BMC Research Notes

DATA NOTE Open Access

# Precision metagenomics reveals microbial landscape in acute upper respiratory infections: a comprehensive dataset

Sadia Almas[3†], Rob E. Carpenter[1,2], Vaibhav K. Tamrakar[4†], Anuradha Singh[4], Aditya Sharma[4] and Rahul Sharma[3,4*]

## Abstract

**Objectives** The comprehension of the microbial composition in upper respiratory tract infections is pivotal for the progression of diagnostic and treatment methodologies. This article presents a dataset derived from Precision Metagenomic next-generation sequencing using hybridization capture-based targeted sequencing. Nasopharyngeal samples from 24 patients with acute URIs were analyzed using the Illumina®/IDbyDNA Respiratory Pathogen ID/AMR panel. The dataset contains a wealth of information on the composition of the microbiota, including the relative abundance of known pathogens and their potential clinical significance.

**Data description** This dataset serves as a valuable asset for future research in respiratory medicine, infectious disease epidemiology, antimicrobial resistance detection, and therapeutic interventions. Its potential for reuse and integration with other omics datasets enhances its significance. The comprehensive nature of the data facilitates research into relationships between the respiratory microbiota and host factors, including clinical outcomes, immune responses, or genetic predispositions. Moreover, the article underscores the interdisciplinary potential by advocating for the integration of this dataset with other relevant datasets such as transcriptomics or metabolomics, enabling a deeper understanding of the intricate interactions in acute upper respiratory infections. The presented dataset contributes to the expanding knowledge in precision metagenomics and holds the promise to propel research and clinical practices in the field of respiratory diseases.

**Keywords** Precision metagenomics, Next generation sequencing (NGS), Respiratory infections, Microbiome, Targeted sequencing

[†]Sadia Almas and Vaibhav K. Tamrakar are contributed equally to this work.

*Correspondence:
Rahul Sharma
rahuldnadx@gmail.com
[1]Advanta Genetics Research Laboratory, 10935 CR 159, Tyler, TX 75703, USA
[2]University of Texas at Tyler, 3900 University Boulevard, Tyler, TX 75799, USA
[3]Scienetix LLC, Tyler 75703, Texas, United States
[4]RetroBioTech And Research Pvt Limited, 309, kusum Vihar Jagatpura, JAIPUR, RAJ 75019, India

## Objective

The dataset offers a comprehensive metagenomic analysis of acute upper respiratory infections (URIs), serving as a valuable tool for researchers and clinicians [1]. It precisely identifies URI-related pathogens and categorizes them based on their potential to induce lower respiratory infections [2, 3].

In terms of clinical benefits, the dataset offers detailed information on the prevalence, abundance, and diversity of known microbial species in the upper respiratory tract. This aids in understanding microbial communities, illuminating the etiology and pathogenesis of URIs, and facilitates the detection of mixed infections. It guides tailored treatment strategies for complex respiratory diseases [4, 5].

Metagenomics enables detection and characterization of previously uncharacterized microorganisms, offering a potential avenue for biomarker discovery associated with different URIs [6]. This aspect enhances our understanding of respiratory infections and may lead to the development of new diagnostic tools or therapeutic interventions [7].

This study enriches research on respiratory diseases by uncovering complex host-microbe interactions and providing insight into immune responses and disease mechanisms. It identifies previously uncultivable or fastidious pathogens, contributing to the characterization of dysbiosis implicated in respiratory diseases, such as chronic obstructive pulmonary disease (COPD) and asthma [8].

Epidemiologists and public health experts benefit from this dataset by gaining insights into the epidemiology and transmission patterns of URI. Microbiota composition and lower respiratory trafficking potential data inform surveillance and outbreak investigations, contributing to epidemiological studies assessing infection spread, potential outbreaks, and transmission dynamics [9]. This dataset represents a paradigm shift in respiratory disease research and epidemiological investigation, guiding public health interventions and the development of targeted preventive strategies.

The insights into acute URI microbiota in the dataset contribute significantly to respiratory research, diagnosis, and treatment. Integration with omics datasets enhances its potential, driving optimization and advancements in microbiological laboratory workflows and clinical diagnostics [10]. This work collaboratively advances our understanding of URI-related complexities in respiratory medicine, infectious diseases, epidemiology, and public health.

## Data description

This study was conducted at Advanta Genetics in Tyler, TX, USA, from January 2022 to June 2022. In this study, 24 nasopharyngeal swab samples were collected from individuals undergoing PCR testing for respiratory pathogens. DNA and RNA, were extracted separately using Zymo Research reagents following the protocol outlined in the Explify Respiratory Pathogen ID/AMR Panel User Guide. T7 bacteriophage DNA was added to each sample to determine the absolute concentration of the target copies. Next-generation metagenomic sequencing (mNGS) libraries were prepared using the Illumina®/IDbyDNA Respiratory Pathogen ID/AMR Panel (RPIP) protocol [1]. The sequenced data were submitted to the National Center for Biotechnology Information (NCBI) in Fastq file format as SRA and Bio project-PRJNA995059 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA995059) [1] (Table 1).

## Methodology

A hybridization-capture-based metagenomic next-generation sequencing (mNGS) workflow was used to detect pathogens from clinical samples. The Illumina®/IDbyDNA Respiratory Pathogen ID/AMR Panel (RPIP) facilitated the sequencing of 24 clinical samples. DNA and RNA were extracted from each sample using Zymo Research reagents following the protocol outlined in the Explify Respiratory Pathogen ID/AMR Panel User Guide.

Each sample was spiked with T7 bacteriophage DNA (Microbiologics, MN, USA) to achieve a final concentration of 1.2 107 plaque-forming units (PFU/mL) to enhance the sensitivity of analysis. The mNGS libraries were prepared following the Illumina®/IDbyDNA Respiratory Pathogen ID/AMR Panel (RPIP) protocol and reagents (Illumina® San Diego, USA). The process involved extraction of RNA and DNA, preparation of cDNA from RNA and subsequent combination with DNA in equal volumes. DNA tagmentation and adapter ligation were performed using an Illumina® RNA Prep with an enrichment kit, leading to the construction of enriched libraries. These libraries were pooled, normalized to a concentration of 1 nM and neutralized before dilution to a loading concentration of 2 pM. Dual-indexed paired-end sequencing with a 75-bp read length was performed on an Illumina MiniSeq® instrument (Illumina® San Diego, USA) high-output flow cell (150 cycles). Although the intended sequencing depth was 1.0 million reads, samples with 0.5 million reads were included in the downstream analysis.

An automated IDbyDNA Explify® Platform data analysis solution (v1.0.1) was used to analyze the sequencing data. Subsequently, the mNGS results were analyzed, and microorganisms were categorized based on their potential to cause infection. Microorganisms were further classified into three phenotypic groups on the basis of their infectivity potential: microorganisms in phenotypic group 1 were categorized as part of the normal flora, colonizers, or contaminants; microorganisms in

**Table 1** All datasets presented in table have been deposited in the National center for biotechnology information (NCBI) under the bioproject accession number PRJNA995059 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA995059) [1]

| Label | Name of data file/ data set | File types (file extension) | Data repository and identifier (identifiers.org URL). |
|---|---|---|---|
| Data set 1 | SRR25428980 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428980) [11] |
| Data set 2 | SRR25428979 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428979) [12] |
| Data set 3 | SRR25428968 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428968) [13] |
| Data set 4 | SRR25428963 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428963) [14] |
| Data set 5 | SRR25428962 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428962) [15] |
| Data set 6 | SRR25428961 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428961) [16] |
| Data set 7 | SRR25428960 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428960) [17] |
| Data set 8 | SRR25428959 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428959) [18] |
| Data set 9 | SRR25428958 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428958) [19] |
| Data set 10 | SRR25428957 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428957) [20] |
| Data set 11 | SRR25428978 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428978) [21] |
| Data set 12 | SRR25428977 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428977) [22] |
| Data set 13 | SRR25428976 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428976) [23] |
| Data set 14 | SRR25428975 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428975) [24] |
| Data set 15 | SRR25428974 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428974) [25] |
| Data set 16 | SRR25428973 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428973) [26] |
| Data set 17 | SRR25428972 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428972) [27] |
| Data set 18 | SRR25428971 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428971) [28] |
| Data set 19 | SRR25428970 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428970) [28] |
| Data set 20 | SRR25428969 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428969) [29] |
| Data set 21 | SRR25428967 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428967) [30] |
| Data set 22 | SRR25428966 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428966) [31] |
| Data set 23 | SRR25428965 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428965) [32] |
| Data set 24 | SRR25428964 | Fastq (.gz) | NCBI Sequence Read archive, (http://identifiers.org/insdc.sra:SRR25428964) [33] |

The bioproject accession serves as a unique identifier for the collection of respiratory Microbiome data included in this table

phenotypic group 2 were frequently associated with respiratory disease; and microorganisms in phenotypic group 3 were deemed pathogenic for respiratory disease. In addition, raw fastq files were analyzed and interpreted to determine the cut-off values based on the specified parameters. Samples with fewer than 0.5 million reads post-sequencing were excluded from further analysis. Analysing the sequencing data specific criteria were applied: >40% target coverage, a median depth of 1X, and RPKM > 10 in the mNGS assay were required for positive classification.

## Limitations

While our published manuscript provides valuable insights into the microbiota of acute upper respiratory infections using a comparative analysis of PCR and metagenomics sequencing methods, it is important to note that the antimicrobial resistance (AMR) results obtained from the 24 nasopharyngeal swab samples through mNGS were not subjected to detailed analysis, thereby limiting our exploration of potential implications in the context of lower respiratory trafficking potential.

### Abbreviations

URIs Upper respiratory infections
NGS Next generation sequencing
mNGS Next-generation metagenomic sequencing
p(mNGS) Precision metagenomic next-generation sequencing
RPIP Respiratory pathogen ID/AMR panel
SRA NCBI-sequence read archive
COPD Chronic obstructive pulmonary disease
RPKM Reads per kilobase million

### Acknowledgements

The reference strains of microbial pathogens were obtained from BEI Resources, NIAID.

### Author contributions

RE.C. Conceptualization and editing, S.A. Methodology and Project administration, VK. T. Data curation, Writing - Original Draft & editing, A.S. and A.S. Data curation and Formal analysis, R.S. Conceptualization, Supervision and Writing.

### Funding

This research received no specific grant from public, commercial, or not-for-profit funding agencies.

### Data availability

The sample-specific fastq files are submitted to the NCBI as Sequence Read Archive (SRA) Data identification number: PRJNA995059. Direct URL to data: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA995059 [1] and mNGS-run BLC files and demultiplexing barcode information are available upon reasonable request.

## Ethics declarations

### Ethics approval and consent to participate

The study was exempted by the institutional review board (IRB) of Advanta Genetics, Tyler, TX, USA. Only de-identified residual samples discarded after routine clinical testing were used in this study.

### Consent for publication

Patient consent was not applicable due to research conducted on de-identified left over samples from routine diagnostics.

### Competing interests

The authors declare no competing interests.

Received: 9 February 2024 / Accepted: 22 September 2025
Published online: 10 December 2025

### References

1. Respiratory Microbiome Database. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA995059. Accessed 20 March 2024.
2. Almas S, et al. Deciphering microbiota of acute upper respiratory infections: a comparative analysis of PCR and mNGS methods for lower respiratory trafficking potential. Adv Respir Med. 2023;91(1):49–65.
3. d'Humières C, et al. The potential role of clinical metagenomics in infectious diseases: therapeutic perspectives. Drugs. 2021;81(13):1453–66.
4. Gaston DC, et al. Evaluation of metagenomic and targeted next-generation sequencing workflows for detection of respiratory pathogens from bronchoalveolar lavage fluid specimens. J Clin Microbiol. 2022;60(7):e0052622.
5. Diao Z, et al. Metagenomics next-generation sequencing tests take the stage in the diagnosis of lower respiratory tract infections. J Adv Res. 2022;38:201–12.
6. Alnimr A. Antimicrobial resistance in ventilator-associated pneumonia: predictive microbiology and evidence-based therapy. Infect Dis Ther. 2023;12(6):1527–52.
7. Carpenter RE, et al. SARS-CoV-2 next generation sequencing (NGS) data from clinical isolates from the East Texas region of the United States. Data Brief. 2023;49:109312.
8. Tiew PY, Meldrum OW, Chotirmall SH. Applying next-generation sequencing and multi-omics in chronic obstructive pulmonary disease. Int J Mol Sci. 2023. https://doi.org/10.3390/ijms24032955.
9. Carpenter RE, et al. Confirming multiplex RT-qPCR use in COVID-19 with next-generation sequencing: strategies for epidemiological advantage. Glob Health Epidemiol Genom. 2022;2022:2270965.
10. Whiteside SA, McGinniss JE, Collman RG. The lung microbiome: progress and promise. J Clin Invest. 2021. https://doi.org/10.1172/JCI150473.
11. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428980.
12. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428979.
13. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428968.
14. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428963.
15. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428962.
16. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428961.
17. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428960.
18. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428959.
19. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428958.
20. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428957.
21. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428978.
22. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428977.
23. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428976.
24. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428975.
25. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428974.
26. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428973.
27. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428972.
28. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428971.
29. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428969.
30. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428967.
31. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428966.
32. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428965.
33. Almas S, Carpenter RE, Tamrakar VK, Singh A, Sharma A, Sharma R. Respiratory Microbiome Dataset. 2023. NCBI Sequence Read Archive. http://identifiers.org/insdc.sra:SRR25428964.

## Publisher's note