ISME Communications

Air monitoring by nanopore sequencing

--Manuscript Draft--

Manuscript Number:	
Full Title:	Air monitoring by nanopore sequencing
Article Type:	Original Article
Keywords:	air microbiome; bioaerosols; metagenomics; nanopore sequencing; long-read sequencing; shotgun sequencing; de novo assembly; infectious disease; antimicrobial resistance; urban air microbiome
Corresponding Author:	Lara Urban Helmholtz Zentrum München Deutsches Forschungszentrum für Umwelt und Gesundheit: Helmholtz Zentrum Munchen Deutsches Forschungszentrum fur Gesundheit und Umwelt GERMANY
Corresponding Author's Institution:	Helmholtz Zentrum München Deutsches Forschungszentrum für Umwelt und Gesundheit: Helmholtz Zentrum Munchen Deutsches Forschungszentrum fur Gesundheit und Umwelt
Other Authors:	Tim Reska
	Sofya Pozdniakova
	Silvia Borras
	Albert Perlas
	Ela Sauerborn
	Lidia Canas
	Michael Schloter
	Xavier Rodo
	Yuanyuan Wang
	Barbro Winkler
	Jörg-Peter Schnitzler
Abstract:	While the air microbiome and its diversity are essential for human health and ecosystem resilience, comprehensive air microbial diversity monitoring has remained rare, so that little is known about the air microbiome's composition, distribution, or functionality. Here we show that nanopore sequencing-based metagenomics can robustly assess the air microbiome in combination with active air sampling through liquid impingement and tailored computational analysis. We provide fast and portable laboratory and computational approaches for air microbiome profiling, which we leverage to robustly assess the taxonomic composition of the core air microbiome of a controlled greenhouse environment and of a natural outdoor environment. We show that long-read sequencing can resolve species-level annotations and specific ecosystem functions through de novo metagenomic assemblies despite the low amount of fragmented DNA used as an input for nanopore sequencing. We then apply our pipeline to assess the diversity and variability of an urban air microbiome, using Barcelona, Spain, as an example; this randomized experiment gives first insights into the presence of highly stable location-specific air microbiomes within the city's boundaries, and showcases the robust microbial assessments that can be achieved through automatable, fast, and portable nanopore sequencing technology.
Additional Information:	
Question	**Response**
Please state the numeric word count of your manuscript, counting from the	4813

beginning of the abstract to the end of the main manuscript (excluding references and figure legends). If substantially longer than the limit (e.g., 5,000 words for an Article), you may be asked to reduce the manuscript's length, before or after review. Word counts for various publication types are detailed in our Author Instructions.	
Has this manuscript been submitted to this journal previously?	Yes
Please indicate the manuscript number for the previous submission. as follow-up to "Has this manuscript been submitted to this journal previously?"	ISMECOMMUN-D-24-00010

Air monitoring by nanopore sequencing

Tim Reska[1,2,3+], Sofya Pozdniakova[4+], Sílvia Borràs[4], Albert Perlas[1,2], Ela Sauerborn[1,2,3], Lídia Cañas[4], Michael Schloter[3,5], Xavier Rodó[4,6], Yuanyuan Wang[7], Barbro Winkler[8], Jörg-Peter Schnitzler[8], Lara Urban[1,2,3*]

[1]Helmholtz AI, Helmholtz Zentrum München, Neuherberg, Germany

[2]Helmholtz Pioneer Campus, Helmholtz Zentrum München, Neuherberg, Germany

[3]Technical University of Munich, School of Life Sciences, Freising, Germany

[4]AIRLAB, Climate and Health (CLIMA) group, ISGlobal, Barcelona, Spain

[5]Institute of Comparative Microbiome Analysis, Helmholtz Zentrum München, Neuherberg, Germany

[6]Catalan Institution for Research and Advanced Studies, ICREA, Barcelona, Spain

[7]Technical University of Munich, School of Engineering and Design, Munich, Germany

[8]Research Unit Environmental Simulation (EUS), Helmholtz Zentrum München, Neuherberg, Germany

[+]contributed equally

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

Abstract

While the air microbiome and its diversity are essential for human health and ecosystem resilience, comprehensive air microbial diversity monitoring has remained rare, so that little is known about the air microbiome's composition, distribution, or functionality. Here we show that nanopore sequencing-based metagenomics can robustly assess the air microbiome in combination with active air sampling through liquid impingement and tailored computational analysis. We provide fast and portable laboratory and computational approaches for air microbiome profiling, which we leverage to robustly assess the taxonomic composition of the core air microbiome of a controlled greenhouse environment and of a natural outdoor environment. We show that long-read sequencing can resolve species-level annotations and specific ecosystem functions through *de novo* metagenomic assemblies despite the low amount of fragmented DNA used as an input for nanopore sequencing. We then apply our pipeline to assess the diversity and variability of an urban air microbiome, using Barcelona, Spain, as an example; this randomized experiment gives first insights into the presence of highly stable location-specific air microbiomes within the city's boundaries, and showcases the robust microbial assessments that can be achieved through automatable, fast, and portable nanopore sequencing technology.

Introduction

The air microbiome encompasses a broad spectrum of bioaerosols, including bacteria, archaea, fungi, viruses, bacterial endotoxins, mycotoxins, and pollen [1]. While its pivotal functions for human health and ecosystem resilience are recognized, little is known about its composition, distribution, and functionality [2]. Past research efforts, particularly those driven by infectious diseases such as COVID-19 and tuberculosis, have shifted the research focus towards potentially pathogenic microbial taxa; however, exposure to a diverse air microbiome has also been increasingly considered as a health-promoting factor, underscoring the need for holistic air microbial diversity monitoring [3].

Most genetics-based air microbiome studies have employed targeted sequencing via metabarcoding due to the low biomass of bioaerosols [1, 4]. While metabarcoding increases the sensitivity of taxonomic detection, it is inherently limited by amplification biases and incomplete databases. In contrast, metagenomics, which is based on shotgun sequencing of native DNA, avoids amplification biases and allows for *de novo* reconstructions of microbial genomes for robust species identification and functional annotation. Such metagenomic approaches have also recently been applied for low biomass bioaerosol analysis [5] and have revealed the complex nature and diverse origins of the air microbiome [4], including vertical-

altitudinal stratification of microbial abundance and distribution [6], and substantial diurnal, seasonal, temperature-, and humidity-dependent fluctuations [7].

These metagenomic assessments of the air microbiome have thus far relied on short-read sequencing technology, which provides accurate sequencing data but hampers *de novo* assemblies, especially of highly repetitive genomic regions, and accurate species- or strain-level identification due to the inherently short sequencing reads; long-read sequencing, on the other hand, has facilitated *de novo* genome assemblies [8] and assessments of highly repetitive genomic regions, including the detection of antimicrobial resistance genes [9], from metagenomic data. Especially recent advances in nanopore sequencing technology have made long-read sequencing increasingly relevant for microbial diversity assessments due to the technology's substantially improving sequencing accuracy [10, 11] while maintaining its long-read sequencing capacity and its automatable [12], fast, and portable deployability for applications in clinical [13] or remote settings [14]. While nanopore sequencing has been used to characterize the microbial diversity of various environments, such as of freshwater [15] and dust [16], no approaches have yet been established to leverage the technology's unique advantages for monitoring the taxonomic and functional diversity of the air microbiome.

Here, we established laboratory and computational approaches to enable robust air microbiome profiling through nanopore metagenomics. We first evaluated the suitability of long-read shotgun sequencing for assessing the air microbiome in a controlled indoor environment, and then applied our approaches to an outdoor environment for validation. We showed 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 human health consequences. We finally applied our laboratory and computational approaches to conduct a randomized air sampling campaign in Barcelona, Spain, to robustly describe its urban air microbiome.

Materials and Methods

We first conducted preliminary tests to compare standard air sampling and DNA extraction approaches for nanopore sequencing-based air metagenomics; this included the testing of standard quartz filter- and liquid impingement-based air samplers and the optimization of respective DNA extraction approaches for subsequent nanopore shotgun sequencing, which relies on minimum DNA input without nucleotide amplification and is sensitive to native DNA contamination (**Supplementary Information**: *Air sampling and DNA extraction optimizations*).

Based on these preliminary tests, we decided to use the Coriolis µ liquid impinger (Bertin Instruments, France; (**Supplementary Information**: *Air sampling and DNA extraction optimizations*) for air sampling, which uses cyclonic forces to concentrate airborne biomass into a collection liquid in a cone. We used 15 mL of ultrapure water with 0.005 % Triton-X (Sigma-Aldrich, Germany) as collection liquid, which functions as a nonionic surfactant to enhance organic compound solubility and surface enlargement due to foam generation. The liquid impinger was positioned at 1.5 m above the ground to sample air within the human breathable zone, which ranges from 1.4 to 1.8 m. We operated the liquid impinger at an air flow rate of 300 L min^{-1} and at a collection liquid refilling rate of 0.8 mL min^{-1} to counter liquid evaporation during sampling. After sampling, we directly transferred the collected liquid into a sterile 15 mL falcon tube. We then divided the liquid across three 5 mL tubes, centrifuged them at 18 000 x g for 25 min, and collected the pellets. The pellets were resuspended, aggregated, and subsequently centrifuged twice at 18 000 x g for 25 min while discarding the supernatant.

We first sampled air in a greenhouse ("Gh"; Helmholtz Munich Environmental Research Unit) as a controlled environment with moderate human activity and continuous air circulation (mean ambient temperature of 23 °C); we sampled air for three consecutive days, either for 1h in three consecutive replicates per day or for 3h with one replicate per day (**Supplementary Table 1**). We next sampled air in a natural environment ("Nat"),

namely on the Helmholtz Munich campus on the outskirts of Munich (48.220889, 11.597028), which is mainly surrounded by natural grassland. We sampled for six consecutive days, following an alternating pattern of 3h or 6h of air sampling; we here tested 6h as sampling duration since we expected a higher variability in the air microbiome in comparison to the controlled greenhouse setting (**Supplementary Table 1**). The liquid impinger was positioned in a shaded area to avoid significant thermal fluctuations. While the weather remained relatively constant and sunny across the six sampling days (ambient temperature ranged from 21°C to 25°C, and humidity from 42% to 71%.), we note that the 6h-sample from day 4 was affected by rain and thunderstorm at the end of the sampling activity. We finally collected urban air samples in Barcelona, Spain, from October 16th to November 3rd, 2023. We sampled five different urban locations: Gracia ("Residential Area", 41.398861, 2.153490), Eixample ("City Center", 41.385500, 2.155103), Poblenou ("Urban Beach", 41.404135, 2.206550), Vall d'Hebron ("Outer Belt", 41.425887, 2.148349), and Observatori Fabra ("Green Belt", 41.419772, 2.122447). We conducted randomized sampling in terms of timing (morning *versus* afternoon) and across days; each location was sampled three times for 3h using two Coriolis µ air samplers, respectively, resulting in altogether 30 air samples (**Supplementary Table 1**).

Based on our preliminary tests, we further decided to use the spin-column based PowerSoil Pro Kit (QIAGEN, 2018, Hilden,

Germany) for DNA extractions, using 30 µL of elution buffer (**Supplementary Information**: *Air sampling and DNA extraction optimizations*). Final DNA concentration was measured on a Qubit 4.0 fluorometer (Invitrogen, 2021), using the high-sensitivity DNA kit and 3 µL of DNA elution as input per sample. We then used the Rapid Barcoding library preparation kit (RBK114-24 V14), R10.4.1 MinION flow cells, and MinKNOW by Oxford Nanopore Technologies (Oxford, UK) to nanopore shotgun-sequence the extracted DNA of the air samples. During library preparation, we used each barcode twice per air sample to increase the DNA input per sample. For sequencing the samples of the controlled and natural environment, we used one R10.4.1 flow cell per sample type (i.e., for all 1h-, 3h-, or 6-samples and replicates, respectively). For sequencing the samples of the urban environment, we pooled all samples from the Outer Belt location onto one flow cell (since they exhibited the lowest DNA concentrations), and the samples of the City Center and Residential Area, as well as of the Green Belt and Urban Beach, onto one flow cell, respectively. The sequencing parameters included a minimum read length of 20 bases, a translocation speed of 400 bases per second, and each sequencing run lasted 24 hours. As we used MinKNOW v23.04.3 for the controlled and natural environment, this sequencing data was generated at a signal measurement frequency of 4 kHz, whereas we used the updated MinKNOW v23.04.5 for the urban environment, which generated sequencing data at 5 kHz.

We included negative controls along our entire protocol to identify contamination of the low-biomass air samples. For sampling negative controls, we treated one liquid impinger cone per sampling event the same way that we treated the actual sampling cone, but we only left them in the impinger for a few minutes and did not actively sample air. For the urban environment, negative sampling controls were collected once per sampling day and sampling location. For DNA extraction and sequencing negative controls, we included one sample of 700 µL nuclease-free water (Thermo Fisher Scientific) per DNA extraction and one sample of 20 µL nuclease-free water per library preparation, respectively. We barcoded all negative controls, i.e. sampling, extraction, and sequencing controls, and included them in the same sequencing library as the respective control samples. We further subjected a positive control of five Gram-positive bacteria, three Gram-negative bacteria, and two fungal species (ZymoBIOMICS Microbial Community Standard, D6300) to our DNA extraction and sequencing protocols to assess any potential biases. The positive control was sequenced on a separate flow cell since the high DNA concentration would have outcompeted the low-biomass air samples.

We next used Guppy v6.3.2 (r10.4.1_e8.2_400bps_hac; [17]) in high-accuracy (HAC) mode for basecalling the controlled and natural environment samples, and Dorado v4.3.0 (dna_r10.4.1_e8.2_400bps_hac@v4.3.0; [18]) for HAC-basecalling of the urban environment samples. We only

processed the data that had passed internal data quality thresholds during sequencing (“passed” sequencing reads). Porechop v0.2.3 [19] was used for removing sequencing adapters and barcodes, and Nanofilt v2.8.0 [20] was applied for filtering reads at a minimum average quality score of 8 and a minimum length of 100 bases for all samples. We then used Kraken2 v2.0.7 [21] with the NCBI nt database (access 29.01.2023) for taxonomic classification across all samples, and downsampled them to a specific read count for comparable taxonomic assessments across samples of one sample type: 5k reads for 1h-samples from the controlled environment, 15k reads for the 3h-samples from the controlled environment, 70k reads for the natural environment samples, and 30k reads for the urban environment samples. We performed Principal Coordinate Analysis (PCoA) on the relative abundances of the genera identified in the urban environment samples, which were downsampled to 30k read, using Python v3.9 with Pandas v1.3.3, NumPy v1.21.2, scikit-learn v0.24.2, scikit-bio v0.5.6, SciPy v1.7.1, and Matplotlib v3.5.2.. The 20 most abundant microbial genera at a minimum relative abundance of 1% as well as the PCoA were visualized using matplotlib v3.5.2 in Python v3.9. We additionally benchmarked several additional bioinformatic analysis tool in application to the controlled and natural environment samples, including DIAMOND BLASTX [22] for protein-based taxonomic classifications and the Chan-Zuckerberg (CZID) computational pipeline [23] for hybrid

taxonomic classifications (i.e., as a combination of read- and contig-based classification).

We generated *de novo* assemblies using metaflye v2.9.1 [24], followed by polishing with minimap2 v2.17 [25] and three rounds of Racon v1.5 [26]. The resulting contigs were then binned into Metagenome-Assembled Genomes (MAGs) using metaWRAP v1.3 [27], which integrates the output of various binning tools. The MAGs were refined and quality-checked using CheckM v1.2.2 [28]. We only maintained MAGs at minimum completeness of 30% and maximum contamination of 10%. For the urban microbiome dataset, we pooled across all samples per sampling location to maximize the number of reads before binning. We finally applied functional annotation to our metagenomic dataset to assess the presence of general metabolic pathways and ecosystem functions (**Supplementary Information**: *Functional annotation*); to identify antimicrobial resistance and virulence genes, we applied AMRFinderPlus v3.12.8 [29] and ABRicate v1.0.1 [30] to the reads, contigs, and bins; for the application to the read level, we converted the fastq files to fasta files using seqkit v2.8.2 [31].

To obtain information about the anthropogenic impact on the different urban sampling locations, we obtained remote sensing data (Sentinel-2 L1C orthoimage products from October 24th 2023) that provides top-of-atmosphere reflectance, which we used to classify the city of Barcelona into Local Climate Zones (LCZs) on based ten bands with 10 m and 20 m ground

sampling distances [32]. We further used the portable aerosol spectrometer Dust Decoder 11-D (GRIMM Aerosol Technik GmbH, Germany) to monitor particle mass fractions (TSP, PM_{10}, and $PM_{2.5}$; TSP=total suspended particles; PM=particulate matter) as well as temperature and relative humidity measurements in 1-minute intervals during each sampling event. We then summarized and analyzed the resulting data using Python v3.9 and SciPy v1.13.0: We applied the Kruskal-Wallis and post-hoc Dunn's tests to identify significant environmental differences between locations, and conducted regression analyses to assess correlations between particle mass fractions and microbial diversity indices (Shannon, Simpson, and richness of microbial genera).

Results

After confirming that Coriolis µ liquid impingement resulted in sufficient high-quality DNA yield for nanopore shotgun sequencing after one hour of sampling (**Materials and Methods**; **Supplementary Information**: *Air sampling and DNA extraction optimizations*), we conducted a pilot study in a controlled environment to determine the robustness of the metagenomic data and assess the impact of sampling duration (**Materials and Methods**). For the 1h-samples, DNA yields ranged from 17.7 ng to 50.7 ng (0.98 ng/m^3 to 2.82 ng/m^3), while the 3-hour samples showed DNA yields ranging from 130.2 ng to 179.4 ng (2.41 ng/m^3 to 3.32 ng/m^3; **Supplementary Table 1**;

pilot_study sheet). Nanopore shotgun sequencing delivered between 7k and 60k high-quality sequencing read at a median read length of 896 bases (**Figure 1A**), respectively, of which 5k to 35k reads were successfully mapped to the taxonomic genus level using Kraken2 and the NCBI nt database (**Figure 1B-C**; **Supplementary Table 1**; *pilot_study* sheet). After downsampling to the same number of reads per sample type (1h- and 3h-samples, respectively), the taxonomic composition of the 20 most abundant taxa indicated that only the 3-hour sampling duration captured a stable "core" air microbiome across days at the genus level (**Figure 1D-E**). These assessments were consistent for protein-level or hybrid read- and assembly-based methods, both at the taxonomic phylum and genus level (**Supplementary Figures 1-2)**. The most abundant genera included soil- and plant-associated bacteria such as *Bradyrhizobium, Paracoccus, Nocardioides, Massilia,* and *Streptomyces* (**Figure 1D-E**; **Materials and Methods**).

Based on these results, we conducted a pilot study in a natural environment over six days; we sampled air for either 3h or 6h, assuming that the natural environment might show more variability than the controlled environment and require longer sampling duration. Briefly, while the extended sampling duration increased total DNA yield, it did not consistently increase the amount of biomass per cubic meter of sampled air, suggesting diminishing returns in efficiency with longer durations (**Supplementary Table 1**; *pilot_study* sheet). Nanopore shotgun sequencing resulted 130k to 200k high-quality

sequencing reads at a slightly higher median read length than the controlled environment of 1481 (**Figure 1F**), of which 70k to 140k reads were successfully mapped to the taxonomic genus level (**Figure 1G-H**; **Supplementary Table 1**; *pilot_study* sheet). After downsampling all samples to 70k reads, analysis of the relative abundance of 20 most abundant taxa revealed a very similar profile for both 3-hour and 6-hour samples. The taxonomic assignments were again consistent across protein-level or hybrid read- and assembly-based methods, both at the taxonomic phylum and genus level (**Supplementary Figures 1-2)**. A distinct air microbiome profile was observed in the natural environment in comparison to the controlled settings, with high predominance of *Pseudomonas* and unique detection of microbial taxa such as *Actinoplanes, Amycolatopsis, Dugnaella, Flavobacterium, Nocardia, Rhodococcus,* and *Variovorax* (**Figure 1I-J**; **Materials and Methods**).

All negative controls resulted in low DNA yields (of <0.1 ng) from typical contaminant species such as *Escherichia*, *Salmonella*, *Shigella*, *Francisella*, and *Pseudomonas* (**Supplementary Figure 3A-B**; **Material and Methods**) [33]. This demonstrates that no external contamination had influenced our assessment of air as a low-biomass ecosystem, thus underscoring the reliability of the presented results. 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. Although the fungal taxa and Gram-positive *Bacillus subtilis*, in particular,

were underrepresented (**Supplementary Figure 3C**; **Material and Methods**).

We finally applied our optimized laboratory and computational approaches to assess an exemplary urban microbiome using nanopore metagenomics (**Figure 2A**; *left*; **Materials and Methods**). Our remote-sensing-based LCZ classification (**Figure 2A**; *right*) indicated that most of our sampling locations (City Center, Residential Area, and Urban Beach) were of the compact low-rise category, a typical feature of central urban environments. The Outer Belt location was classified as compact mid-rise category, which features taller buildings at the outskirts of the city. The Green Belt location was classified as scattered trees category, featuring more natural elements. In terms of air pollution assessed through particle mass fractions (**Supplementary Table 2**; **Materials and Methods**), we found significant differences in TSP, PM_{10} and $PM_{2.5}$, between our sampling locations (**Supplementary Figure 4**). The total air pollution measured by TSP was highest in the three compact low-rise sampling locations, while TSP was lowest in the Outer Belt. The relatively medium levels of TSP in the Green Belt were dominated by relatively high levels of PM_{10} (**Supplementary Figure 4**).

Nanopore shotgun sequencing delivered between 33k and 422k high-quality sequencing read at a median read length of between 598 and 2358 bases (**Figure 2B**), respectively, of which 21k to 312k reads were successfully mapped to the

taxonomic genus level using Kraken2 and the NCBI nt database (**Figure 2C**; **Supplementary Table 1**; *urban_study* sheet). The City Center exhibited the longest DNA fragments, and the Outer Belt location the shortest DNA fragments (**Figure 2B**). The relatively high fragmentation in the Outer Belt coincided with generally low DNA yields across all the location's samples and replicates (**Supplementary Table 1**; *urban_study* sheet).

For taxonomic comparisons across replicates and samples, we again downsampled the number of reads (here to 30k reads per sample) and compared the relative distribution of the 20 most abundant microbial genera per location at a minimum relative abundance cutoff of 1% displaying (**Materials and Methods**). We observed that the microbial compositions were highly location-specific across all six samples per location, including across the three randomized sampling events and the two respective sampling replicates (**Figure 2D**; **Materials and Methods**). The core urban air microbiome consisted of microbial genera such as *Streptomyces, Sphingomonas, Pseudomonas, Nocardioides*, and *Microbacterium*, which were detected across all samples. Specifically the Green Belt was characterized by the presence of several unique taxa such as *Rubrobacter, Gemmatirosa, Capillimicrobium*, and *Amycolatopsis*, whereas dominant "urban" taxa such as *Paracoccus, Kocuria, Deinociccus*, and *Cellulomonas* were not detected at all (**Figure 2D**). Principal Coordinate Analysis (PCoA) clearly distinguishes the five different urban locations, with the first PCoA axis separating the Green Belt and City

Center locations from the remaining ones; the second PCoA axis then further delineates the individual sampling locations (**Figure 2E**).

Despite the location-specific differences in air microbial composition (**Figure 2A**), in LCZ-based land usage (**Figure 2D**) and in air pollution measured by particle mass fractions (**Supplementary Figure 4**), we found no significant correlations between any environmental variable and microbial diversity measurements (**Materials and Methods**).

To next obtain as highly contiguous *de novo* genome assemblies as possible, we pooled all samples per location before contig assembly and binning (**Materials and Methods**). Taxonomic classification of these bins showed that only the most abundant taxa could be assembled (**Table 1**). Functional annotation of the reads, contigs, and bins detected typical microbial metabolic functions (**Supplementary Information**: *Functional annotation*). We next focused on the annotation of antimicrobial resistance and virulence genes with potential human health consequences (**Supplementary Table 3**; **Materials and Methods**). One of the most frequently detected genes was the *VanR-O* gene, which is responsible for vancomycin resistance. When comparing resistance gene prevalence across urban locations, the Urban Beach location exhibited the highest density of resistance genes; the *blaCARB-8* and *blaCARB-16* genes, which confer beta-lactam resistance, and the *blaOXA-17* gene, which confers oxacillin resistance,

were detected at the read level. Additionally, *blaL1*, which confers to a broad range of beta-lactam antibiotics, the *blaOXY* gene, which confers oxacillin resistance, and the *blaPSZ* gene, which confers resistance to penicillins and cephalosporins, were identified at the contig level (**Supplementary Table 3**).

Discussion

Metagenomic approaches have provided unprecedented insights into the nature, origin, and complexity of the air microbiome [4, 6, 7, 34]. While past studies have relied on traditional short-read sequencing, we here describe the first long-read nanopore sequencing technology-based approaches to robustly assess the air microbiome. Although nanopore sequencing has been applied to various environmental samples, such as water and soil [15, 35, 36], its applicability to air samples was expected to pose a particular challenge due to the ultra-low biomass of air and the amplification-free nature of nanopore sequencing [5]. We here showed that nanopore shotgun sequencing in combination with active air sampling through liquid impingement and tailored computational analyses can reproducibly describe the air microbiome of different environments (**Figure 1**) while leveraging the latest nanopore chemistry improvements which offer high sequencing accuracy and reduced minimum DNA input requirements [10, 11].

We further showed that only three hours of active air sampling resulted in robust air microbiome assessments in a controlled and natural environment, with consecutive application of our

laboratory and computational approaches to the urban air microbiome in Barcelona, Spain, revealing surprisingly stable location-specific signatures of microbial composition and diversity (**Figure 2**). These stable signatures could importantly be identified across replicates (using two air samplers per sampling event) and despite stringent randomization across sampling days and morning and afternoon sampling events. Several microbial taxa such as *Sphingomonas* and *Streptomyces*, which are known for their evolutionary adaptability, were nevertheless present in all air microbiomes, and could potentially be part of the stable air microbiome of this urban environment. Ordination of the taxonomic composition was able to capture the majority of variance in this multidimensional data (>80%; **Figure 2E**) and nicely visualizes the distinct clusters that separate each urban location and specifically the Green Belt and City Center locations from the remaining ones. The relative similarity of Green Belt and City Center samples might be attributable to the phenomenon of orographic uplift, where air masses ascend from lower regions (here the Barcelona City Center) to higher elevated areas (here the closeby Green Belt). As a result of this upward movement, certain airborne particles and microorganisms might have been transported from the City Center to the Green Belt location [37, 38].

The individual samples of the Green Belt location cluster together most tightly (**Figure 2E**). be because of several microbial taxa that were uniquely detected at this location, which

represents the only natural environment in our study according to our remote-sensing-based assessments; those unique taxa are known to be associated with soil or have been frequently found in forests and green spaces [39]. Besides this finding, we however found no evidence of correlation of the urban air microbiome with measurements of anthropogenic impact (as assessed through the remote-sensing-based Local Climate Zones, LCSz; **Figure 2A**) or of air pollution (as assessed through particle mass fraction measurements; **Supplementary Figure 4**). This might be due to complex interactions between air microbiomes, as exemplified by our hypothesis of the impact of orographic uplift, or because of lack of depth when describing our environmental variables. For example, air pollution by TSP was higher in the Green Belt than in the Outer Belt, which would have not been expected according to the remote-sensing-based anthropogenic impact inferences. However, these elevated levels of TSP in the Green Belt might have originated from natural air components such as pollen, which would require more in-depth environmental monitoring to dissect.

The annotation of antimicrobial resistance and virulence genes in our metagenomic data shows that we can use the same dataset to assess potential anthropogenic impacts on microbial diversity while concurrently understanding potential public health consequences [40]. We detected evidence of antimicrobial resistance across all sampled environments (**Supplementary Table 3**), but especially the detections of clinically relevant beta-lactamases such as *blaCARB-8*,

blaOXA-1, and *blal-1*, and of genes conferring resistance to other antibiotics such as carbenicillin and oxacillin [41], in Barcelona's urban air microbiome underscore the possibility of monitoring airborne virulence dissemination using nanopore-based metagenomics.

Genome assembly and binning of the long nanopore reads further allows us to be more confident in the presence of specific microbial species and of their pathogenic potential through the identification of Metagenome-Assembled Genomes (MAGs) (**Table 1**). We obtained high-quality genome assemblies (**Materials and Methods)** of the pathogenic species *Stenotrophomonas maltophilia* and *Salmonella enterica* from the urban microbiome data (**Table 1**). The *Stenotrophomonas* species is known as an emerging difficult-to-treat human pathogen [42]and many of the *Salmonella enterica* serovars can cause disease in humans through zoonotic or foodborne transmission [43]. While we require good coverage of a microbial genome to create such assemblies for taxonomic species or strain identification, also just the presence of individual pathogen-associated sequencing reads might be used for obtaining first information on the potential presence of microorganisms of public health concern. For example, given the presence of sequencing reads of the Brucella genus, an animal pathogen that can affect dogs, in several of our urban air samples, we further analyzed our taxonomic annotation, which was based on the entire NCBI nt database, and were indeed able to detect the presence of Canis lupus familiaris in the same

551 air samples [44]. While this might point to a potential impact of
552 animal domestication and specifically frequent dog walking in
553 Barcelona on public health [45], such complex
554 interdependencies can only be confirmed in a controlled and/or
555 experimental setting.

556 While we were able to build *de novo* assemblies from our
557 nanopore-based air metagenomic data, most of the MAGs were
558 incomplete (<30%) and/or showed high levels of contamination
559 (>10%) (**Table 1**). Given the low amount of DNA input and
560 therefore relatively small size of the resulting metagenomic
561 datasets in combination with the expectedly high fragmentation
562 of DNA in air samples, this might just be an inherent
563 shortcoming when it comes to assessing the air microbiome
564 – albeit applying long-read sequencing technology. We here
565 found a particular small median DNA fragment and sequencing
566 read length for the Outer Belt location (**Figure 2B**), which might
567 point towards the impact of environmental conditions or specific
568 taxonomic compositions (and variables such as the
569 microorganisms' genome size and cell wall composition) on the
570 final fragment and read length distribution. It is further expected
571 that non-viable microorganisms, which might significantly
572 contribute to the air microbiome, result in more fragmented DNA
573 in the air samples; this means that substantial differences in
574 read lengths between microbial taxa might also be attributed to
575 their differential viability in the air environment – a hypothesis
576 that we might be able to resolve in the future using viability-
577 resolved metagenomic approaches [46].

We emphasize that our sampling, laboratory and computational approaches constitute one feasible and reproducible way of using nanopore shotgun sequencing to profile the air microbiome. While we tested some additional established air sampling and DNA extraction methodology, we have not conducted an extensive study of all possible approaches. We specifically emphasize that the detection of fungi and Gram-positive bacteria could be improved when using different sample processing and DNA extraction techniques. This is also reflected by the application of our approaches to a positive control, which shows that fungal taxa and Gram-positive *Bacillus subtilis*, in particular, were underrepresented. As sturdier cell walls would require more aggressive DNA extraction approaches, this would, however, also lead to increased DNA fragmentation, especially in Gram-negative bacteria, and therefore more difficult downstream analyses. A good trade-off could be the sequencing of several, differently processed DNA extracts and subsequent data pooling to assess the microbial diversity of any air sample more holistically.

In conclusion, our study establishes a robust framework for air microbiome assessments using nanopore metagenomics. We envision that nanopore sequencing for air monitoring can provide a basis for fast, robust, and automated characterizations of the air microbiome in both urbanized and remote settings. This characterization importantly extends beyond taxonomic composition to include functions related to human and ecosystem health, such as pathogen and drug resistance and

virulence gene detection, which can enhance our understanding of infectious disease transmission patterns and their relationship with exerted anthropogenic pressures.

Data and code availability

All raw data and MAGs have been made publicly available via ENA (study accession number: PRJEB76446). All code has been made publicly available via Github: https://github.com/ttmgr/Air_Metagenomics.

Acknowledgements

Author contributions

T.R, S.P., S.B. and L.U. designed the experiment. T.R, S.P. conducted the sampling, processing, and data analysis under L.U.'s supervision. T.R, S.P., and L.U. wrote the manuscript with the contribution of all authors who revised and approved the manuscript.

Financial Disclosure

This study was funded by a Helmholtz Principal Investigator Grant awarded to L.U.

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 (eds). *Advances in Ecological Research*. 2022. Academic Press, pp 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**: 661–664.

4. Naumova NB, Kabilov MR. About the Biodiversity of the Air Microbiome. *Acta Naturae* 2022; **14**: 50–56.

5. 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**: 1–11.

6. 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. *Proceedings of the National Academy of Sciences* 2022; **119**: e2117293119.

7. 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**: 23299–23308.

8. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore

sequencing. *Nat Biotechnol* 2020; **38**: 701–707.

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**: 444.

10. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, Wollenberg RD, et al. 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. *Nat Methods* 2022; **19**: 823–826.

11. Hall MB, Wick RR, Judd LM, Nguyen ANT, Steinig EJ, Xie O, et al. Benchmarking reveals superiority of deep learning variant callers on bacterial nanopore sequence data. 2024. bioRxiv. , 2024.03.15.585313

12. Raymond-Bouchard I, Maggiori C, Brennan L, Altshuler I, Manchado JM, Parro V, et al. Assessment of Automated Nucleic Acid Extraction Systems in Combination with MinION Sequencing As Potential Tools for the Detection of Microbial Biosignatures. *Astrobiology* 2022; **22**: 87–103.

13. Sauerborn E, Corredor C, Reska T, Perlas A, Atum SV da F, Goldman N, et al. Detection of hidden antibiotic resistance through real-time genomics. 2023.

14. 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**: 228–232.

15. Urban L, Holzer A, Baronas JJ, Hall MB, Braeuninger-Weimer P, Scherm MJ, et al. Freshwater monitoring by nanopore sequencing. *eLife* 2021; **10**: e61504.

16. 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. *Sci Rep* 2020; **10**: 3209.

17. Guppy protocol - Guppy software overview. *Oxford Nanopore Technologies*. https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb_2003_v1_revax_14dec2018. Accessed 13 Jun 2024.

18. nanoporetech/dorado: Oxford Nanopore's Basecaller. *GitHub*. https://github.com/nanoporetech/dorado. Accessed 10 Jun 2024.

19. Wick R. rrwick/Porechop. 2024.

20. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 2018; **34**: 2666–2669.

21. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019; **20**: 1–13.

22. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015; **12**: 59–60.

23. Kalantar KL, Carvalho T, de Bourcy CFA, Dimitrov B, Dingle G, Egger R, et al. IDseq—An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *GigaScience* 2020; **9**: giaa111.

24. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods* 2020; **17**: 1103–1110.

25. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018; **34**: 3094–3100.

26. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research* 2017; **27**: 737.

27. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 2018; **6**: 1–13.

28. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015; **25**: 1043–1055.

29. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep* 2021; **11**: 1–9.

30. tseemann/abricate: :mag_right: Mass screening of contigs for antimicrobial and virulence genes. *GitHub*. https://github.com/tseemann/abricate. Accessed 10 Jun 2024.

31. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS ONE* 2016; **11**.

732 32. Rosentreter J, Hagensieker R, Waske B. Towards large-
733 scale mapping of local climate zones using multitemporal
734 Sentinel 2 data and convolutional neural networks. *Remote*
735 *Sensing of Environment* 2020; **237**: 111472.
736 33. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO,
737 Moffatt MF, et al. Reagent and laboratory contamination can
738 critically impact sequence-based microbiome analyses. *BMC*
739 *Biol* 2014; **12**: 87.
740 34. Luhung I, Uchida A, Lim SBY, Gaultier NE, Kee C, Lau
741 KJX, et al. Experimental parameters defining ultra-low biomass
742 bioaerosol analysis. *npj Biofilms Microbiomes* 2021; **7**: 1–11.
743 35. Cummings PJ, Olszewicz J, Obom KM. Nanopore DNA
744 Sequencing for Metagenomic Soil Analysis. *J Vis Exp* 2017;
745 55979.
746 36. Urban L, Miller AK, Eason D, Vercoe D, Shaffer M,
747 Wilkinson SP, et al. Non-invasive real-time genomic monitoring
748 of the critically endangered kākāpō. *eLife* 2023; **12**: RP84553.
749 37. Ribeiro I, Martilli A, Falls M, Zonato A, Villalba G. Highly
750 resolved WRF-BEP/BEM simulations over Barcelona urban
751 area with LCZ. *Atmospheric Research* 2021; **248**: 105220.
752 38. Segura R, Badia A, Ventura S, Gilabert J, Martilli A,
753 Villalba G. Sensitivity study of PBL schemes and soil
754 initialization using the WRF-BEP-BEM model over a
755 Mediterranean coastal city. *Urban Climate* 2021; **39**: 100982.
756 39. Lladó S, López-Mondéjar R, Baldrian P. Forest Soil
757 Bacteria: Diversity, Involvement in Ecosystem Processes, and
758 Response to Global Change. *Microbiol Mol Biol Rev* 2017; **81**:

759 e00063-16.
760 40. Urban L, Perlas A, Francino O, Martí-Carreras J, Muga
761 BA, Mwangi JW, et al. Real-time genomics for One Health. *Mol*
762 *Syst Biol* 2023; **19**: e11686.
763 41. Zhuang M, Achmon Y, Cao Y, Liang X, Chen L, Wang
764 H, et al. Distribution of antibiotic resistance genes in the
765 environment. *Environmental Pollution* 2021; **285**: 117402.
766 42. Lipuma JJ, Currie BJ, Peacock SJ, Vandamme PAR.
767 Burkholderia, Stenotrophomonas, Ralstonia, Cupriavidus,
768 Pandoraea, Brevundimonas, Comamonas, Delftia, and
769 Acidovorax. *Manual of Clinical Microbiology*. 2015. John Wiley
770 & Sons, Ltd, pp 791–812.
771 43. Andino A, Hanning I. Salmonella enterica: Survival,
772 Colonization, and Virulence Differences among Serovars.
773 *ScientificWorldJournal* 2015; **2015**: 520179.
774 44. Pinn-Woodcock T, Frye E, Guarino C, Franklin-Guild R,
775 Newman AP, Bennett J, et al. A one-health review on brucellosis
776 in the United States. *Journal of the American Veterinary Medical*
777 *Association* 2023; **261**: 451–462.
778 45. Hensel ME, Negron M, Arenas-Gamboa AM. Brucellosis
779 in Dogs and Public Health Risk. *Emerg Infect Dis* 2018; **24**:
780 1401–1406.
781 46. Urel H, Benassou S, Reska T, Marti H, Rayo E, Martin
782 EJ, et al. Nanopore- and AI-empowered metagenomic viability
783 inference. 2024. bioRxiv. , 2024.06.10.598221

List of Tables

Table 1. *De novo* genome assembly results across all air samples from the controlled (Gh), natural (Nat) environment, and the urban microbiome dataset. Contigs were assembled and then used to identify metagenome-assemblies (MAGs), their taxonomic origin, completeness, and contamination.

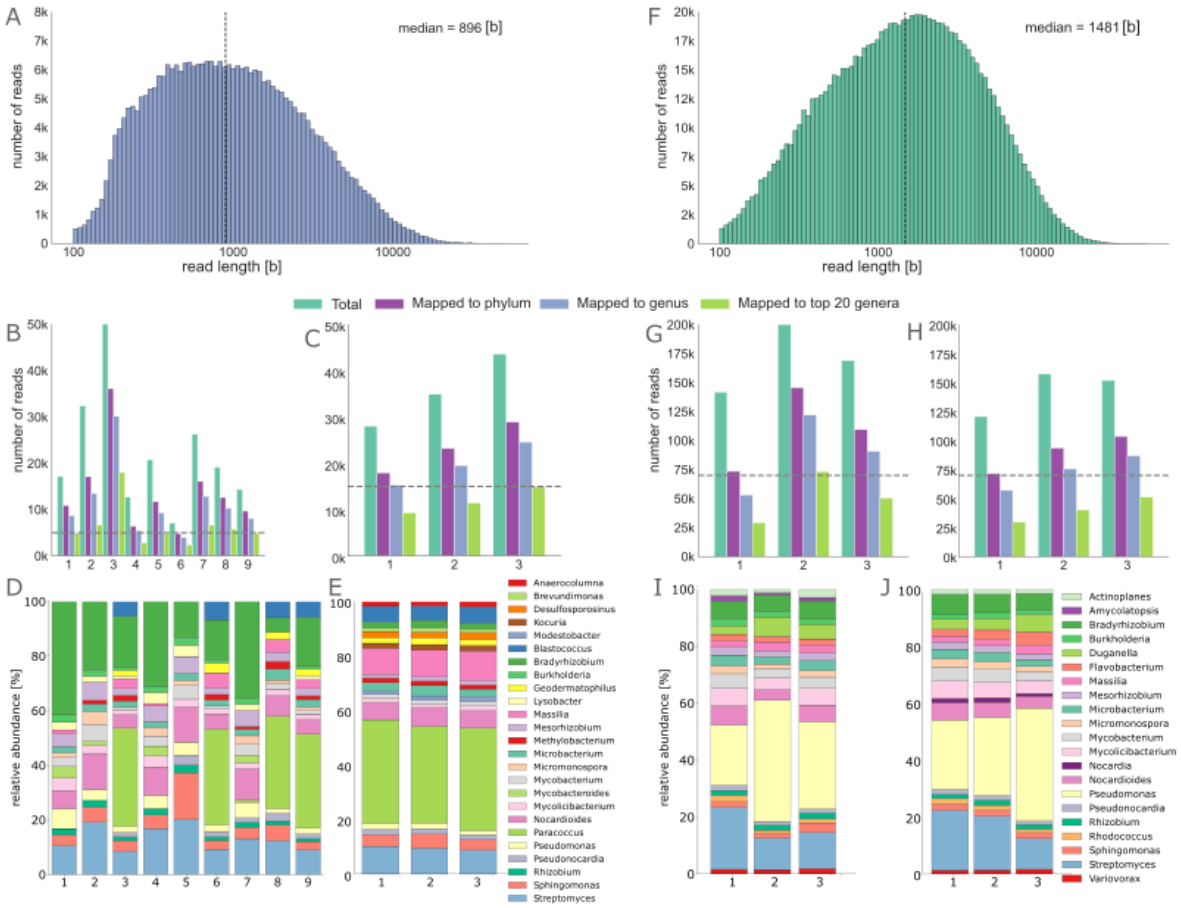
Sample	# contigs (mean)	N50 contigs (mean)	# MAGs	Species	Completeness [%]	Contamination [%]
Gh1h	21	5 928	/	/	/	/
GH3h	121	15 330	2	*Paracoccus aerius* *Paracoccus denitrificans*	64.59 63.41	2.94 1.46
Nat3h	204	7 401	/	/	/	/
Nat6h	117	7 282	/	/	/	/
City center	1170	23 151	/	/	/	/
Residential area	470	11 098	/	/	/	/
Green belt	1171	15 215	1	*Burkholderia sp.*	36.66	2.38
Urban beach	7732	21 049	1	*Stenotrophomonas maltophilia*	48.33	6.69
Outer belt	1874	10 282	1	*Salmonella enterica*	41.72	10.59

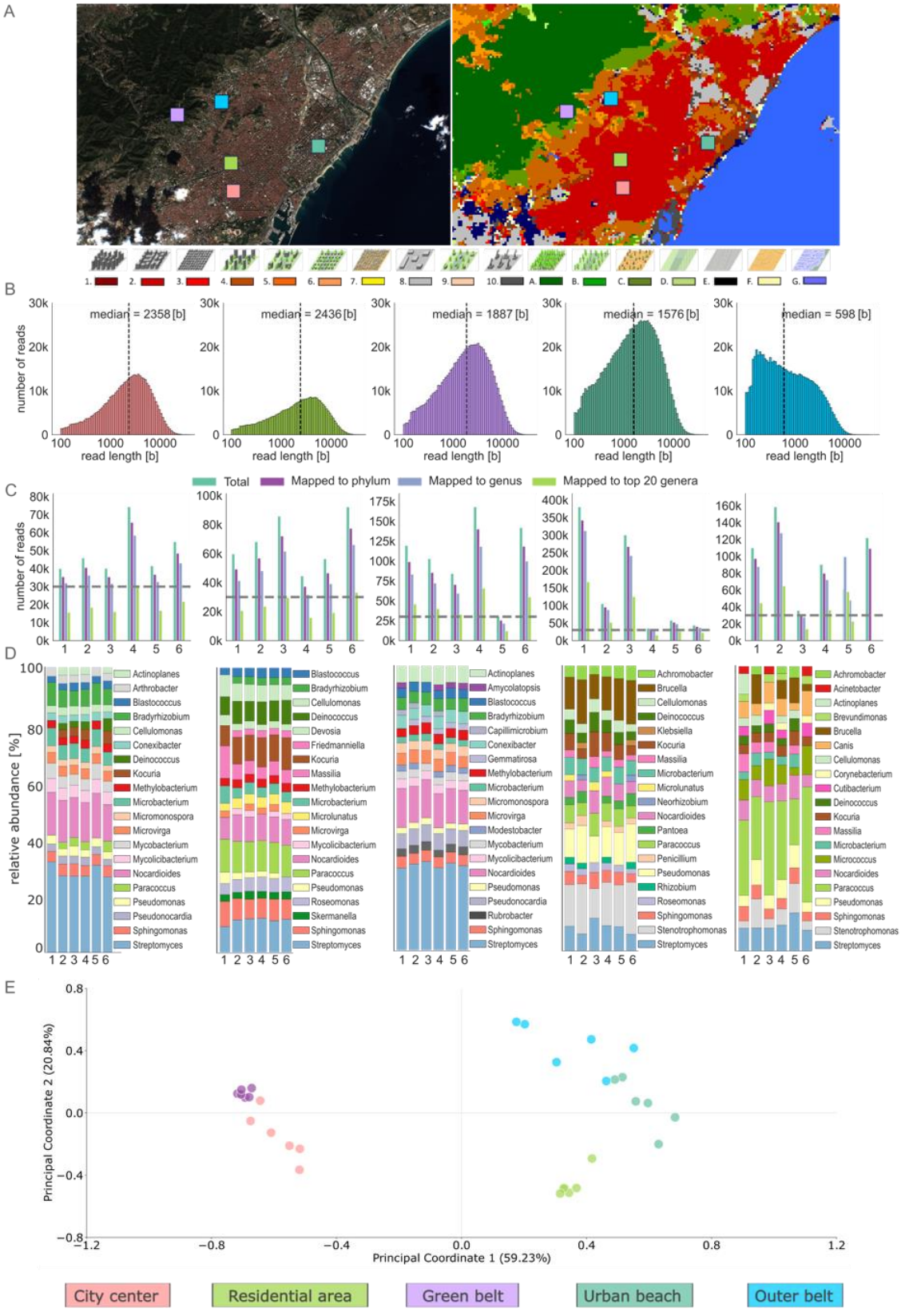
List of Figures

Figure 1. Robust air microbiome assessments of a controlled (*left*; **A-E**) and natural (*right*; **F-J**) environment through nanopore shotgun sequencing. **A.** Nanopore sequencing read length distribution across 1h- and 3-Gh samples. **B-C**. 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 samples; the downsampling threshold across samples is indicated by the dashed horizontal line. Taxonomic composition of the **D.** 1h- and **E.** 3h-samples after downsampling based on the 20 most dominant genera across samples. **F.** Nanopore sequencing read length distribution across 3h- and 6h-Nat samples. **G-H**. 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 samples; the downsampling threshold across samples is indicated by the dashed horizontal line. Taxonomic composition of the **I.** 3h- and **J.** 6h-samples after downsampling based on the 20 most dominant genera across samples.

Figure 2. Local climate zone (LCZ) classification and metagenomic analysis of Barcelona using satellite images and nanopore shotgun sequencing data. **A**. Sentinel-2 image and LCZs classification map of Barcelona on 24.10.2023, with a legend of LCZs classes at the bottom. The colored squares indicate the five sampling locations. The legend at the bottom depicts various LCZs represented by 3D models and their corresponding colors. (1-Compact high-rise; 2-Compact mid-rise; 3-Compact low-rise; 4-Open high-rise; 5-Open mid-rise; 6-Open low-rise; 7-Lightweight low-rise; 8-Large low-rise; 9-Sparsely built; 10-Heavy industry; A-Dense trees; B-Scattered trees; C-Bush scrub; D-Low plants; E-Bare rock or paved; F-Bare soil or sand; G-Water). **B.** Histograms showing the distribution of read lengths [b] for each sampling site, with the median read length indicated on the top of each histogram. **C.** Bar plots displaying the number of reads mapped at various taxonomic levels for each sample site, with the downsampling threshold indicated by the dashed horizontal line. **D.** Relative abundance of the top 20 most abundant bacterial

genera at the read level, downsampled to 30K reads before taxonomic classification using Kraken2. **E**. Principal Coordinates Analysis (PCoA) of the relative abundances of the bacterial genera identified at the five sampling locations.




A
B
median = 2358 [b]
median = 2436 [b]
median = 1887 [b]
median = 1576 [b]
median = 598 [b]
number of reads
read length [b]
C
Total
Mapped to phylum
Mapped to genus
Mapped to top 20 genera
number of reads
D
relative abundance [%]
Actinoplanes
Arthrobacter
Blastococcus
Bradyrhizobium
Cellulomonas
Conexibacter
Deinococcus
Kocuria
Methylobacterium
Microbacterium
Micromonospora
Microvirga
Mycobacterium
Mycolicibacterium
Nocardioides
Paracoccus
Pseudomonas
Pseudonocardia
Sphingomonas
Streptomyces
Bradyrhizobium
Cellulomonas
Deinococcus
Devosia
Friedmanniella
Kocuria
Massilia
Methylobacterium
Microbacterium
Microlunatus
Microvirga
Mycolicibacterium
Nocardioides
Paracoccus
Pseudomonas
Roseomonas
Skermanella
Sphingomonas
Streptomyces
Amycolatopsis
Capillimicrobium
Gemmatirosa
Modestobacter
Rubrobacter
Achromobacter
Brucella
Klebsiella
Neorhizobium
Pantoea
Penicillium
Rhizobium
Stenotrophomonas
Acinetobacter
Brevundimonas
Canis
Corynebacterium
Cutibacterium
Micrococcus
E
Principal Coordinate 2 (20.84%)
Principal Coordinate 1 (59.23%)
City center
Residential area
Green belt
Urban beach
Outer belt

829

Figure 1

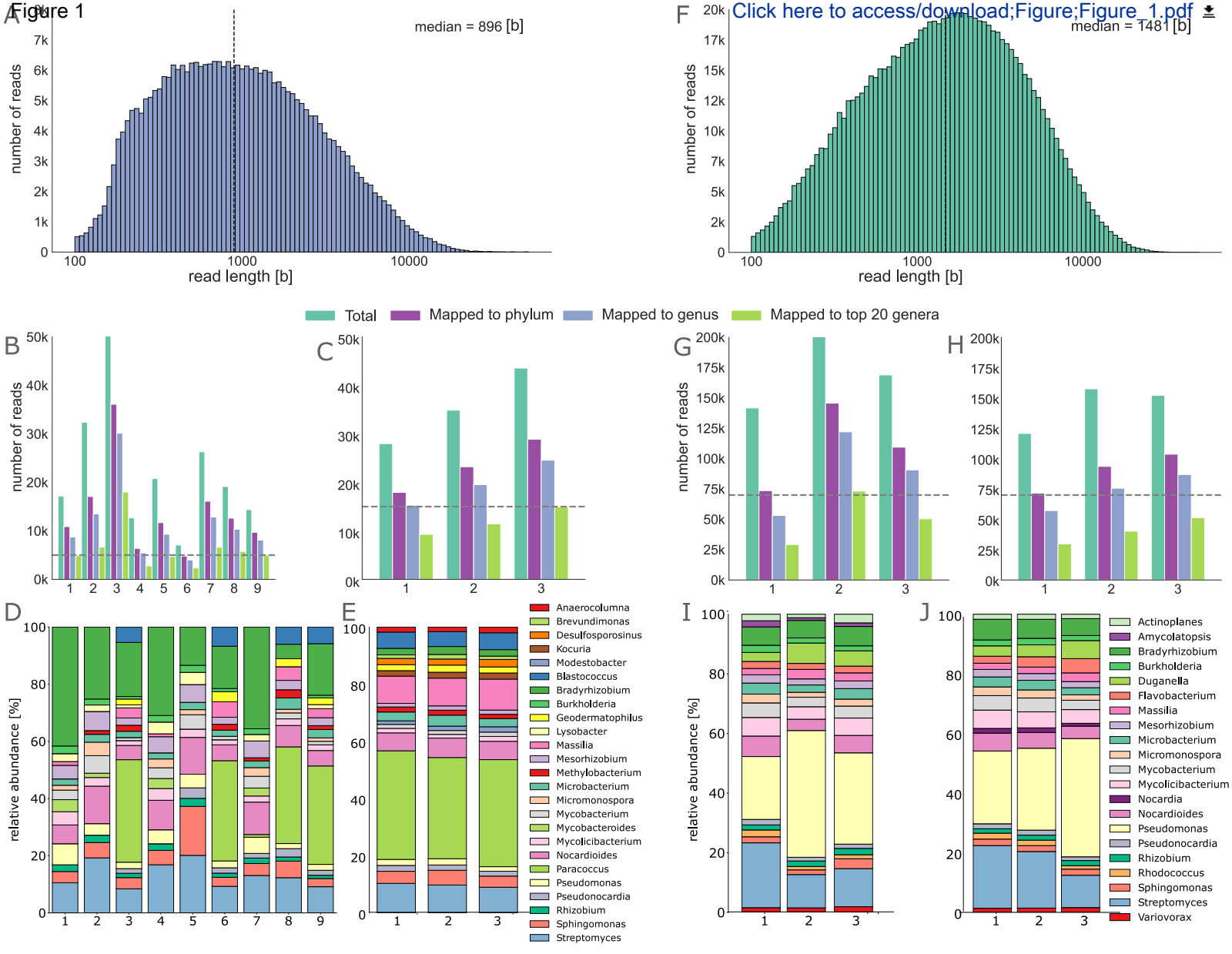
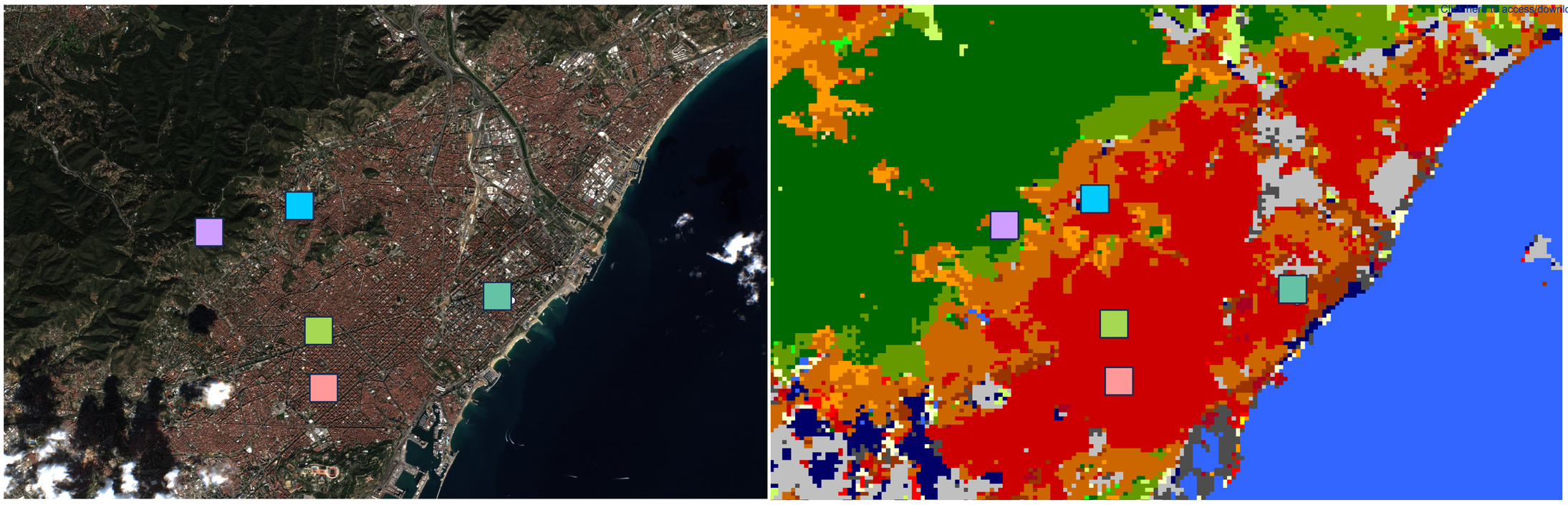
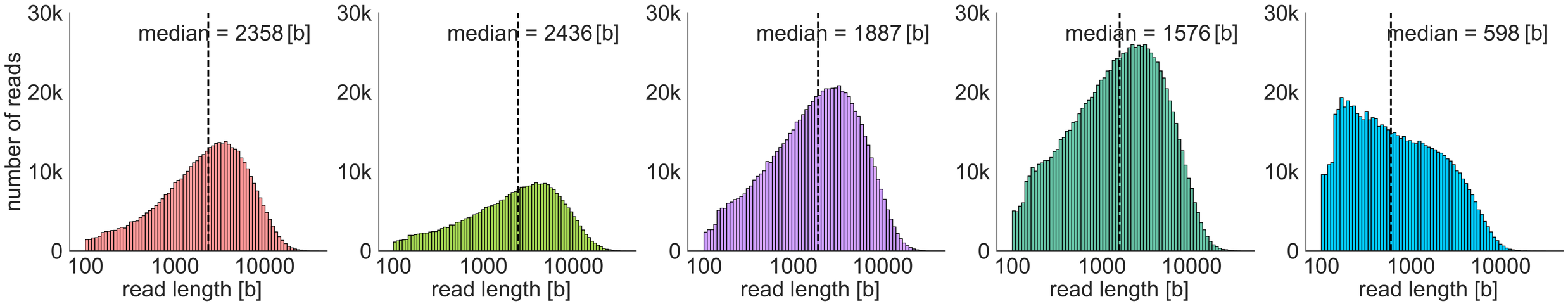


Figure 2

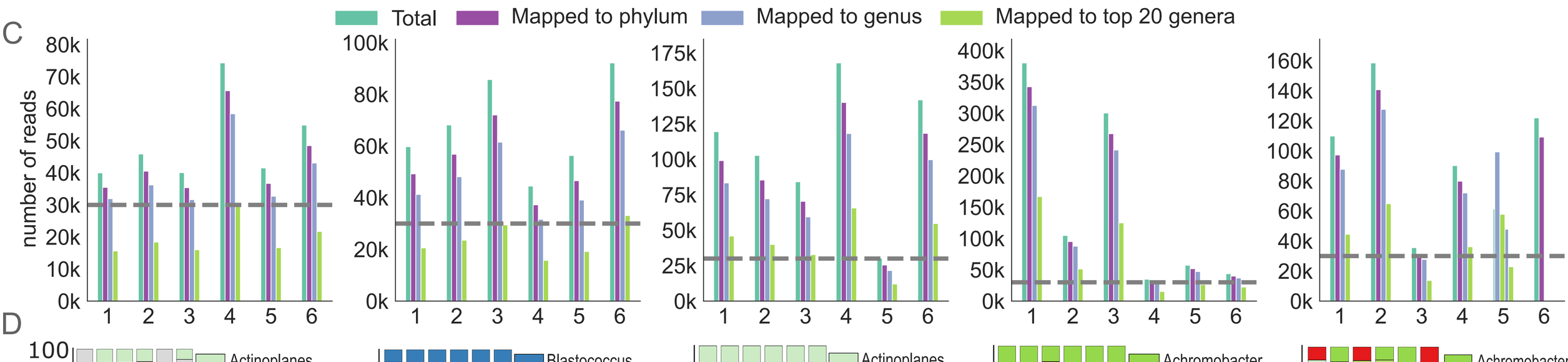
A



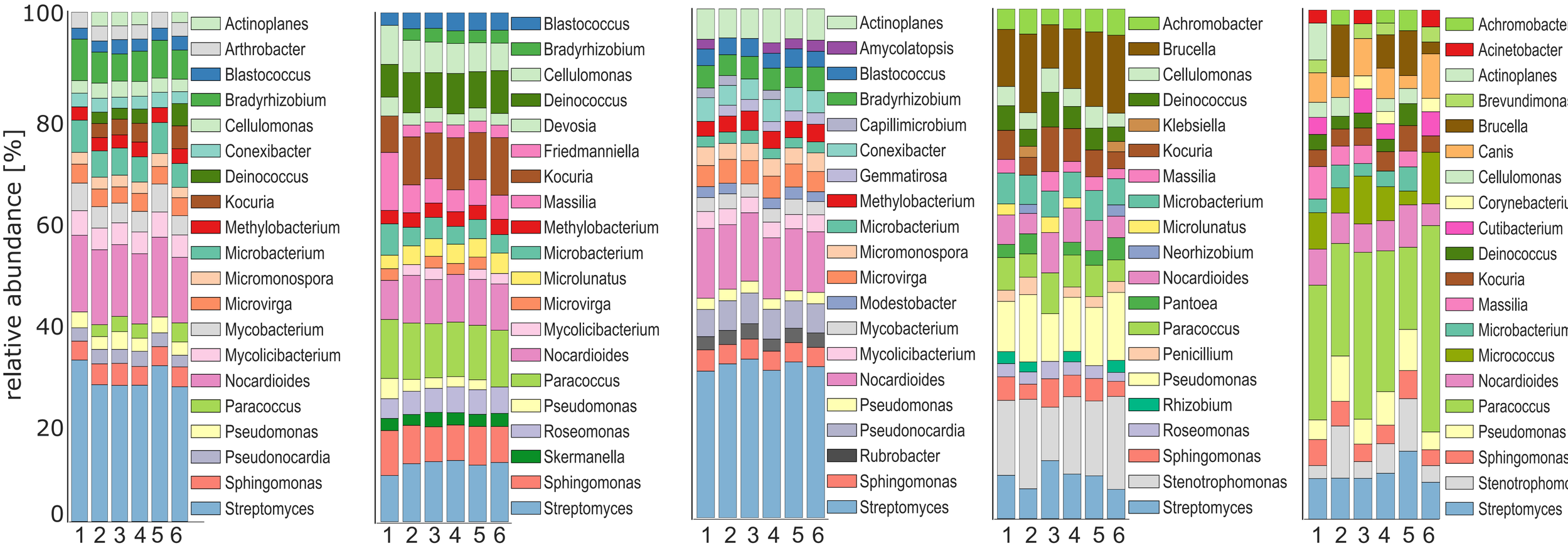
B



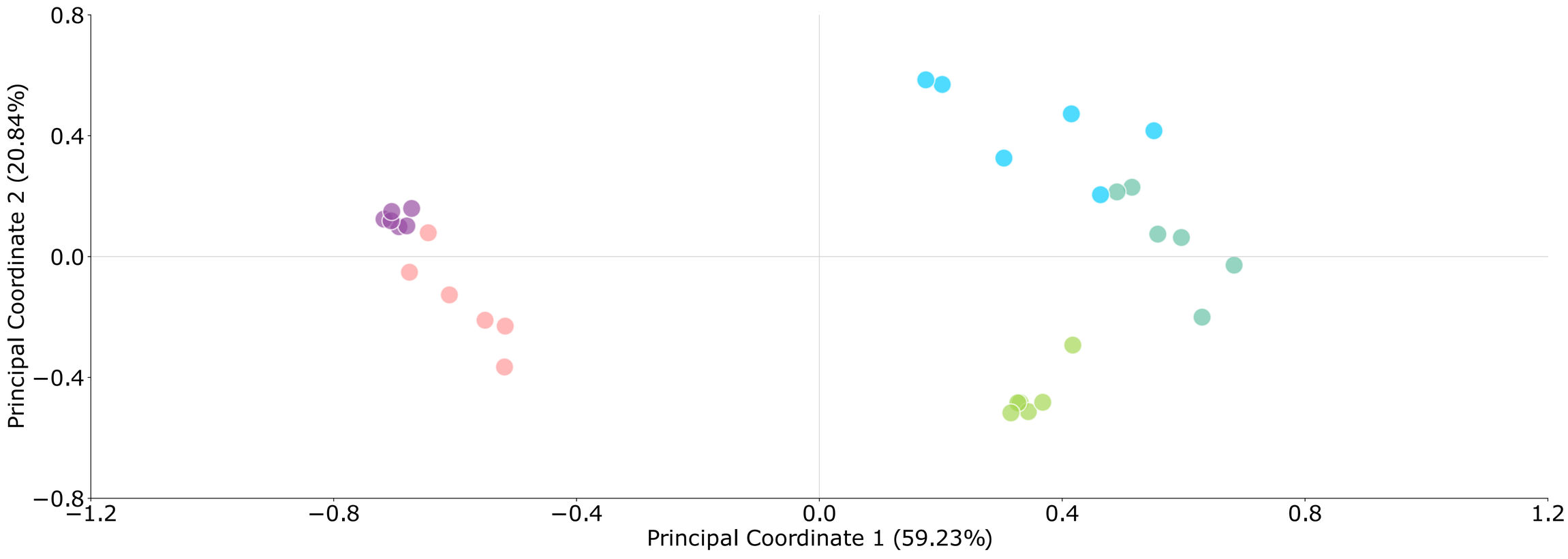
C



D



E



Supplemental Figures

Supplemental Table 1

Supplemental Table 2

Supplemental Table 3

Munich, June 14th 2024

Dear Marnix, dear Janet, dear ISME Communications Editorial Board, dear reviewers,

We thank the editors and reviewers of our manuscript **Air monitoring by nanopore sequencing** for their constructive feedback. We herewith resubmit an extensively revised version of our manuscript, which should address all of the reviewers' comments and importantly contains completely new data which we specifically generated and analyzed to demonstrate the robustness of our laboratory and computational approaches. We specifically applied our approaches to robustly characterize the urban air microbiome of the city of Barcelona, Spain, which is known to suffer from high air pollution and where one of our research centers (ISGlobal is based). We further provide more detail on our assessments of potential human health consequences (please refer to our reply to reviewer #1's 2nd comment). Please find all our detailed responses below.

> *Reviewer #1: Comments to the authors: This is a well-presented and performed study detailing the use of nanopore sequencing for characterising the air microbiome. It includes suitable positive and negative controls and provides a useful guide to those working in a similar field that would like to employ nanopore sequencing in their own studies. However, there are some improvements that could be made:*
>
> *1. The authors refer to supplementary information when stating that several methods were tested - however, it would be useful to briefly include the outcome of method comparisons and outcomes (ie which methods were tried and which final one was selected and why) in the main text (referring to more detail in supplementary information). Note the specific section within the supplementary info needs to be stated so the reader can more easily locate the relevant section of the supplementary information. This is the case in other parts of the manuscript - a specific part of the supplementary information should be referenced - not just (Supplementary Information).*

Thank you very much for your positive feedback! We followed your suggestion and now provide a short overview of the methods and results that we have so far only mentioned in the Supplementary Information. We now specifically have two text sections in the Supplementary Information, (1) "Air sampling and DNA extraction optimizations" (p. 5, ll 56-86.) and (2) "Functional annotation" (p. 6-7, ll 86-121) which provide some information for interested researchers, but which (1) did not result in enough data to make robust conclusions, or (2) allowed us to make hypotheses about microbial ecosystem functions which we will, however, not be able to follow up in this study.

Materials and Methods:

p. 5, ll. 115-124:
"We first conducted preliminary tests to compare standard air sampling and DNA extraction approaches for nanopore sequencing-based air metagenomics; this included the testing of standard quartz filter- and liquid impingement-based air samplers and the optimization of respective DNA extraction approaches for subsequent nanopore shotgun sequencing, which relies on minimum DNA

input without nucleotide amplification and is sensitive to native DNA contamination (Supplementary Information: Air sampling and DNA extraction optimizations)."

p. 11, ll. 271-274
"We finally applied functional annotation to our metagenomic dataset to assess the presence of general metabolic pathways and ecosystem functions (Supplementary Information: Functional annotation);"

Results:

p. 12, ll. 300-304
"After confirming that Coriolis µ liquid impingement resulted in sufficient high-quality DNA yield for nanopore shotgun sequencing after one hour of sampling (Materials and Methods; Supplementary Information: Air sampling and DNA extraction optimizations),"

p. 17, ll. 431-434
"Functional annotation of the reads, contigs, and bins detected typical microbial metabolic functions (Supplementary Information: Functional annotation)."

> *2. The authors perform functional annotation, but do not really explore these findings within the context of the air microbiome. Table 1 is referred to as having information on putative biodegradation mechanisms - but no such information is clearly included - above the genes identified. It would be useful for the reader to include brief descriptions of the functional roles of these genes/groups of genes in this or another table. While the authors do elaborate on the functional roles of some of the identified genes in the text, more could be included on how these are related to functions of the air microbiome and in particular putative biodegradation mechanisms that they highlight numerous times. They particularly state that this method can be used to describe functions related to/to evaluate human and environmental health consequences, and targeted applications such as infectious disease transmission. However, they do not go into any examples here from their own results.*
> *Expanding on this would improve the impact of the paper, really demonstrating the use and benefits of the functional annotation.*

Thanks to your feedback, we discussed the general gene annotation results in more detail, especially with the microbiologist (Michael Schloter) on our team, and we concluded that our statements on biodegradation functions can only be of hypothetical nature and that we currently don't have means to follow up on them experimentally. As the statements on such microbial functions were not the focus of the manuscript, we now shifted them to the Supplementary Information (p. 6-7, ll. 87-121), where we discuss findings in our new urban air microbiome in a more suggestive manner.

Importantly, we also analyzed our data for evidence of human and environmental consequences. First, we now importantly include "real-world" microbial data by monitoring the urban air in Barcelona, Spain (p. 7, ll. 164-176); this allows us to assess if we can indeed find relevant pathogens and

antimicrobial resistance and virulence genes in public spaces, which would be relevant for public health assessments.

We next included Supplementary Table 3 that lists all the resistance and virulence genes that we detected in all our air samples. We further included the input of a medical doctor (Ela Sauerborn) who pinpointed resistance genes that are known to have human health consequences; she specifically found many beta-lactamases:

p. 20-21, ll 516-528.
"The annotation of antimicrobial resistance and virulence genes in our metagenomic data shows that we can use the same dataset to assess potential anthropogenic impacts on microbial diversity while concurrently understanding potential public health consequences. We detected evidence of antimicrobial resistance across all sampled environments (Supplementary Table 3), but especially the detections of clinically relevant beta-lactamases such as blaCARB-8, blaOXA-1, and blal-1, and of genes conferring resistance to other antibiotics such as carbenicillin and oxacillin, in Barcelona's urban air microbiome underscore the possibility of monitoring airborne virulence dissemination using nanopore-based metagenomics."

We next focused on the detection of specific pathogenic taxa. We hereby obtained high-quality assemblies of *Stenotrophomonas maltophilia* and *Salmonella enterica*, which were both detected in the urban air microbiome (Table 1).

p. 21-22, ll 529-555.
"Genome assembly and binning of the long nanopore reads further allows us to be more confident in the presence of specific microbial species and of their pathogenic potential through the identification of Metagenome-Assembled Genomes (MAGs) (Table 1). We obtained high-quality genome assemblies (Materials and Methods) of the pathogenic species Stenotrophomonas maltophilia and Salmonella enterica from the urban microbiome data (Table 1). The Stenotrophomonas species is known as an emerging difficult-to-treat human pathogen, and many of the Salmonella enterica serovars can cause disease in humans through zoonotic or foodborne transmission. While we require good coverage of a microbial genome to create such assemblies for taxonomic species or strain identification, also just the presence of individual pathogen-associated sequencing reads might be used for obtaining first information on the potential presence of microorganisms of public health concern. For example, given the presence of sequencing reads of the Brucella genus, a typical canine pathogen, in several of our urban air samples, we further analyzed our taxonomic annotation, which was based on the entire NCBI nt database, and were indeed able to detect the presence of Canis lupus familiaris in the same air samples. While this might point to a potential impact of animal domestication and specifically frequent dog walking in Barcelona on public health, such interdependencies would have to be investigated in a controlled and/or experimental setting."

> *3. The authors could also comment on how this form of sequencing compares to others in terms of detecting microbes within air.*

As we now discuss in more detail in the Introduction, the most commonly used method for characterizing air microbiomes is the amplicon-based approach followed by short-read Illumina

sequencing (p. 4, ll. 78-83), which copes better with the ultra-low biomass nature of the air but introduces well-known biases (amplification bias, database completeness, taxonomic resolution, etc.). We now also discuss the advantage of metagenomics and long-read shotgun sequencing in comparison to such amplicon-based approaches in more detail the Introduction (p. 4, ll. 83-93).

> *4. The authors refer to the "stable core microbiome" and that at least 3hour sampling was required to detect a "stable core microbiome". This suggests to the non-expert reader that a 3h plus sampling time is the best to use for consistency. However, it is likely that the microbiome constantly changes at local levels (within the breathing range of humans) - perhaps on a minute by minute basis - and so sampling for longer periods of time reduces temporal resolution - which may be important for certain environments, particularly when measuring/analysing anthropogenic sources of bioaerosols. It would be interesting to know the detection limits for this method - ie the shortest sampling time that still generates a robust signal, and indeed if/how the microbiome changes over short periods of time. The authors themselves state how the air microbiome can have diurnal patterns. I appreciate the more stable results demonstrate the reproducibility of the method - but the reader/potential user of such a technique may need to consider what sampling time is appropriate to them depending on their research question. This could be clarified in the manuscript.*

Thank you for this important input. Our maximal temporal resolution in this study was 1h since even when sampling for an hour, we still obtained very little DNA, which was however just sufficient to create interpretable nanopore sequencing data. We showed in our pilot study that 1h-sampling resulted in a stable microbiome across sampling events per day and sampling events across days in a controlled environment (greenhouse; Figure 1D), but that the across-day 3h-sampling results seemed to be even more stable (Figure 1E). We subsequently showed that 3h-samples are also sufficient to capture similar microbial signatures across days in a natural environment (i.e. outdoors; Figure 1I; Figure 1J for 6h-samples).

Having said all this, we now most importantly include real-time data from different locations in Barcelona to describe the city's air microbiome. Here, we show that – across two technical replicates and randomized across days over a time period of two weeks and randomized between morning and afternoon sampling – we can still assess highly location-specific air microbiomes using 3h-sampling and our approaches (Figure 2D). As such robust assessments of air microbiomes was our primary goal here – to subsequently think about human and environmental consequences that we might draw from such data –, we believe that the longer sampling time is warranted – while we absolutely agree that these assessments accumulate across any smaller-scale variability, which we are not interested in in this case and would also not be able to measure given minimum DNA inputs.

> *5. Minor comments:*

Abstract - "we here show" to "Here we show"
Response: Thank you, corrected. (p. 2, ll. 30-33)

Abstract - "assemblies from the long sequencing" - remove "the"

Response: This section is not anymore present in the new version of our manuscript.

Intro - "can create such long sequencing reads" to "can sequence long reads"
Response: This section is not anymore present in the new version of our manuscript.

Intro - "that would allow to leverage" to "that would leverage"
Response: This section is not anymore present in the new version of our manuscript.

Intro - "to allow for robust" to "to enable robust"
Response: Thank you, corrected. (p. 5, ll. 100-102)

Intro - "air, and concurrently provides" to "air, providing"

Response: Could you please specify where you found it?

Intro "genomes annotations" to "genome annotations"

Response: Could you please specify where you found it?

Figure 1 legend - remove "across the C. 1h- and D. 3-Gh samples" or change to just "across samples"

Response: Thank you, corrected. (p. 33, ll. 795-799)

Figure 1 legend - remove "across the I. 3h- and J. 6-Gh samples" or change to just "across samples"

Response: Thank you, corrected. (p. 33, ll. 801-805)

Results and discussion - "while especially the fungal taxa and Gram-positive Bacillus subtilis were underrepresented by our genomic data" change to "although the fungal taxa and Gram-positive Bacillus subtilis, in particular, were underrepresented…". End sentence after "our genomic data" and start new sentence with "This might reflect"

Response: Thank you, corrected. (p. 14-15, ll. 363-366)

Within methods - authors refer to supplementary information - this needs to specify which part of supplementary info they are referring to specifically. There are also a couple of examples of this throughout the manuscript - ensure when referring to supplementary info - it is clear which specific section

Response: Thank you, we improved the text and reference to the methods and supplementary information accordingly.

Reviewer #2:

This study is well conducted and written up, I have no hesitation in recommending it for publication with a few minor corrections and queries to the authors. The application of shotgun sequencing using the nanopore platform is novel and challenging. In addition to the data published in the main paper, there is a wealth of methodological optimization included in the supplementary information that will be extremely valuable to the field.

I Have a couple of questions for the authors:

1) Do they think fungi are underrepresented in the data? To me there are fewer fungi that I would expect. What was the proportion of reads assigned to fungi (and other groups like arthropods and plants which are also collected with Coriolois)? The plots are for the most dominant taxa only, so maybe the fungi are there but missed on the plots, can they make a plot that filters for fungi only?
Do you think there is a bias for bacteria? Is there a reason for a bias towards bacteria (extraction, bioinformatics/QC of data?). It would have been good to have a comparison with meta-barcoding, or shotgun sequencing on an Illumina platform.

Thank you very much for your very positive feedback on our manuscript! Indeed, we believe that fungal taxa (and possible others with study cell walls) are underrepresented in our data due to choices we had to make with respect to sample processing and DNA extraction. We now detail this in much more detail in the Discussion:

p. 23 ll. 578-596
"We emphasize that our sampling, laboratory and computational approaches constitute one feasible and reproducible way of using nanopore shotgun sequencing to profile the air microbiome. While we tested some additional established air sampling and DNA extraction methodology, we have not conducted an extensive study of all possible approaches. We specifically emphasize that the detection of fungi and Gram-positive bacteria could be improved when using different sample processing and DNA extraction techniques. This is also reflected by the application of our approaches to a positive control, which shows that fungal taxa and Gram-positive Bacillus subtilis, in particular, were underrepresented. As sturdier cell walls would require more aggressive DNA extraction approaches, this would, however, also lead to increased DNA fragmentation, especially in Gram-negative bacteria, and therefore more difficult downstream analyses. A good trade-off could be the sequencing of several, differently processed DNA extracts and subsequent data pooling to assess the microbial diversity of any air sample more holistically."

We here importantly emphasize the results of the application of our protocols to a mock community that contains fungal as well as Gram-positive and -negative taxa. We believe that the results of this mock community analyses very well reflect the biases in our data, and discuss a trade-off for more holistic microbial assessments in the future. As other approaches lead to their own biases in similar mock communities (e.g. strong overamplification of Enterobacteriaceae bia 16S rRNA-based metabarcoding approaches, see e.g. Urban et al. (2021), eLife: https://elifesciences.org/articles/61504#fig2), we believe that the applications of our approaches to

the mock community are most meaningful to point towards certain biases in our data. With respect to direct comparisons between Illumina- and nanopore-based shotgun sequencing, we believe that past research, also on mock communities (e.g., Liu et al. (2022) BMC Microbiome: https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-022-01415-8), has already shown the improved performance of nanopore sequencing for certain metagenomic approaches.

> *2) The MAGs have low completeness (table 1 max is ~65%). Contamination does not seem to be a problem (indeed looks very good in comparison). Still but the best MAGs are "medium" quality (https://doi.org/10.1038/nbt.3893) and I guess the majority you found are very poor and therefore not in the table at all?*
>
> *Why do you think this is? Too little coverage? is more biomass needed? Where the reads too fragmented, as the bioinformatics seem robust to me. Contamination is low maybe because of the long reads. Would greater completion of the MAGs be a benefit of shotgun sequencing on Illumina here? Please comment on this as a possible limitation of long- v short-read shotgun methods in the discussion.*

Thank you; we now discuss this in much more detail in the Discussion as well (please see below). However, we would not expect any greater completion of MAGs with short-read sequencing data, but the exact opposite. Many other studies have shown that long-read sequencing data substantially improved de novo assemblies (e.g., Liu et al. (2022) BMC Microbiome: https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-022-01415-8).

p. 22, ll 556-577.:
“While we were able to build de novo assemblies from our nanopore-based air metagenomic data, most of the MAGs were incomplete (<30%) and/or showed high levels of contamination (>10%) (Table 1). Given the low amount of DNA input and therefore relatively small size of the resulting metagenomic datasets in combination with the expectedly high fragmentation of DNA in air samples, this might just be an inherent shortcoming when it comes to assessing the air microbiome – albeit applying long-read sequencing technology. We here found a particular small median DNA fragment and sequencing read length for the Outer Belt location (Figure 2B), which might point towards the impact of environmental conditions or specific taxonomic compositions (and variables such as the microorganisms’ genome size and cell wall composition) on the final fragment and read length distribution. It is further expected that non-viable microorganisms, which might significantly contribute to the air microbiome, result in more fragmented DNA in the air samples; this means that substantial differences in read lengths between microbial taxa might also be attributed to their differential viability in the air environment – a hypothesis that we might be able to resolve in the future using viability-resolved metagenomic approaches.”

.

> *3) Other points:*

> Not clear in the abstract what natural means to me. It is a vague term, I get what you are trying to say from the context and as a contrast to a greenhouse, but maybe be more specific about your natural and controlled environments (specify them in brackets or similar) for the abstract.

Response: Thank you. The corresponding changes have been made throughout the manuscript.

> Background paragraph 2. "… due to the air's ultra-low biomass (1, 4)." The air its self cannot have low biomass, rephrase please.

Response: Thank you, corresponding changes have been made p. 3 ll. 63-65

> Background paragraph 2. Would be good to include some of the limitations of shotgun sequencing for characterizing community composition and diversity, particularly when paired with long-read. For example, bias to taxa with large genomes, sequencing of non-target DNA (human etc.) taking up read depth and reducing coverage, lack of standardized unit of diversity (e.g. 16s gene) for robust estimation of alpha-diversity.

Response: Thank you for your comment. We have incorporated this information into the introduction, highlighting the limitations of shotgun sequencing line p. 4 ll. 78-83.

> Will the accession numbers for data be provided, and will the draft MAG genomes be available also? accession number

Response: The data will be available under the accession number: PRJEB76446 including the draft MAG genomes.

Reviewer #3:

In this letter article, Tim et al. claimed to establish an experimental and bioinformatic analysis protocol for monitoring air microbiome with metagenomic approach using nanopore sequencing. However, the article's focus is unclear, as neither sample handling nor bioinformatics analysis is elaborated upon in depth. The sampling method appears to be a standard approach for aerosol collection, not specifically optimized to match the rapid sequencing capabilities of nanopore technology. Additionally, it does not address the challenge of low biomass in aerosol samples, which is a critical factor challenging effective nanopore library preparation for aerosol samples. Moreover, the paper does not present any novel analytical algorithms, nor does it attempt to streamline existing tools into a user-friendly package or enhance databases. From their data, I infer a precarious sequencing endeavor characterized by exceedingly short reads and low throughput. Indeed, long reads can be annotated down to the species level and aid in metagenome-assembled genome (MAG) recovery, but these are general advantages of long-read sequencing, not unique to aerosol samples. The article fails to address the real challenges in aerosol microbiome monitoring, such as developing protocols to collect sufficient biomass or methods to enrich DNA from low-biomass samples. In summary, the research presented lacks innovation and is unlikely to make a significant contribution to the advancement of the field.

Thank you for taking the time to read our manuscript. We, however, have to disagree with the reviewer's basic assumption that our newly established laboratory and computational approaches are not novel. We believe that reviewer #1's and #2's comments on the other hand, show how we have established a completely novel end-to-end protocol to use nanopore sequencing for air metagenomics that previously did not exist and has brought many challenges (see, e.g., reviewer #2's comment: "The application of shotgun sequencing using the nanopore platform is novel and challenging."). We have now made our advances clearer throughout the manuscript, and also expanded on the background in our Introduction section.

We now discuss in more detail that our approach is obviously not the only one to assess air monitoring by nanopore sequencing (e.g. p. 23, ll 578-581.: "We emphasize that our sampling, laboratory and computational approaches constitute one feasible and reproducible way of using nanopore shotgun sequencing to profile the air microbiome."), but we highlight the specific challenges of establishing such an approach throughout the manuscript (e.g. low biomass, fragmented DNA for MAG generation, DNA extraction without nanopore blocking given the low biomass, computational taxonomic assignment, etc.) and provide solutions to allow for air monitoring by nanopore sequencing.

We finally hope that the additional very robust results of the first "real-world" application of our protocol clearly showcase how the replicable assessments of an exemplary urban air microbiome can inform future studies and potential public health consequences. We also believe that alone the finding of surprisingly location-specific air microbiomes within a city's boundary will be of substantial interest to the research community.