

Air monitoring by nanopore sequencing

Tim Reska^{1,2,3,†}, Sofya Pozdniakova^{4,†}, Sílvia Borràs⁴, Albert Perlas^{1,2}, Ela Sauerborn^{1,2,3}, Lúdia Cañas⁴, Michael Schlöter^{3,5}, Xavier Rodó^{4,6}, Yuanyuan Wang⁷, Barbro Winkler⁸, Jörg-Peter Schnitzler⁸, Lara Urban^{1,2,3,*}

¹Helmholtz AI, Helmholtz Zentrum München, 85764 Neuherberg, Germany

²Helmholtz Pioneer Campus, Helmholtz Zentrum München, 85764 Neuherberg, Germany

³Technical University of Munich, School of Life Sciences, 85354 Freising, Germany

⁴AIRLAB, Climate and Health (CLIMA) group, ISGlobal, 08003 Barcelona, Spain

⁵Institute of Comparative Microbiome Analysis, Helmholtz Zentrum München, 85764 Neuherberg, Germany

⁶Catalan Institution for Research and Advanced Studies, ICREA, 08010 Barcelona, Spain

⁷Technical University of Munich, School of Engineering and Design, 80333 Munich, Germany

⁸Research Unit Environmental Simulation (EUS), Helmholtz Zentrum München, 85764 Neuherberg, Germany

*Corresponding author: Lara Urban, Helmholtz AI, Helmholtz Munich, Ingolstädter Landstr. 1, Neuherberg 85764, Germany. Email: lara.h.urban@gmail.com

†Contributed equally

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.

Keywords: urban air microbiome, bioaerosols, metagenomics, nanopore sequencing, long-read sequencing, shotgun sequencing, de novo assembly, infectious disease, antimicrobial resistance

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].

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].

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.

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

Received: 18 June 2024. Revised: 12 July 2024. Accepted: 15 July 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of the International Society for Microbial Ecology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

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 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 18000 $\times$ g for 25 min, and collected the pellets. The pellets were resuspended, aggregated, and subsequently centrifuged twice at 18000 $\times$ 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 1 h in three consecutive replicates per day or for 3 h 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 3 h or 6 h of air sampling; we here tested

6 h 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 6 h-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 16th October to 3rd November 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 3 h 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 μ L 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 μ L 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 1 h-, 3 h-, 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 h. 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 μ L nuclease-free water (Thermo Fisher Scientific) per DNA extraction and one sample of 20 μ L 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: 5 k reads for 1 h-samples from the controlled environment, 15 k reads for the 3 h-samples from the controlled environment, 70 k reads for the natural environment samples, and 30 k 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 30 k 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 24 October 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 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, PM₁₀, 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 posthoc 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 1 h-samples, DNA yields ranged from 17.7 to 50.7 ng (0.98 to 2.82 ng/m³), while the 3-hour samples showed DNA yields ranging from 130.2 to 179.4 ng (2.41 to 3.32 ng/m³; Supplementary Table 1; pilot_study sheet). Nanopore shotgun sequencing delivered between 7 and 60 k high-quality sequencing read at a median read length of 896 bases (Fig. 1A), respectively, of which 5 to 35 k reads were successfully mapped to the taxonomic genus level using Kraken2 and the NCBI nt database (Fig. 1B-C; Supplementary Table 1; pilot_study sheet). After downsampling to the same number of reads per sample type (1- and 3 h-samples, respectively), the taxonomic composition of the 20 most abundant taxa indicated that only the 3-h sampling duration captured a stable "core" air microbiome across days at the genus level (Fig. 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 Figs 1–2). The most abundant genera included soil- and plant-associated bacteria such as *Bradyrhizobium*, *Paracoccus*, *Nocardioides*, *Massilia*, and *Streptomyces* (Fig. 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 3 or 6 h, 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 in 130 to 200 k high-quality sequencing reads at a slightly higher median read length than the controlled environment of 1481 (Fig. 1F), of which 70 to 140 k reads were successfully mapped to the taxonomic genus level (Fig. 1G-H; Supplementary Table 1; pilot_study sheet). After downsampling all samples to 70 k reads, analysis of the relative abundance of 20 most abundant taxa revealed a very similar profile for both 3-h and 6-h 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 Figs 1–2). A distinct air microbiome profile was observed in the natural environment in comparison to the controlled settings, with a high predominance of *Pseudomonas* and unique detection of microbial taxa such as *Actinoplanes*, *Amycolatopsis*, *Dugnaella*, *Flavobacterium*, *Nocardia*, *Rhodococcus*, and *Variovorax* (Fig. 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*,

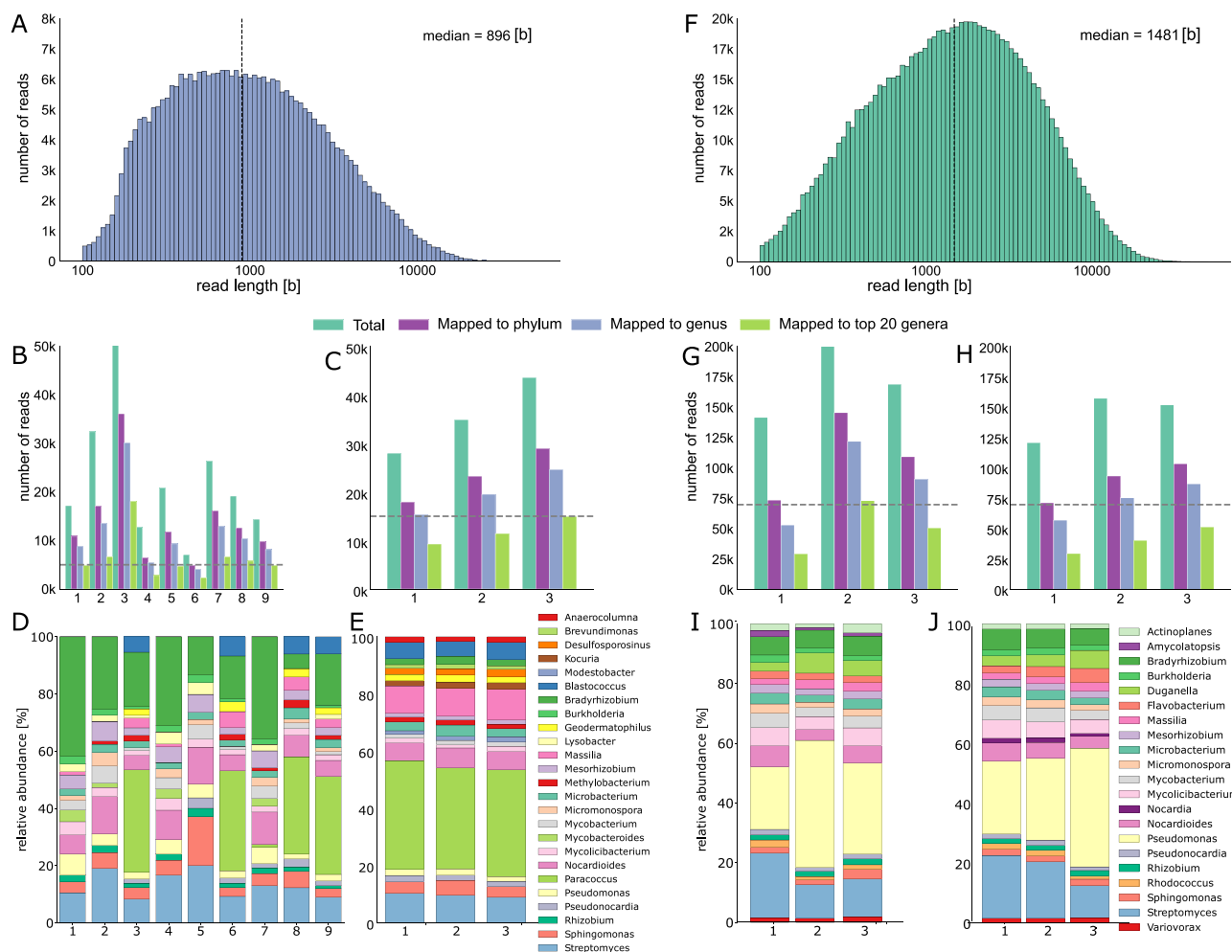


Figure 1. Robust air microbiome assessments of a controlled (left; **A-E**) and natural (right; **F-J**) environment through nanopore shotgun sequencing. **A.** Nanopore sequencing read length distribution across 1 h- and 3-Gh samples. **B-C.** Number of total sequencing reads, and of reads mapping to 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.** 1 h- and **E.** 3 h-samples after downsampling based on the 20 most dominant genera across samples. **F.** Nanopore sequencing read length distribution across 3 h- and 6 h-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.** 3 h- and **J.** 6 h-samples after downsampling based on the 20 most dominant genera across samples.

Shigella, *Francisella*, and *Pseudomonas* (Supplementary Fig. 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 Fig. 3C; Material and Methods).

We finally applied our optimized laboratory and computational approaches to assess an exemplary urban microbiome using nanopore metagenomics (Fig. 2A; left; Materials and Methods). Our remote-sensing-based LCZ classification (Fig. 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 a compact mid-rise category, which features taller buildings on 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, PM₁₀ and PM_{2.5}, between our sampling locations (Supplementary Fig. 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 PM₁₀ (Supplementary Fig. 4).

Nanopore shotgun sequencing delivered between 33 and 422 k high-quality sequencing read at a median read length of between 598 and 2358 bases (Fig. 2B), respectively, of which 21 to 312 k reads were successfully mapped to the taxonomic genus level using Kraken2 and the NCBI nt database (Fig. 2C; Supplementary Table 1; urban_study sheet). The City Center exhibited the longest DNA fragments, and the Outer Belt location the shortest DNA fragments (Fig. 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).

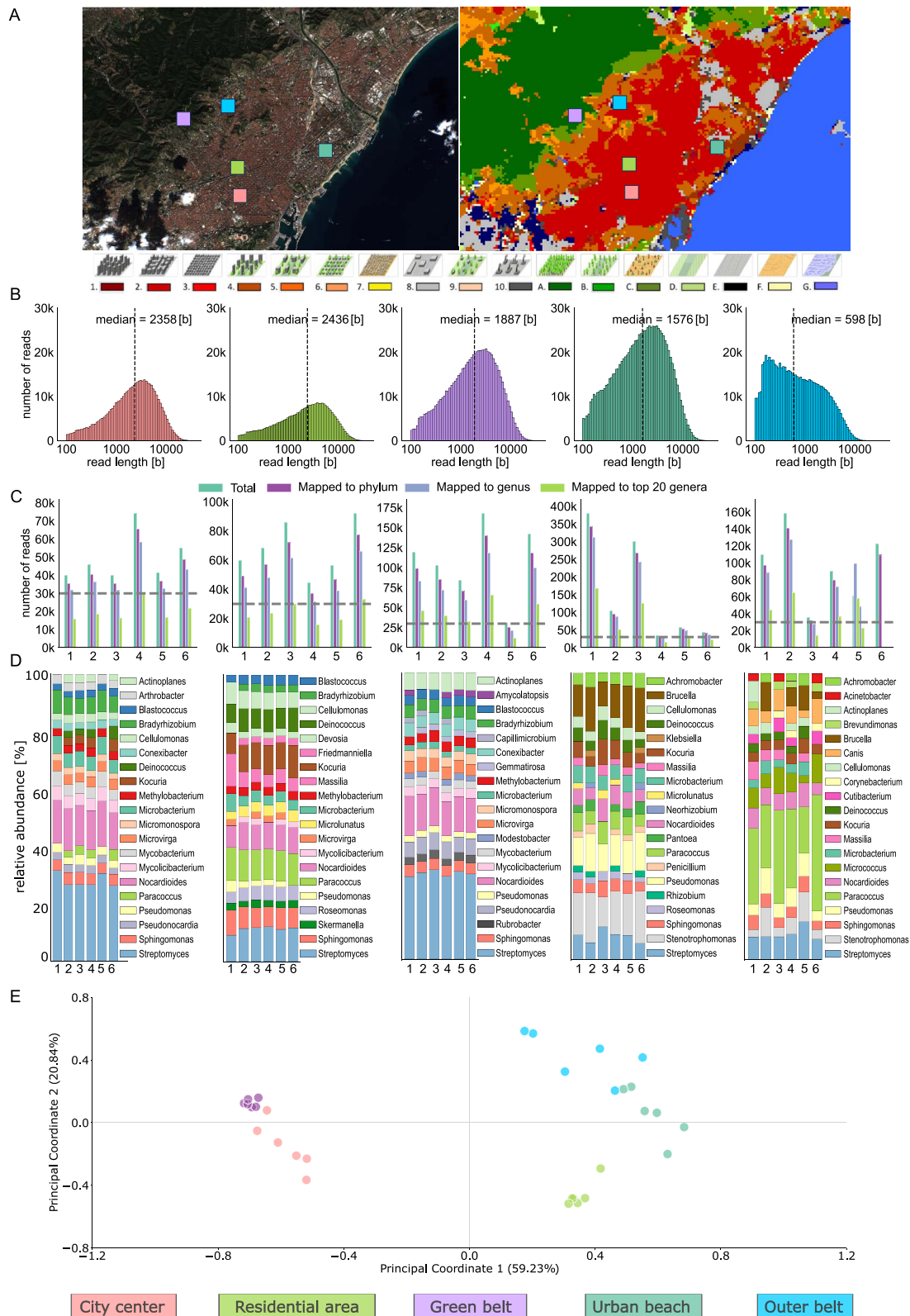


Figure 2. LCZ classification and metagenomic analysis of Barcelona using satellite images and nanopore shotgun sequencing data. **A.** Sentinel-2 image and LCZs classification map of Barcelona on 24 October 2023, with a legend of LCZs classes at the bottom. The colored squares indicate the five sampling 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 30 K reads before taxonomic classification using Kraken2. **E.** PCoA of the relative abundances of the bacterial genera identified at the five sampling locations.

For taxonomic comparisons across replicates and samples, we again downsampled the number of reads (here to 30 k 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 (Fig. 2D; Materials and Methods). The core urban air microbiome consisted of microbial genera such as *Streptomyces*, *Sphingomonas*, *Pseudomonas*, *Nocardioideae*, 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*, *Deinococcus*, and *Cellulomonas* were not detected at all (Fig. 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 (Fig. 2E).

Despite the location-specific differences in air microbial composition (Fig. 2A), in LCZ-based land usage (Fig. 2D), and in air pollution measured by particle mass fractions (Supplementary Fig. 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 [4–7]. 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, 34, 35], 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 (Fig. 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 (Fig. 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%; Fig. 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 close by 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 [36, 37].

The individual samples of the Green Belt location cluster together most tightly (Fig. 2E), 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 [38]. 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; Fig. 2A) or of air pollution (as assessed through particle mass fraction measurements; Supplementary Fig. 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 [39]. 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 *blaL1*, and of genes conferring resistance to other antibiotics such as carbenicillin and oxacillin [40], in Barcelona’s urban air

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	5928	/	/	/	/
GH3h	121	15 330	2	<i>Paracoccus aerius</i>	64.59	2.94
				<i>Paracoccus denitrificans</i>	63.41	1.46
Nat3h	204	7401	/	/	/	/
Nat6h	117	7282	/	/	/	/
City center	1170	23 151	/	/	/	/
Residential area	470	11 098	/	/	/	/
Green belt	1171	15 215	1	<i>Burkholderia</i> sp.	36.66	2.38
Urban beach	7732	21 049	1	<i>Stenotrophomonas maltophilia</i>	48.33	6.69
Outer belt	1874	10 282	1	<i>Salmonella enterica</i>	41.72	10.59

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 [41] and many of the *S. enterica* serovars can cause disease in humans through zoonotic or foodborne transmission [42]. 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 air samples [43]. While this might point to a potential impact of animal domestication and specifically frequent dog walking in Barcelona on public health [44], such complex interdependencies can only be confirmed in a controlled and/or experimental setting.

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 (Fig. 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 [45].

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 methodologies, 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 *B. 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.

Acknowledgements

None.

Author contributions

Tim Reska, Sofya Pozdniakova, Sílvia Borràs, and Lara Urban conceptualized and designed the experiment. Tim Reska and Sofya Pozdniakova handled investigation, sampling, processing, and formal analysis. Lara Urban supervised the project. The manuscript was written by Tim Reska, Sofya Pozdniakova, and Lara Urban, with all authors participating in review and editing, finalizing the document for approval.

Supplementary material

Supplementary material is available at ISME Communications online.

Conflicts of interest

None declared.

Funding

This study was funded by a Helmholtz Principal Investigator grant awarded to L.U. Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt and the Institutional Identifier is: 501100013295.

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.

References

- Whitby C, Ferguson RMW, Colbeck I et al. Chapter three - compendium of analytical methods for sampling, characterization and quantification of bioaerosols. In: Bohan D.A., Dumbrell A. (eds.), *Advances in Ecological Research*. Cambridge, Massachusetts: Academic Press, 2022, 101–229.
- 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. <https://doi.org/10.1016/j.wasman.2015.04.015>
- Robinson JM, Breed MF. The aerobiome-health axis: a paradigm shift in bioaerosol thinking. *Trends Microbiol* 2023;**31**:661–4. <https://doi.org/10.1016/j.tim.2023.04.007>
- Naumova NB, Kabilov MR. About the biodiversity of the air microbiome. *Acta Nat* 2022;**14**:50–6. <https://doi.org/10.32607/actanaturae.11671>
- Luhung I, Uchida A, Lim SBY et al. Experimental parameters defining ultra-low biomass bioaerosol analysis. *npj Biofilms Microbiomes* 2021;**7**:1–11. <https://doi.org/10.1038/s41522-021-00209-4>
- Drautz-Moses DI, Luhung I, Gusareva ES et al. Vertical stratification of the air microbiome in the lower troposphere. *Proc Natl Acad Sci* 2022;**119**:e2117293119. <https://doi.org/10.1073/pnas.2117293119>
- Gusareva ES, Acerbi E, Lau KJX et al. Microbial communities in the tropical air ecosystem follow a precise diel cycle. *Proc Natl Acad Sci USA* 2019;**116**:23299–308. <https://doi.org/10.1073/pnas.1908493116>
- Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat Biotechnol* 2020;**38**:701–7. <https://doi.org/10.1038/s41587-020-0422-6>
- Chan WS, Au CH, Chung Y 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. <https://doi.org/10.1186/s13104-020-05287-9>
- Sereika M, Kirkegaard RH, Karst SM 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**:823–6. <https://doi.org/10.1038/s41592-022-01539-7>
- Hall MB, Wick RR, Judd LM et al. Benchmarking reveals superiority of deep learning variant callers on bacterial nanopore sequence data. 2024bioRxiv; 2024.03.15.585313.
- Raymond-Bouchard I, Maggiori C, Brennan L 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. <https://doi.org/10.1089/ast.2020.2349>
- Sauerborn E, Corredor C, Reska T et al. Detection of hidden antibiotic resistance through real-time genomics. *Nat Commun* 2024;**15**:5494. <https://doi.org/10.1038/s41467-024-49851-4>
- Quick J, Loman NJ, Durauffour S et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016;**530**:228–32. <https://doi.org/10.1038/nature16996>
- Urban L, Holzer A, Baronas JJ et al. Freshwater monitoring by nanopore sequencing. *elife* 2021;**10**:e61504. <https://doi.org/10.7554/eLife.61504>
- Nygaard AB, Tunsjø HS, Meisal R et al. 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. <https://doi.org/10.1038/s41598-020-59771-0>
- Guppy protocol - Guppy software overview. *Oxford Nanopore Technologies*. https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb_2003_v1_revax_14dec2018 (13 June 2024, date last accessed).
- Nanoporetech/dorado: Oxford Nanopore's Basecaller. *GitHub*. <https://github.com/nanoporetech/dorado> (10 June 2024, date last accessed).
- Wick R. Rrwick/Porechop. 2024. <https://github.com/rrwick/Porechop>
- De Coster W, D'Hert S, Schultz DT et al. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 2018;**34**:2666–9. <https://doi.org/10.1093/bioinformatics/bty149>
- Wood DE, Lu J, Langmead B. Improved metagenomic analysis with kraken 2. *Genome Biol* 2019;**20**:1–13. <https://doi.org/10.1186/s13059-019-1891-0>
- Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015;**12**:59–60. <https://doi.org/10.1038/nmeth.3176>
- Kalantar KL, Carvalho T, de Bourcy CFA et al. IDseq—an open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *GigaScience* 2020;**9**:giaa111. <https://doi.org/10.1093/gigascience/giaa111>
- Kolmogorov M, Bickhart DM, Behsaz B et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods* 2020;**17**:1103–10. <https://doi.org/10.1038/s41592-020-00971-x>
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;**34**:3094–100. <https://doi.org/10.1093/bioinformatics/bty191>
- Vaser R, Sović I, Nagarajan N et al. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 2017;**27**:737–46. <https://doi.org/10.1101/gr.214270.116>
- Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 2018;**6**:1–13. <https://doi.org/10.1186/s40168-018-0541-1>
- Parks DH, Imelfort M, Skennerton CT et al. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;**25**:1043–55. <https://doi.org/10.1101/gr.186072.114>

29. Feldgarden M, Brover V, Gonzalez-Escalona N 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. <https://doi.org/10.1038/s41598-021-91456-0>
30. Tseemann/Abriicate::mag_right: Mass Screening of Contigs for Antimicrobial and Virulence Genes. GitHub. <https://github.com/tseemann/abriicate> (10 June 2024, date last accessed).
31. Shen W, Le S, Li Y et al. SeqKit: a cross-Platform and ultra-fast toolkit for FASTA/Q file manipulation. *PLoS One* 2016; **11**:e0163962. <https://doi.org/10.1371/journal.pone.0163962>
32. Rosentreter J, Hagensieker R, Waske B. Towards large-scale mapping of local climate zones using multitemporal sentinel 2 data and convolutional neural networks. *Remote Sens Environ* 2020;**237**:111472. <https://doi.org/10.1016/j.rse.2019.111472>
33. Salter SJ, Cox MJ, Turek EM et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014;**12**:87. <https://doi.org/10.1186/s12915-014-0087-z>
34. Cummings PJ, Olszewicz J, Obom KM. Nanopore DNA sequencing for metagenomic soil analysis. *J Vis Exp* 2017;**130**:55979. <https://doi.org/10.3791/55979>
35. Urban L, Miller AK, Eason D et al. Non-invasive real-time genomic monitoring of the critically endangered kākāpō. *elife* 2023;**12**:RP84553. <https://doi.org/10.7554/eLife.84553>
36. Ribeiro I, Martilli A, Falls M et al. Highly resolved WRF-BEP/BEM simulations over Barcelona urban area with LCZ. *Atmos Res* 2021;**248**:105220. <https://doi.org/10.1016/j.atmosres.2020.105220>
37. Segura R, Badia A, Ventura S et al. Sensitivity study of PBL schemes and soil initialization using the WRF-BEP-BEM model over a Mediterranean coastal city. *Urban Clim* 2021;**39**:100982. <https://doi.org/10.1016/j.uclim.2021.100982>
38. Lladó S, López-Mondéjar R, Baldrian P. Forest soil bacteria: diversity, involvement in ecosystem processes, and response to global change. *Microbiol Mol Biol Rev* 2017;**81**:e00063–16. <https://doi.org/10.1128/MMBR.00063-16>
39. Urban L, Perlas A, Francino O et al. Real-time genomics for one health. *Mol Syst Biol* 2023;**19**:e11686. <https://doi.org/10.15252/msb.202311686>
40. Zhuang M, Achmon Y, Cao Y et al. Distribution of antibiotic resistance genes in the environment. *Environ Pollut* 2021;**285**:117402. <https://doi.org/10.1016/j.envpol.2021.117402>
41. Lipuma JJ, Currie BJ, Peacock SJ et al. Burkholderia, Stenotrophomonas, Ralstonia, Cupriavidus, Pandoraea, Brevundimonas, Comamonas, Delftia, and Acidovorax. In: Jorgensen JH, Carroll KC, Funke G, Pfaller MA (eds.), *Manual of Clinical Microbiology*. Hoboken, New Jersey, USA: John Wiley & Sons, Ltd, 2015, 791–812.
42. Andino A, Hanning I. Salmonella enterica: survival, colonization, and virulence differences among Serovars. *ScientificWorldJournal* 2015;**2015**:520179, 1–16. <https://doi.org/10.1155/2015/520179>
43. Pinn-Woodcock T, Frye E, Guarino C et al. A one-health review on brucellosis in the United States. *J Am Vet Med Assoc* 2023;**261**: 451–62. <https://doi.org/10.2460/javma.23.01.0033>
44. Hensel ME, Negron M, Arenas-Gamboa AM. Brucellosis in dogs and public health risk. *Emerg Infect Dis* 2018;**24**:1401–6. <https://doi.org/10.3201/eid2408.171171>
45. Urel H, Benassou S, Reska T et al. Nanopore- and AI-empowered metagenomic viability inference. *bioRxiv* 2024.06.10.598221. <https://doi.org/10.1101/2024.06.10.598221>