Air monitoring by nanopore sequencing

Tim Reska[1,2,3+], Sofya Pozdniakova[4+], Silvia Borras[4], Michael Schloter[3,5], Lidia Cañas[4], Amit Fenn[1,2], Xavier Rodó[4,6], Barbro Winkler[7], Jörg-Peter Schnitzler[7], Lara Urban[1,2,3*]

[1]Helmholtz AI, Helmholtz Zentrum Muenchen, Neuherberg, Germany
[2]Helmholtz Pioneer Campus, Helmholtz Zentrum Muenchen, Neuherberg, Germany
[3]Technical University of Munich, School of Life Sciences, Freising, Germany
[4]Climate and Health (CLIMA) Program, ISGlobal, Barcelona, Spain
[5]Institute of Comparative Microbiome Analysis, Helmholtz Zentrum Muenchen, Neuherberg, Germany
[6]Catalan Institution for Research and Advanced Studies ICREA, Barcelona, Spain
[7]Research Unit Environmental Simulation (EUS), Helmholtz Zentrum Muenchen, Neuherberg, Germany

[+]contributed equally
[*]corresponding author: lara.h.urban@gmail.com

Abstract

While the air microbiome and its diversity are known to be essential for human health and ecosystem resilience, the current lack of holistic air microbial diversity monitoring means that little is known about the air microbiome's composition, distribution, or functional effects. We here show that nanopore shotgun sequencing can robustly assess the air microbiome in combination with active air sampling through liquid impingement and tailored computational analysis. We provide laboratory and computational protocols for air microbiome profiling, and reliably assess the taxonomic composition of the core air microbiome in controlled and natural environments. Based on *de novo* assemblies from the long sequencing reads, we further identify the taxa of several air microbiota down to the species level and annotate their ecosystem functions, including their potential role in natural biodegradation processes. As air monitoring by nanopore sequencing can be employed in an automatable, fast, and portable manner for *in situ* applications all around the world, we envision that this approach can help holistically explore the role of the air microbiome.

Background

The air microbiome encompasses a broad spectrum of bioaerosols, ranging from bacteria, archaea, fungi, and viruses, to bacterial endotoxins, mycotoxins, and pollen (1). While it is known to serve pivotal functions for human health and ecosystem resilience, its composition, distribution, and dose-response relationships remain poorly understood (2). Infectious diseases such as COVID-19 or tuberculosis have shifted past research efforts towards pathogenic microbial taxa; however, exposure to a diverse air microbiome has also increasingly been considered as a health promoting factor, which underlines the need for holistic air microbial diversity monitoring (3). Recent advances in the molecular assessment of microbiomes have revealed first insights into the complex nature and diverse natural and anthropogenic origin of bioaerosols (4), such as the vertical-altitudinal stratification of microbial abundance and distribution (5), and substantial diurnal, seasonal, temperature, and humidity fluctuations (6).

Most genomics-based air microbiome studies have been based on targeted sequencing via metabarcoding due to the air's ultra-low biomass (1, 4). While metabarcoding can increase the sensitivity of taxonomic detection, it is inherently limited due to amplification biases and incomplete databases. Metagenomics, on the contrary, is based on shotgun sequencing of native DNA and is therefore not subjected to amplification biases while allowing for *de novo* reconstruction of potentially complete microbial genomes for robust species identification and functional annotation, even of previously unidentified taxa (7). Recent advances in long-read metagenomics further facilitate *de novo* genome assemblies (8) and assessments of highly repetitive genomic regions, for example in antimicrobial resistance genes (9). Nanopore sequencing technology can create such long sequencing reads while simultaneously being easily employable and automatable (10) as well as fast and portable for potential applications in clinical (11) or remote settings (12). While this technology has been used to characterize the microbial diversity of various environments such as freshwater (13) and dust (14), no protocol exists to date that would allow to leverage the advantages of nanopore long-read sequencing for monitoring the taxonomic and functional diversity of the air microbiome.

Here, we established laboratory and computational protocols to allow for robust air microbiome profiling through nanopore metagenomics. We first evaluated the suitability of long-read shotgun sequencing for robustly assessing the air microbiome in a controlled environment, and then applied the protocol to a natural environment for validation. We show that nanopore sequencing is a robust tool to describe the composition and diversity of microbial taxa in the air, and to concurrently annotate *de novo* microbial genomes to evaluate potential consequences for human and environmental health.

Results and discussion

After testing several technological options for air microbiome assessments using nanopore sequencing (**Supplementary Information**), we applied our optimized protocol in a greenhouse (Gh) as a controlled environment (**Figure 1A**) for three consecutive days. We sampled air using liquid impingement methodology for one or three hours (Materials and methods; **Supplementary Table 1**). Nanopore metagenomics delivered a median read length of 965 bases (**Figure 1B**) and resulted in between 7k and 60k high-quality sequencing reads, of which between 5k and 35k could successfully be mapped to

the taxonomic genus level using Kraken2 and the NCBI prokaryote, viral and fungal database (**Figures 1C-D**; Material and methods). After downsampling to the same number of reads across one-hour (1h) and three-hour (3h) samples, the taxonomic composition of the 20 most abundant taxa showed that only three hours, and not one hour, of sampling resulted in a stable "core" air microbiome across days for genus level assessments (**Figures 1E-F**). These taxonomic assessments were similar when using protein-level or hybrid read- and assembly-based read assignment methods, both on the taxonomic phylum (**Supplementary Figure 1**) and genus level (**Supplementary Figure 2**). The most abundant genera were soil- and plant-associated bacteria such as *Paracoccus*, *Bradyrhizobium*, *Massilia*, or *Streptomyces*. Other genera such as *Geodermatophilus* and *Sphingomonas* are known for their resilience to adverse environmental conditions (15, 16).

As three hours of air sampling resulted in robust air microbiome assessments in the controlled setting, we repeated this experiment across consecutive days in a natural environment (Nat), namely a natural grassland close to Munich (**Figure 1G**). As we expected larger microbiome composition fluctuations in the natural setting, we additionally interspersed the 3h-sampling days with 6h-sampling days (Materials and methods; **Supplementary Table 1**). Nanopore metagenomics delivered a median read length of 1560 bases (**Figure 1H**) and resulted in between 130k and 200k high-quality sequencing reads, of which between 70k and 140k could successfully be mapped to the taxonomic genus level (**Figures 1I-J**). After downsampling all samples to 70k reads, the taxonomic composition of the 20 most abundant taxa showed that both three and six hours of sampling resulted in a stable core microbiome (**Figures 1K-L**).

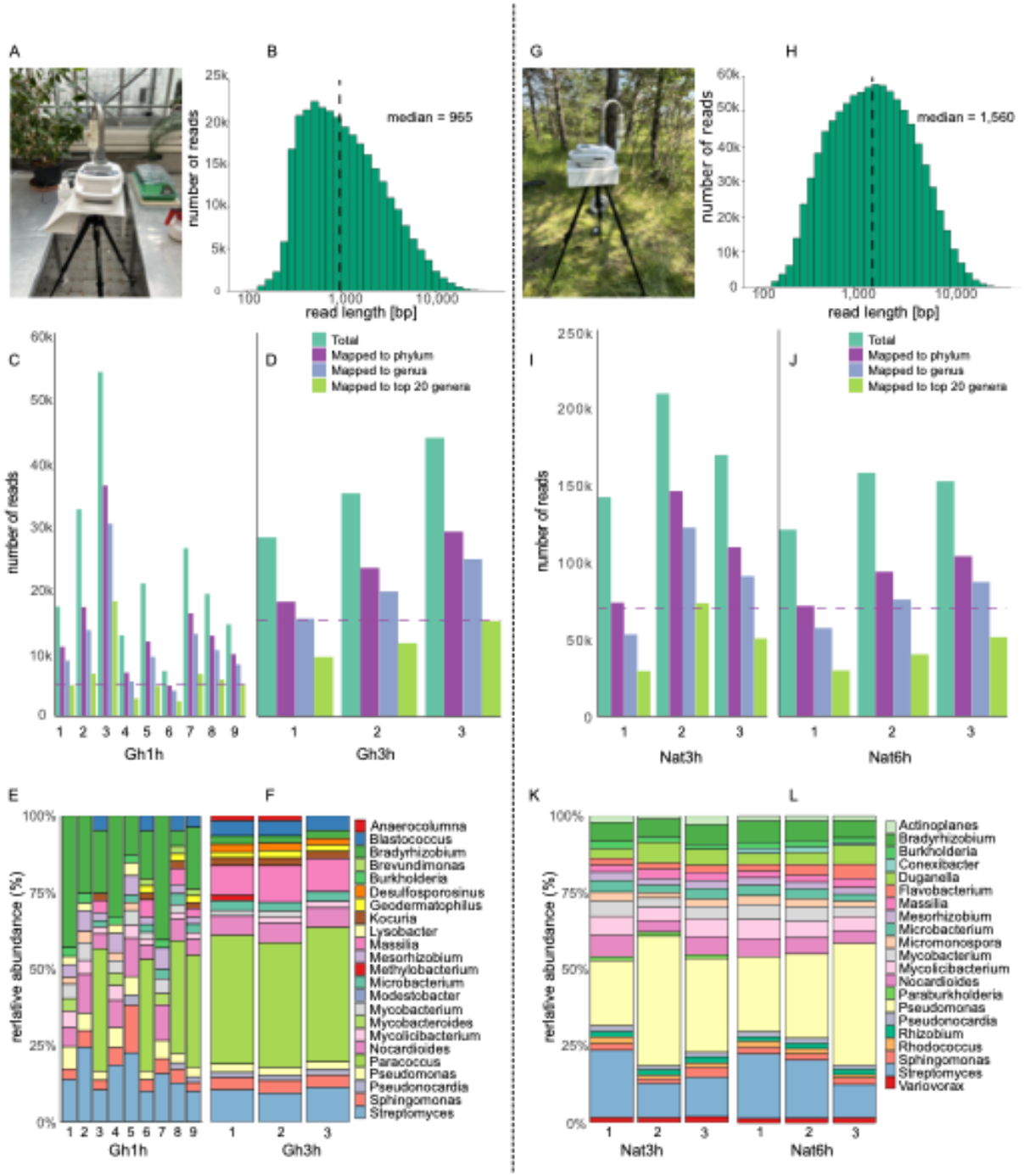

Figure 1. Robust air microbiome assessments through nanopore shotgun sequencing. **A.** Liquid impingement and nanopore sequencing set-up in a greenhouse (Gh) as controlled environment. **B.** Nanopore sequencing read length distribution across 1h- and 3-Gh samples. **C-D.** Number of total sequencing reads, and of reads mapping to taxonomic phylum and genus level as well as to the 20 most dominant genera using Kraken2 (Material and methods) across the **C.** 1h- and **D.** 3-Gh samples; the downsampling threshold across samples is indicated by the dashed horizontal line. **E-F.** Taxonomic composition of the **E.** 1h- and **F.** 3h-samples after downsampling based on the 20 most dominant genera across samples. **G.** Air microbiome monitoring in a natural environment (Nat). **H.** Nanopore sequencing read length distribution across 3h- and 6h-Nat samples. **I-J.** Number of total sequencing reads, and of reads mapping to taxonomic phylum and genus level as well as to the 20 most dominant genera using Kraken2 across the **I.** 3h- and **J.** 6-Gh samples; the downsampling threshold across samples is indicated by the dashed horizontal line. **K-L.** Taxonomic composition of the **K.** 3h- and **L.** 6h-samples after downsampling based on the 20 most dominant genera across samples.

All samplings, DNA extractions, and sequencing negative controls resulted in low DNA yields of <0.1 ng; nanopore shotgun sequencing of the controls identified the typical contaminant species *Escherichia*, *Salmonella*, *Shigella*, *Francisella*, and *Pseudomonas* (**Supplementary Figure 3A-B**) (17), showing that no external contamination had impacted our assessment of air as a low-biomass ecosystem. The application of our protocol to a well-defined mock community further showed that all bacterial and fungal species could be detected with approximately correct abundance estimates; while especially the fungal taxa and Gram-positive *Bacillus subtilis* were underrepresented by our genomic data, this might reflect our relatively gentle DNA extraction approach, which was used to prevent further DNA fragmentation (**Supplementary Figure 3C**; Materials and methods).

Our *de novo* assembly and binning approaches resulted in several assemblies with completeness above 20 % and contamination below 5 % (**Table 1**; Material and methods). Functional annotation of the contigs and metagenome-assembled genomes (MAGs) allowed us to predict a wide array of functions of the air microbiome (**Supplementary Figure 4**), extending to putative biodegradation mechanisms (**Table 1**). Such *de novo* analyses can be crucial for profiling previously understudied microbiomes such as bioaerosols. We, for example, obtained a *de novo* assembly of *Sphingomonas alba*, which has previously only been defined through a soil isolate (18) and might therefore represent a novel strain with important functional variation. Our genome annotation identified genes (*fre, por, ribBA*) from flavin-based metabolic cycles, and a gene (*cher1*) which plays a role in biofilm production (**Table 1**), which might confirm the role of *Sphingomonas* as a degrader of organic material (18) and could identify it as a biomarker of air pollution.

Table 1. *De novo* genome assembly results across all samples from the controlled (Gh) and natural (Nat) environment. Contigs were assembled and then used to identify metagenome-assemblies (MAGs), their taxonomic origin, completeness, and contamination. Functional annotation was filtered to provide insights into natural biodegradation capacity (Material and methods).

Sample	# contigs	N50 contigs	MAG species	complete [%]	contam. [%]	genes
GH3h	121	15,330	*Paracoccus aerius*	64.59	2.94	arsC; cadA; chrA; cnrA; copA; msrA
			P. denitrificans	63.41	1.46	arsC; cadA; chrA; cnrA; copA; msrA
			P. aerius	26.17	2.42	arsC; cadA; chrA; cnrA; copA; msrA
			P. aerius	23.76	0.15	arsC; cadA; chrA; cnrA; copA; msrA
			P. denitrificans	20.14	1.78	arsC; cadA; chrA; cnrA; copA; msrA
Nat3h	204	7,401	*Pseudomonas putida*	23.29	1.66	bphA; msrB; msrC; pcaG
			P. putida	23.82	1.76	bphA; msrB; msrC; pcaG
Nat6h	117	7,282	*Sphingomonas alba*	21.43	4.78	flr; cher1; ribBA

We have shown that nanopore shotgun sequencing can reliably assess the air microbiome in combination with active air sampling through liquid impingement and a tailored computational analysis pipeline. We have leveraged the latest nanopore sequencing chemistry (R10.4.1), whose relatively longer nanopores allow for high sequencing accuracy (>99%) and genome assembly without additional polishing (19), and whose minimum DNA input requirements have further dropped (from previously 400 ng to now 50 ng). Despite the ultralow abundance of biomass in the air, we have shown that only a few hours of active air sampling results in reproducible air microbiome assessments which remain stable across consecutive days – in a controlled setting in a greenhouse, as well as in a natural environment.

Our reproducible computational pipeline provides preliminary results with evidence for a core air microbiome that remains stable across temporal scales, with some core taxa such as *Sphingomonas* being known for their quick evolutionary adaptation potential, which allows them to exist in various environments (20). Our *de novo* genome assemblies importantly allow for in-depth taxonomic and functional profiling of potentially previously undiscovered microbial taxa. We simultaneously emphasize that the contribution of certain microbial taxa such as fungi are most likely underestimated in our study (7); while a more rigorous DNA extraction approach could provide access to fungal DNA despite their often sturdy cell walls, we anticipate that this could further fragment our DNA and hamper our computational analyses. As a solution, we envision the simultaneous sequencing of several DNA extracts, which can subsequently be combined in a holistic computational analysis.

We envision that the application of nanopore sequencing for air monitoring can provide the basis for fast, robust, and automated characterizations of the air microbiome at centralized and remote settings. This characterization importantly goes beyond the air's taxonomic composition but can describe functions related to human and ecosystem health, such as bioremediation mechanisms which can contribute to the degradation of pollutants. We further envision that more targeted applications, for example for pathogen and drug resistance and virulence gene detection, can be leveraged to better understand infectious disease transmission patterns.

References

1. Whitby C, Ferguson RMW, Colbeck I, Dumbrell AJ, Nasir ZA, Marczylo E, et al. Chapter Three - Compendium of analytical methods for sampling, characterization and quantification of bioaerosols. In: Bohan DA, Dumbrell A, editors. Advances in Ecological Research. 67: Academic Press; 2022. p. 101-229.
2. O'Connor DJ, Daly SM, Sodeau JR. On-line monitoring of airborne bioaerosols released from a composting/green waste site. Waste Manag. 2015;42:23-30.
3. Robinson JM, Breed MF. The aerobiome-health axis: a paradigm shift in bioaerosol thinking. Trends Microbiol. 2023;31(7):661-4.
4. Naumova NB, Kabilov MR. About the Biodiversity of the Air Microbiome. Acta Naturae. 2022;14(4):50-6.
5. Drautz-Moses DI, Luhung I, Gusareva ES, Kee C, Gaultier NE, Premkrishnan BNV, et al. Vertical stratification of the air microbiome in the lower troposphere. Proc Natl Acad Sci U S A. 2022;119(7).
6. Gusareva ES, Acerbi E, Lau KJX, Luhung I, Premkrishnan BNV, Kolundžija S, et al. Microbial communities in the tropical air ecosystem follow a precise diel cycle. Proc Natl Acad Sci U S A. 2019;116(46):23299-308.
7. Luhung I, Uchida A, Lim SBY, Gaultier NE, Kee C, Lau KJX, et al. Experimental parameters defining ultra-low biomass bioaerosol analysis. NPJ Biofilms Microbiomes. 2021;7(1):37.
8. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. Nature Biotechnology. 2020;38(6):701-7.
9. Chan WS, Au CH, Chung Y, Leung HCM, Ho DN, Wong EYL, et al. Rapid and economical drug resistance profiling with Nanopore MinION for clinical specimens with low bacillary burden of Mycobacterium tuberculosis. BMC Research Notes. 2020;13(1):444.
10. Raymond-Bouchard I, Maggiori C, Brennan L, Altshuler I, Manchado JM, Parro V, Whyte LG. Assessment of Automated Nucleic Acid Extraction Systems in Combination with MinION Sequencing As Potential Tools for the Detection of Microbial Biosignatures. Astrobiology. 2022;22(1):87-103.
11. Sauerborn E, Reska T, Hamway Y, Urban L. Real-time metagenomics for rapid clinical pathogen identification and resistance prediction in gram-negative bloodstream infections. 2023.
12. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530(7589):228-32.
13. Urban L, Holzer A, Baronas JJ, Hall MB, Braeuninger-Weimer P, Scherm MJ, et al. Freshwater monitoring by nanopore sequencing. eLife. 2021;10:e61504.
14. Nygaard AB, Tunsjø HS, Meisal R, Charnock C. A preliminary study on the potential of Nanopore MinION and Illumina MiSeq 16S rRNA gene sequencing to characterize building-dust microbiomes. Scientific Reports. 2020;10(1):3209.
15. Zhang X, Xia X, Dai M, Cen J, Zhou L, Xie J. Microplastic pollution and its relationship with the bacterial community in coastal sediments near Guangdong Province, South China. Sci Total Environ. 2021;760:144091.
16. Castro JF, Nouioui I, Sangal V, Trujillo ME, Montero-Calasanz MDC, Rahmani T, et al. Geodermatophilus chilensis sp. nov., from soil of the Yungay core-region of the Atacama Desert, Chile. Syst Appl Microbiol. 2018;41(5):427-36.
17. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology. 2014;12(1):87.
18. Siddiqi MZ, Rajivgandhi G, Lee S-Y, Im W-T. Characterization of four novel bacterial species of the genus Sphingomonas, Sphingomonas anseongensis, Sphingomonas alba, Sphingomonas brevis and Sphingomonas hankyongi sp.nov., isolated from wet land. International Journal of Systematic and Evolutionary Microbiology. 2023;73(5).

19. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, Wollenberg RD, Albertsen M. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. Nature Methods. 2022;19(7):823-6.
20. Lombardino J, Bijlani S, Singh NK, Wood JM, Barker R, Gilroy S, et al. Genomic Characterization of Potential Plant Growth-Promoting Features of Sphingomonas Strains Isolated from the International Space Station. Microbiol Spectr. 2022;10(1):e0199421.