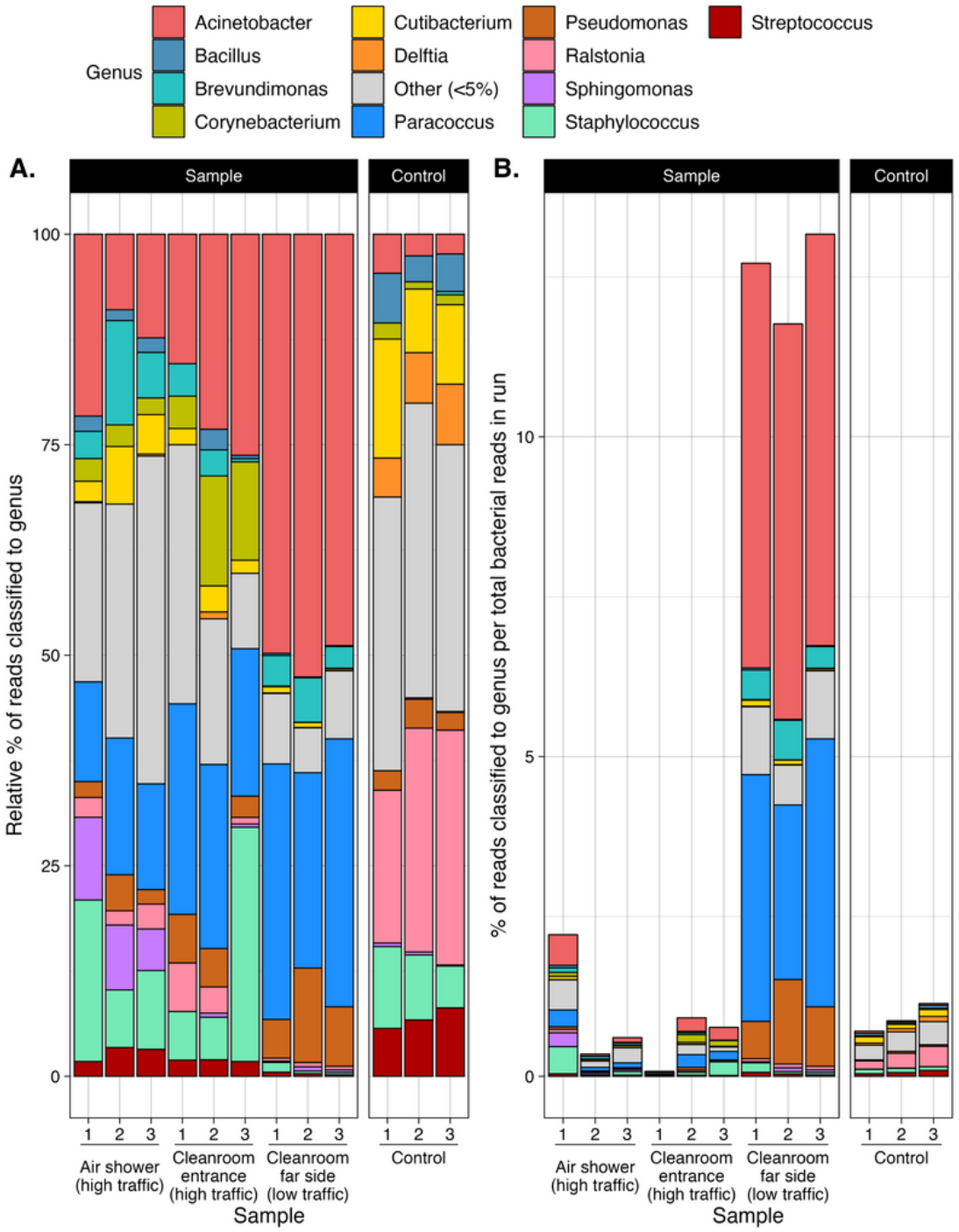


Figure 4
JPL-SAF floor nanopore bacterial reads annotated to the genus level, with taxonomic classifications shown as A) relative percentage of the total sample and B) percent of the total number good quality reads (see Table 1 for totals).

For bacterial reads classified down to the species level (Figure 5), the process control samples revealed only 2 species making up more than 0.25% of species-level bacterial reads, including *Cutibacterium acnes* and

Bacillus thuringiensis. In contrast, 22 species of bacteria accounting for more than 0.25% of species-level bacterial reads were present in floor samples, including 10 species of *Acinetobacter* and 8 species of *Paracoccus.* Using both genus- and species-level data, with and without standardization by number of reads per sample, the microbial communities from FCR samples were significantly different from those of process control samples (Permutational multivariate analysis of variance (PERMANOVA), $P < 0.05$; Figure 6, Figure S2).

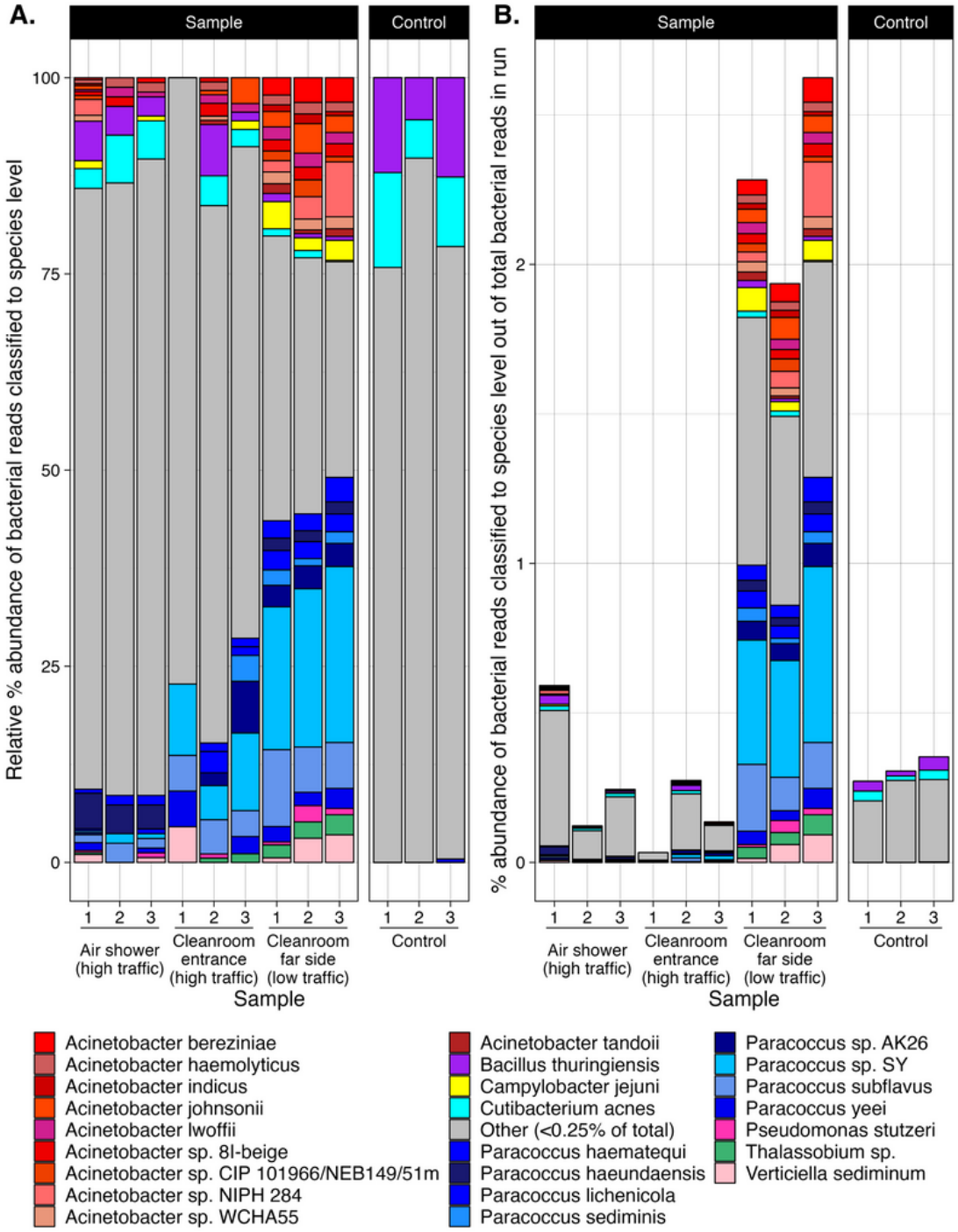


Figure 5

JPL-SAF floor nanopore bacterial reads annotated to the species level expressed as A) relative percentage of the total sample and B) percent of the total number of reads classified to the domain Bacteria.

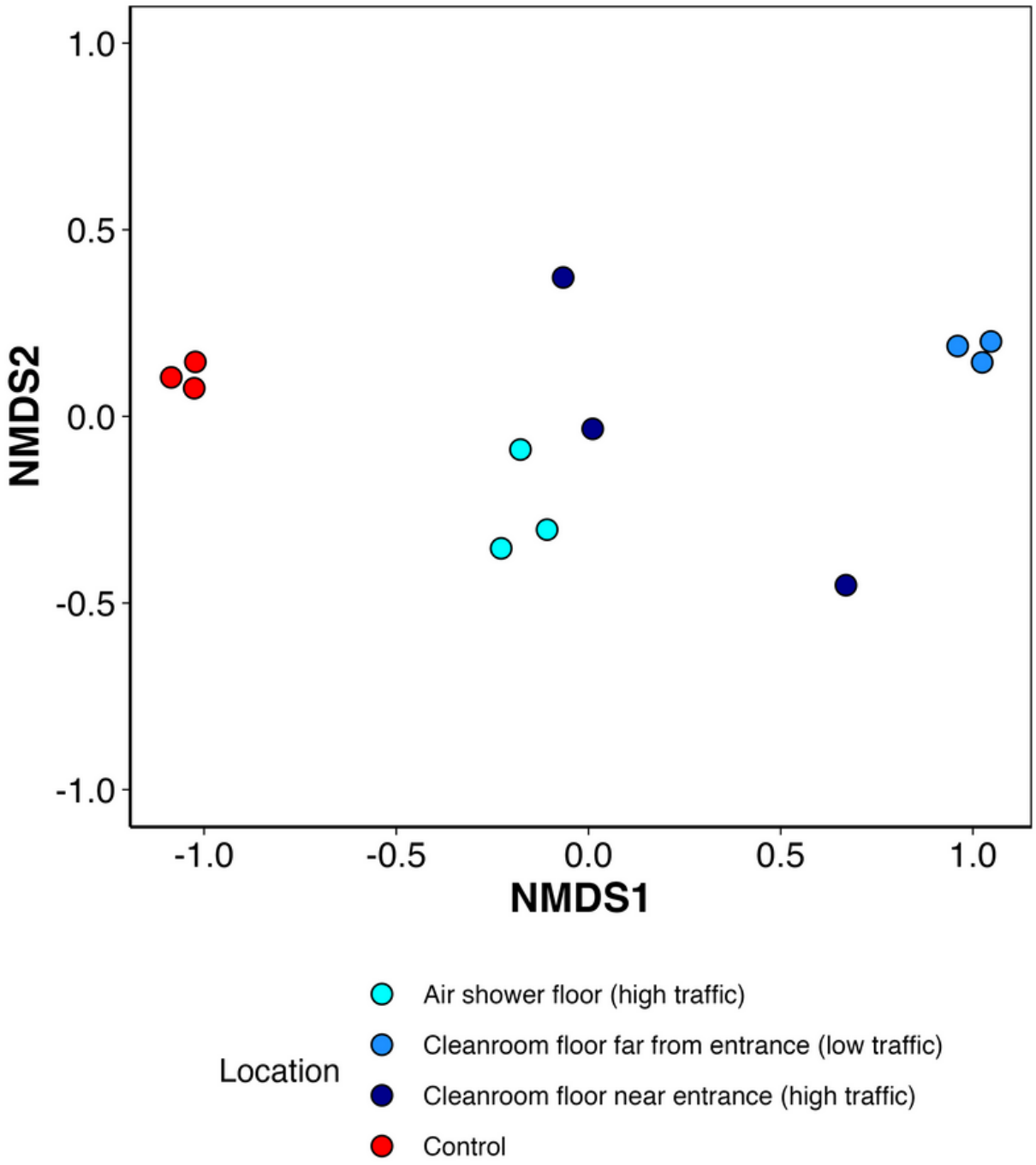


Figure 6
Non-Metric Multidimensional Scaling (NMDS) ordination (Bray–Curtis) using counts of all reads annotated to the genus level. Samples from the low-traffic area which amplified well (FCR) were distinct from control samples.

Floor sample cultivable bioburden

Colony counts and morphotypes ranged from below detection level to 3 colonies per sample (Table 1). For all sample categories, including the 3 sampling areas and processing controls, at least 2 out of 3 samples showed at least 1 cultivable microbe; ~50% of isolates were from 1 of the top 10 genera detected in nanopore reads, supporting the finding that the nanopore method was also able to account for cultivable bacteria. Notably, 1 fungal species was isolated, although only 0.6% of total metagenomic reads were classified as fungi.

DISCUSSION

Use of the SALSA device allowed for successful sequencing of microbial DNA recovered from the ultra-low biomass JPL-SAF cleanroom. Annotation of nanopore sequencing results produced taxonomic information down to the species level, which is the ultimate target for meaningful input into NASA planetary protection probabilistic assessment models. Because the accuracy of sequencing data from the R9.4.1 Nanopore flow cells does not approach Illumina-level quality, species-level data in this study would not be acceptable for input into these models; however, the recently released R10.4.1 flow cells (as of the year 2023) have shown 99%+ accuracy and can now produce species-level data with certainty. Previous swab- and wipe-based approaches to sequencing JPL-SAF surfaces, even with extreme care taken to avoid contamination, showed that due to insufficient sample collection efficiency, it was extremely difficult to distinguish true signal from background contamination.[11],[45] Total sample aspiration without the need to adsorb and elute cells and eDNA to the fibers of swabs or wipes promises to greatly decrease the difficulty and complexity of sampling low-biomass environments for downstream applications such as shotgun metagenomic sequencing. Previous experimentation with the SALSA device has shown that it increases recovery compared to swab techniques, and when sampling the JPL-SAF floor, samples collected by the SALSA device over a 100-cm^2 area yielded an order of magnitude greater number of 16S rRNA gene copies compared to swab assays (Figure 3).

A bias inherent in all microbiome studies is that many or most prokaryotes are not culturable by standard means and are therefore only detectable by metagenomic or amplicon sequencing. Conversely, some prokaryotic species are extremely hardy (e.g., endospores) and difficult to lyse and are only detected via culture. The accuracy of taxonomic classification of metagenomic sequencing data is improving but still poor in many cases; therefore, the culturable microbiome must be considered the “minimum truth” of a sample.[46], [47] In this study, as in most environmental studies utilizing metagenomics, many culturable bacteria were not detected in nanopore microbiome datasets and vice versa. For example, *Acinetobacter* species were not isolated but were the most abundant bacterial genera in the sequencing data, whereas *Methylobacterium* were cultured but not seen in the nanopore sequencing data. Cultivation bias in the culturing of environmental microbes is a well-known issue and could be due to a wide variety of issues: media selection, culture conditions, or nonviability because of the cleaning/maintenance of the cleanroom. In the latter case, microbial cells might be lysed, rendering them non-cultivable but leaving their nucleic acids available for nanopore sequencing. Hence, to understand the true viable microbial burden, during sample processing, naked nucleic acids should be

removed as previously established;[48] although in the case of planetary protection sampling for future life-detection missions, eDNA should still be processed. Similarly, fungal signatures were almost completely absent in nanopore sequencing (0.6% of reads), which could be attributed to poor nucleic acids extraction, although this figure is supported by previous findings in which fungal presence was ~1 to 2% of the total microbial burden in NASA cleanrooms.[21] The cultivation approach revealed the presence of spore formers (*Bacillus* and *Paenibacillus*), biofilm-producing members (*Methylobacterium*), and fungi (*Aspergillus pseudoreflectus*), but their sequences were not retrieved or poorly retrieved via nanopore sequencing, which might be due to their low relative abundance in the samples and/or difficulty in cell lysis. To confirm these issues, more thorough studies are needed.

There is clearly a strong requirement to improve workflows for ultra-low biomass shotgun metagenomic sequencing, including developing and using sterile DNA-free reagents and consumables. While DNA-free metapolyzyme is already available for low-biomass and space applications, other DNA-free reagents, including DNA extraction kits themselves as well as DNA-free library synthesis kits, PBS, Tris, and other consumables, are needed. This is also true for the SALSA device. Even in the event that DNA-free components are developed and used, negative process control samples will still be required to accurately assess the true content of cleanroom samples.[49] The isolation of single colonies of spore-forming bacteria, but no non-spore–forming bacteria, from 2 out of 3 SALSA process control samples suggests that contamination might have arisen from sample processing steps. Numerous control studies (S. Wong, unpublished) have demonstrated that the current sterilization level of the SALSA device is sufficient to produce 0 colonies in carefully controlled laboratory conditions; however, these procedures will need to be reevaluated for field usage and metagenome-based analysis.

While all JPL-SAF samples (including those taken from the entryway and anteroom) amplified an order of magnitude higher than process control samples using conventional 16S rRNA gene qPCR, only 4 samples amplified using the RPB004 nanopore library synthesis protocol. The entryway (high-traffic area) floor samples showed the highest original concentrations of DNA but had lower DNA concentrations than the process controls after LongAmp PCR. While this may be attributed to low DNA input, it has been observed for other sample types (Tighe and Venkateswaran, personal communication) and may not be solely attributed to ultra-low DNA input but related to the presence of inhibitors, lack of efficient fragmentation (due to DNA input size or sequence context), or LongAmp PCR Taq performance. In particular, this could be due to inhibitory surfactants and other chemicals used to clean the JPL-SAF floor remaining in DNA samples, as these high-traffic areas are much more frequently cleaned. Without consistent amplification of these samples, meaningful semiquantitative assessment of functional genes and species may be skewed by vastly different numbers of sequences per sample. The standard nanopore Rapid PCR barcoding protocol (RPB004) was used throughout this study, but it is clear that increased performance is needed. Further consideration using dUTP/UNG PCR chemistry to prevent PCR-forward contamination should also be adopted.

Overall, this study clearly demonstrates that the biomass of the samples to be monitored for microbial abundance should be increased by collecting samples from larger surface areas (eg, novel SALSA sampling device) in combination with sample concentration and subsequent metagenomic sequencing. The coupling of a long-read, modified (using higher inputs and entirely of PCR product) Oxford Nanopore protocol for characterizing low-biomass surfaces combined with the SALSA approach has the potential to provide microbiome data for ultra-low biomass environments within less than 24 hours of collection.

ACKNOWLEDGMENTS

The research was carried out at the JPL, California Institute of Technology, under a contract with NASA (80NM0018D0004). The authors would like to thank Dan Vellone at the University of Vermont for his technical assistance with Oxford Nanopore and Gayane Kazarians from JPL for facilitating the sampling of the JPL-SAF. The authors received no specific funding for this work. The development of the SALSA was funded through the support of the Small Business Innovation Research Program of the National Institutes of Health (contract # 75D30121P11936) and NASA (contract# 80NSSC22PA932). ©2023 California Institute of Technology. Government sponsorship acknowledged.

AUTHOR CONTRIBUTIONS

SW and HW developed the sampling device and performed sample collection training. KV and ST conceived the idea. AS and NS performed spacecraft assembly floor surface sampling. AS, ST, and NS performed nanopore sequencing. ST performed nanopore training. AT and MW performed plating of samples. NS and AS performed bioinformatic and statistical analysis. AS, SW, and ST wrote the manuscript, and KV contributed to editing the manuscript.

CONFLICTS OF INTEREST

SW and HW are employed by AI Biosciences, which is developing the SALSA instrument under Small Business Innovation Research (Centers for Disease Control and Prevention, NASA). However, all authors declare that there are no conflicts of interest. This manuscript was prepared as an account of work sponsored by NASA, an agency of the US Government. The US Government, NASA, California Institute of Technology, JPL, and their employees make no warranty, expressed or implied, or assume any liability or responsibility for the accuracy, completeness, or usefulness of information, apparatus, product, or process disclosed in this manuscript or represent that its use would not infringe upon privately held rights. The use of, and references to, any commercial product, process, or service does not necessarily constitute or imply endorsement, recommendation, or favoring by the US Government, NASA, California Institute of Technology, or JPL. The views and opinions presented herein by the authors of this manuscript do not necessarily reflect those of the US Government, NASA, California Institute of Technology, or JPL and shall not be used for advertisements or product endorsements.

SUPPLEMENTARY MATERIAL

FigureS1.tiff 87 MB

Figure S1. Taxonomic assignment of total Nanopore reads.

FigureS2.tiff 63 MB

Figure S2. Two-dimensional NMDS ordination of read counts for all reads classified to the genus level using Bray–Curtis distance measure standardized by total number of reads. Multiple regression of taxonomic categories to the 2-dimensional ordination was performed using the envfit function in vegan; vectors for taxonomic categories that significantly ($P < 0.01$) correlated with NMDS axes after false discovery rate (FDR) P-value corrections are visualized (note: the number of dimensions necessary to describe the data was checked via the dimcheckMDS function; more than 2 dimensions reduced stress to near zero).

FigureS3.tiff 4 MB

Figure S3. Quality scores for the Nanopore run, averaging a Qscore of approximately 13.

References

1. Kerkhof LJ. Is Oxford Nanopore sequencing ready for analyzing complex microbiomes? *FEMS Microbiol Ecol*. 2021;97(3):fiab001. doi:10.1093/femsec/fiab001 ↩
2. Johnson SS, Zaikova E, Goerlitz DS, Bai Y, Tighe SW. Real-time DNA sequencing in the Antarctic dry valleys using the Oxford Nanopore sequencer. *J Biomol Tech*. 2017;28(1):2-7. doi:10.7171/jbt.17-2801-009 ↩
3. Quick J, Loman NJ, Duraffour S, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature*. 2016;530(7589):228-232. doi:10.1038/nature16996 ↩
4. Tedersoo L, Albertsen M, Anslan S, Callahan B. Perspectives and benefits of high-throughput long-read sequencing in microbial ecology. *Appl Environ Microbiol*. 2021;87(17):e0062621. doi:10.1128/AEM.00626-21 ↩
5. Delahaye C, Nicolas J. Sequencing DNA with nanopores: troubles and biases. *PLoS One*. 2021;16(10):e0257521. doi:10.1371/journal.pone.0257521 ↩

6. Bruijns BB, Tiggelaar RM, Gardeniers H. The extraction and recovery efficiency of pure DNA for different types of swabs. *J Forensic Sci*. 2018;63(5):1492-1499. doi:10.1111/1556-4029.13837 ↩
7. Canfield JR, Jollie M, Worst T, Oechsle C. Comparison of swab types & elution buffers for collection and analysis of intact cells to aid in deconvolution of complex DNA mixtures. *Forensic Sci Int*. 2022;340:111448. doi:10.1016/j.forsciint.2022.111448 ↩
8. Ogai K, Nagase S, Mukai K, et al. A comparison of techniques for collecting skin microbiome samples: swabbing versus tape-stripping. *Front Microbiol*. 2018;9:2362. doi:10.3389/fmicb.2018.02362 ↩
9. Bonsu DO, Higgins D, Henry J, Austin JJ. Evaluation of the efficiency of Isohelix™ and Rayon swabs for recovery of DNA from metal surfaces. *Forensic Sci Med Pathol*. 2021;17(2):199-207. doi:10.1007/s12024-020-00329-x ↩
10. Kwan K, Cooper M, La Duc M, et al. Evaluation of procedures for the collection, processing, and analysis of biomolecules from low-biomass surfaces. *Appl Environ Microbiol*. 2011;77(9):2943-2953. doi:10.1128/AEM.02978-10 ↩
11. Bargoma E, La Duc M, Kwan K, Vaishampayan P, Venkateswaran K. Differential recovery of phylogenetically disparate microbes from spacecraft-qualified metal surfaces. *Astrobiology*. 2013;13(2):189-202. doi:10.1089/ast.2012.0917 ↩
12. Kittigul L, Ekchaloemkiet S, Utrarachkij F, et al. An efficient virus concentration method and RT-nested PCR for detection of rotaviruses in environmental water samples. *J Virol Methods*. 2005;124(1-2):117-122. doi:10.1016/j.jviromet.2004.11.013 ↩
13. Checinska Sielaff A, Urbaniak C, Mohan GBM, et al. Characterization of the total and viable bacterial and fungal communities associated with the International Space Station surfaces. *Microbiome*. 2019;7(1):50. doi:10.1186/s40168-019-0666-x ↩
14. Ruppert KM, Kline RJ, Rahman MS. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA. *Glob Ecol Conserv*. 2019;17:e00547. doi:10.1016/j.gecco.2019.e00547 ↩
15. Minich JJ, Zhu Q, Janssen S, et al. KatharoSeq enables high-throughput microbiome analysis from low-biomass samples. *mSystems*. 2018;3(3):e00218-17. doi:10.1128/mSystems.00218-17 ↩
16. Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog*. 2016;8:24. doi:10.1186/s13099-016-0103-7 ↩

17. Weyrich LS, Farrer AG, Eisenhofer R, et al. Laboratory contamination over time during low-biomass sample analysis. *Mol Ecol Resour*. 2019;19(4):982-996. doi:10.1111/1755-0998.13011 ↩
18. de Goffau MC, Lager S, Salter SJ, et al. Recognizing the reagent microbiome. *Nat Microbiol*. 2018;3(8):851-853. doi:10.1038/s41564-018-0202-y ↩
19. Mojarro A, Hachey J, Ruvkun G, Zuber MT, Carr CE. CarrierSeq: a sequence analysis workflow for low-input nanopore sequencing. *BMC Bioinformatics*. 2018;19(1):108. doi:10.1186/s12859-018-2124-3 ↩
20. Wood JM, Singh NK, Guan L, Seuylemezian A, Benardini JN, Venkateswaran K. Performance of multiple metagenomics pipelines in understanding microbial diversity of a low-biomass spacecraft assembly facility. *Front Microbiol*. 2021;12:685254. doi:10.3389/fmicb.2021.685254 ↩
21. Hendrickson R, Urbaniak C, Minich JJ, et al. Clean room microbiome complexity impacts planetary protection bioburden. *Microbiome*. 2021;9(1):238. doi:10.1186/s40168-021-01159-x ↩
22. Wood JM, Singh NK, Guan L, Seuylemezian A, Benardini JN, Venkateswaran K. Performance of multiple metagenomics pipelines in understanding microbial diversity of a low-biomass spacecraft assembly facility. *Frontiers Microbiol*. 2021;12:685254. doi:10.3389/fmicb.2021.685254 ↩
23. Danko DC, Sierra MA, Benardini JN, et al. A comprehensive metagenomics framework to characterize organisms relevant for planetary protection. *Microbiome*. 2021;9(1):82. doi:10.1186/s40168-021-01020-1 ↩
24. Mojarro A, Hachey J, Bailey R, et al. Nucleic acid extraction and sequencing from low-biomass synthetic mars analog soils for in situ life detection. *Astrobiology*. 2019;19(9):1139-1152. doi:10.1089/ast.2018.1929 ↩
25. Heavens D, Chooneea D, Giolai M, et al. How low can you go? Driving down the DNA input requirements for nanopore sequencing. *bioRxiv*. October 18, 2021. doi:10.1101/2021.10.15.464554 ↩
26. Bargoma E, La Duc MT, Kwan K, Vaishampayan P, Venkateswaran K. 2013. Differential recovery of phylogenetically disparate microbes from spacecraft-qualified metal surfaces. Astrobiology 13:189-202 ↩
27. Bargoma E, La Duc MT, Kwan K, Vaishampayan P, Venkateswaran K. Differential recovery of phylogenetically disparate microbes from spacecraft-qualified metal surfaces. *Astrobiology*. 2013;13(2):189-202. doi:10.1089/ast.2012.0917 ↩
28. Wise NM, Wagner SJ, Worst TJ, Sprague JE, Oechsle CM. Comparison of swab types for collection and analysis of microorganisms. *Microbiologyopen*. 2021;10(6):e1244. doi:10.1002/mbo3.1244 ↩
29. Benardini JN III, La Duc MT, Beaudet RA, Koukol R. Implementing planetary protection measures on the Mars Science Laboratory. *Astrobiology*. 2014;14(1):27-32. doi:10.1089/ast.2013.0989 ↩

30. Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. *PCR Protocols: A Guide to Methods and Applications*. Academic Press; 2012 ↩
31. Simpson AC, Suzuki T, Miller DR, Venkateswaran K. Microbial burden estimation of food items, built environments, and the International Space Station using film media. *Microorganisms*. 2022;10(9):1714. doi:10.3390/microorganisms10091714 ↩
32. Yonkus JA, Whittle E, Alva-Ruiz R, et al. “Answers in hours”: a prospective clinical study using nanopore sequencing for bile duct cultures. *Surgery*. 2022;171(3):693-702. doi:10.1016/j.surg.2021.09.037 ↩
33. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 2019;20(1):257. doi:10.1186/s13059-019-1891-0 ↩
34. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*. 2019;26(6):1925-1927. doi:10.1093/bioinformatics/btz848 ↩
35. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 2021;18(4):366-368. doi:10.1038/s41592-021-01101-x ↩
36. Beier S, Tappu R, Huson DH. Functional analysis in metagenomics using MEGAN 6. In: Charles TC, Liles MR, Sessitsch A, eds. *Functional Metagenomics: Tools and Applications*. Springer; 2017:65-74 ↩
37. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(D1):D733-D745. doi:10.1093/nar/gkv1189 ↩
38. Mizrachi I. GenBank. In: Beck J, Benson D, Coleman J, et al., eds. *The NCBI Handbook [Internet]*. 2nd ed. National Center for Biotechnology Information (US). 2013. ↩
39. Kanehisa M, Araki M, Goto S, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res*. 2008;36(suppl_1):D480-D484. doi:10.1093/nar/gkm882 ↩
40. Overbeek R, Olson R, Pusch GD, et al. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res*. 2014;42(D1):D206-D214. doi:10.1093/nar/gkt1226 ↩
41. Huerta-Cepas J, Szklarczyk D, Heller D, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47(D1):D309-D314. doi:10.1093/nar/gky1085 ↩
42. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. *J Open Source Softw*. 2019;4(43):1686. doi:10.21105/joss.01686 ↩

43. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci*. 2003;14(6):927-930. doi:10.1111/j.1654-1103.2003.tb02228.x ↩

44. Blachowicz A, Mhatre S, Singh NK, et al. The isolation and characterization of rare mycobiome associated with spacecraft assembly cleanrooms. *Front Microbiol*. 2022;13:777133. doi:10.3389/fmicb.2022.777133 ↩

45. Sierra MA, Ryon KA, Tierney BT, et al. Microbiome and metagenomic analysis of Lake Hillier Australia reveals pigment-rich polyextremophiles and wide-ranging metabolic adaptations. *Environ Microbiome*. 2022;17(1):60. doi:10.1186/s40793-022-00455-9 ↩

46. Hiergeist A, Gläsner J, Reischl U, Gessner A. Analyses of intestinal microbiota: culture versus sequencing. *ILAR J*. 2015;56(2):228-240. doi:10.1093/ilar/ilv017 ↩

47. Vaishampayan P, Probst AJ, La Duc MT, et al. New perspectives on viable microbial communities in low-biomass cleanroom environments. *ISME J*. 2013;7(2):312-324. doi:10.1038/ismej.2012.114 ↩

48. Lorenz MG. DNA-free reagents and materials for molecular pathogen analysis of bacterial and fungal pathogens. *Res Mol Microbiol*. 2019;1(19):1-8. ↩