IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, June 2025
Impact Factor: 7.67

Unveiling Airborne Microbiomes through Metagenomic Sequencing

Divyansh Verma[1], Riya Srivastav[2], Rachna Chaturvedi[1] and Jyoti Prakash[1*]
Amity Institute of Biotechnology, Amity University, Lucknow, UP, India[1]
Kishan Lal Public College, Rewari, Haryana[2]
*Corresponding Author: jprakash@lko.amity.edu
ORCID: - 0000-0002-9326-355X

Abstract: *This study aimed to characterise the variety of bacteria present in air samples collected in the Sahara Desert, Barbados, Cyprus, and during the Air Quality and Climate Change in the Arabian Basin (AQABA) ship voyage that travelled across the Mediterranean Sea and around the Arabian Peninsula. A metagenomic analysis was conducted using 16s rRNA gene sequences from the MGnify database (study ID: PRJEB26788). Human health, atmospheric chemistry, and environmental processes are all significantly influenced by airborne microbial communities. The microbial diversity and composition of air samples taken from a variety of sites, including desert, marine, and urban settings, were examined in this work using metagenomic analysis. To identify important bacterial taxa including Streptomyces, Nocardioidaceae, Burkholderiaceae, and Mycobacteriaceae, we processed and analysed sequencing data using programs like FastQC, Trim Galore, Kraken2, and Krona. The findings demonstrated the impact of environmental conditions on airborne microbiomes by revealing a notable variance in microbial abundance between locales.Our results also imply that long-distance atmospheric transport plays a role in the worldwide spread of microbial species, which may have consequences for the stability of ecosystems and the spread of disease. This research highlights the importance of metagenomics for high-resolution microbial community analysis and emphasizes the need for continued surveillance of airborne microorganisms to better understand their ecological functions and health implications.*

Keywords: 16S rRNA gene sequences, airborne, metagenomics, analysis.

I. INTRODUCTION

A wide variety of microbial communities, such as bacteria, fungi, viruses, and other airborne particles, are abundant in the dynamic and complex ecosystem that is air. Since these airborne microbial populations are important for ecological processes, public health, and atmospheric chemistry, it is imperative to comprehend their makeup and functions.[1].Because many species are difficult to cultivate in typical laboratory settings, traditional approaches to researching airborne microorganisms, such as culture-based techniques, only offer a partial picture of microbial diversity. This restriction has prompted the use of metagenomic techniques to thoroughly examine the microbial composition of air samples.[2].

By directly extracting and sequencing DNA from environmental samples, metagenomics eliminates the requirement for cultivation and makes it possible to identify both known and unidentified microbes.[3].For the investigation of air microflora, many metagenomic methods are frequently employed, including metatranscriptomics, 16S/18S rRNA gene amplicon sequencing, and shotgun metagenomic sequencing. While amplicon sequencing focuses on particular marker genes to determine taxonomic composition, shotgun sequencing captures all of the genetic material in a sample to provide a thorough picture of the microbial community.[4].However, by exposing the functional activity of airborne microorganisms, metatranscriptomics aids researchers in comprehending the metabolic pathways and possible virulence factors found in the air microbiome.[5].

The various microbial species that makeup air microflora come from a variety of sources, including soil, water, plants, human activity, and industrial emissions. These microbes can affect biogeochemical cycles, affect atmospheric

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, June 2025
Impact Factor: 7.67

processes, and hurt human health through pathogenic or allergic interactions.[6].Temperature, humidity, and geographic location are some of the environmental variables that affect the microbial diversity in the air. Mapping these dynamic microbial communities and determining their involvement in disease transmission and ecosystem resilience has been made possible in large part by metagenomic research.[7].

Despite its revolutionary potential, metagenomic analysis of air samples has many drawbacks, such as low biomass, contamination hazards, and the requirement for well-designed DNA extraction and sampling procedures.[8]. Our capacity to analyse metagenomic data is being improved by developments in bioinformatics pipelines and sequencing technologies, which offer a more profound understanding of the ecology of airborne microbes and make it easier to identify disease outbreaks early.[9].

II. MATERIALS AND METHODOLOGY

This study(**PRJEB26788**) investigates the microbial composition and metabolic potential of filter, soil dust, and biocrust samples taken in Barbados, Cyprus, the Sahara Desert, and during the Air Quality and Climate Change in the Arabian Basin (AQABA) ship expedition that travelled through the Mediterranean Sea and around the Arabian Peninsula using metagenomic 16S rRNA gene sequencing data from the MGnify database.

Dust, which travels across regional and even transcontinental distances as aerosols, is believed to be mostly produced in deserts. Dust has been linked to the development of respiratory, allergic, and cardiovascular disorders, demonstrating its detrimental effects on human health. There is evidence that dust particles can carry bacteria, viruses, protozoa, archaea, and fungal spores.

Using a systematic procedure, the main goal of this study was to identify and characterise the bacterial flora found in these air samples. The bioinformatics tools used to study and analyse the microbial flora are: 1) **FASTQC-**Researchers can evaluate the quality of high-throughput sequencing reads before downstream analysis by using FASTQC, a popular quality control tool that offers a thorough overview of raw sequencing data. It produces comprehensive reports on several quality measures, such as the distribution of sequence lengths, GC content, per-base sequence quality, and the existence of adapter sequences or over-represented k-mers. FASTQC becomes much more accessible when coupled with the Galaxy server, allowing users to perform quality checks on sequencing data via an intuitive web-based interface without requiring intricate command-line procedures[10].The data pretreatment step is streamlined by Galaxy's interactive procedures, which let researchers see FASTQC results and decide in real-time whether to trim or filter data. This combination of FASTQC and Galaxy helps researchers confirm the integrity of their sequencing data, thus boosting the dependability of subsequent analyses[11].2)**TRIM GALORE-** Trim Galore is a popular bioinformatics program for preprocessing and quality control of high-throughput sequencing data. It automates adapter trimming and quality filtering by combining the features of Cutadapt and FastQC, guaranteeing dependable and clean sequences for further investigation[12]. With user-defined thresholds, Trim Galore can identify and eliminate low-quality bases and residual adapter sequences from both single-end and paired-end reads. It is particularly helpful for epigenomics research since it can also manage the trimming of biased sequences, such as methylation-specific adapters in bisulfite sequencing. When Trim Galore is combined with the Galaxy server, a web-based platform for biomedical research involving large amounts of data, it provides researchers with an easy-to-use interface that doesn't require complicated command-line inputs to accomplish thorough data cleaning.[13]. 3) **KRAKEN2-** A popular taxonomy classification tool for metagenomic investigation, Kraken2 allows for the quick and precise identification of microbial genomes from intricate environmental samples[14]. By comparing brief DNAsequences to a pre-built database using a k-mer-based methodology, categorisation is quick and memory-efficient. Through an intuitive interface, Kraken2 is available on the Galaxy server, enabling researchers to examine sequences without the need for sophisticated computing abilities. The technology is particularly helpful for air microbiome investigations since it allows users to customise databases, which allows them to customise their analysis to certain microbial communities or environmental niches. Following classification, Galaxy's suite of data analysis tools allows users to further visualise and analyse the comprehensive reports that Kraken2 produces, which include taxonomic abundance and confidence scores. Researchers may now more easily and accurately investigate microbial diversity and its ecological implications thanks to this integration, which simplifies the metagenomic approach.[13]. 4) **CONVERT KRAKEN-**The CONVERT KRAKEN

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, June 2025
Impact Factor: 7.67

tool, which makes it easier to convert Kraken output files into more comprehensible forms, is one of the robust arrays of tools for metagenomic data analysis that researchers may access through the Galaxy platform. Although Kraken is a popular taxonomy categorisation method for metagenomics, it can be difficult to understand its unprocessed output. By reformatting classification results into easier-to-understand tables and reports, the CONVERT KRAKEN tool facilitates downstream analysis and helps researchers visualise and comprehend the makeup of microbial communities. The productivity of metagenomic processes is improved by this tool's seamless integration with other Galaxy tools and reduction of result complexity, which makes it particularly useful when working with huge datasets[14]. 5) **KRONA PIE CHART-** Metagenomic data interpretation requires visualisation tools, and the Krona pie chart—accessible via the Galaxy server—is an effective tool for this. Researchers may easily examine the hierarchical structure of complicated microbial communities with Krona's interactive, multi-layered pie charts. Because it graphically depicts relative abundances and phylogenetic relationships within a sample, this tool is particularly useful for analysing taxonomic classifications using sequencing data. Even researchers with less bioinformatics experience can utilise the Galaxy server since it offers an intuitive interface for merging Krona results and carrying out metagenomic operations. Scientists may easily browse massive metagenomic datasets by using Krona pie charts. This makes it easier to identify rare and dominant taxa in airborne microbial populations, which is essential for comprehending the effects on the environment and human health[15].

RESULT

The composition of the bacterial communities in the three datasets was revealed by the metagenomic analysis of air microbiomes. Trim Galore effectively removed adapters and low-quality reads, and quality evaluation using FastQC confirmed the high integrity of the sequencing data. Kraken 2 was used for taxonomic categorisation to identify the primary bacterial taxa present in each dataset, with notable variations in microbial diversity. Krona pie charts facilitated a comparative study by visualising differences in the relative abundances of significant microbiological groupings.

Sample 1

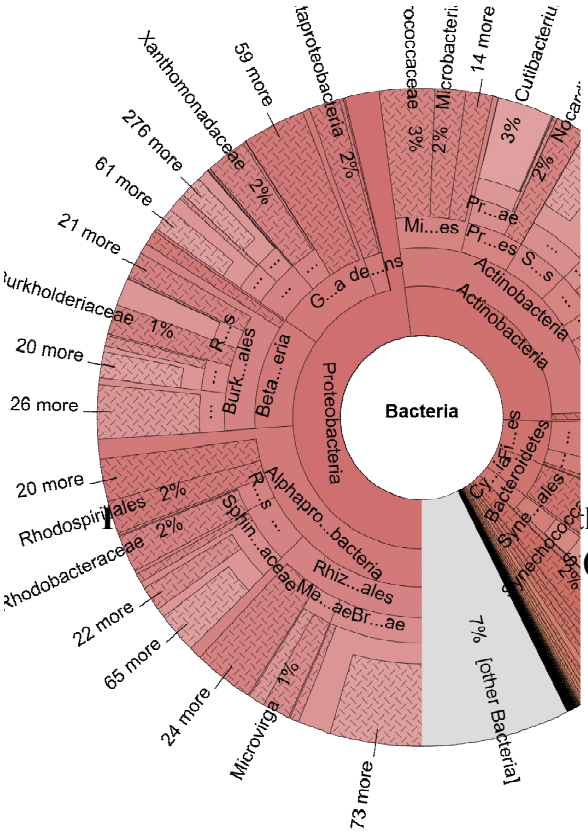

Figure 1: The air microbial community's (Sample-SAMEA4668129)

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, June 2025
Impact Factor: 7.67

Table 1: **The relative abundance of the four most prevalent bacterial taxa in Dataset 1**

S.No	Bacteria	Percentage
1	*Cutibacteriumacnes*	3% (Opportunistic pathogen)
2	Streptomyces	4%
3	Micrococcaceae	3%
4	Nocardioidaceae	2% (Opportunistic pathogen)

Streptomyces (4%), a bacterium species renowned for producing antibiotics, dominated the varied bacterial community identified by the metagenomic study of Dataset 1. Notable presences of two possible opportunistic infections, *Cutibacterium acnes* (3%) and *Nocardioidaceae* (2%) were also observed. *Micrococcaceae* (3%) were prevalent bacteria found in the environment and on the skin. The air sample contains a variety of environmental and human-associated species, according to the relative abundance values. This implies that human activity and natural sources may have an impact on the microbial content of the air.

REFERENCES

[1] P. e. a. Amato, "Microbial communities in the atmosphere and their potential impacts on the environment and human health," *Frontiers in Microbiology,* pp. 8, 738., 2017.

[2] R. M. e. a. Bowers, " Characterization of airborne microbial communities at a high-elevation site and their potential to act as atmospheric ice nuclei," *Applied and Environmental Microbiology, 75(15),* p. 5121–5130., 2009.

[3] J. Handelsman, "Metagenomics: Application of genomics to uncultured microorganisms," *Microbiology and Molecular Biology Reviews, 68(4),* p. 669–685., 2004.

[4] N. e. a. Shah, "Metagenomics for air quality monitoring: Current status and future perspectives," *Environmental Science & Technology, 53(4),* p. 1960–1972., 2019.

[5] A. e. a. Franzetti, "Airborne microbial communities and their functional profiles in urban environments," *Science of the Total Environment, 721,* p. 137918., 2020.

[6] V. e. a. Bertolini, "Temporal variability and effect of environmental variables on airborne bacterial communities in an urban area of northern Italy," *Applied Microbiology and Biotechnology, 97(14),* p. 6561–6570., 2013.

[7] V. R. e. a. Després, "Primary biological aerosol particles in the atmosphere: A review. Tellus B: Chemical and Physical Meteorology, 64," *Tellus B: Chemical and Physical Meteorology, 64,* p. 15598, 2012.

[8] S. e. a. Yooseph, "A metagenomic framework for the study of airborne microbial communities," *PLoS ONE, 8(12),* p. e81862., 2013.

[9] H. e. a. Behzad, "The world of the viral air microbiome," *Frontiers in Microbiology, 9,* p. 2333., 2018.

[10] E. e. a. Afgan, "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update," *Nucleic Acids Research, 46(W1),* p. W537–W544., 2018.

[11] D. e. a. Blankenberg, "Galaxy: A web-based genome analysis tool for experimentalists," *Current Protocols in Molecular Biology, 89(1),* p. 19.10.1–19.10.21., 2010.

[12] F. Krueger, "Trim Galore: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files," *Babraham Bioinformatics..*

[13] E. e. a. Afgan, "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update," *Nucleic Acids Research, 46(W1),* p. W537–W544., 2018.

[14] D. E. e. a. Wood, "Improved metagenomic analysis with Kraken 2," *Genome Biology, 20(1),* p. 257, 2019.

[15] B. D. e. a. Ondov, " Interactive metagenomic visualization in a web browser," *BMC Bioinformatics, 12,* p. 385., 2011.