Characterization of Biological Aerosols Over Northeast India: A Metagenomic Approach

Binita Pathak

Dibrugarh University

Ankita Khataniar

Dibrugarh University Centre for Studies in Biotechnology: Dibrugarh University Centre for Biotechnology and Bioinformatics

Barlin Das

Dibrugarh University

Sristisri Upadhyaya

Dergaon Kamal Dowerah College

Ankita Medhi

University of Delhi Miranda House

Pradip Kumar Bhuyan

Dibrugarh University

Alak Kumar Buragohain

Royal Global University

Debajit Borah (✉ dborah89@gmail.com)

Royal Global University

Research Article

Keywords: Bacterial diversity, metagenome analysis, air quality, human health, bioaerosol characterization

Posted Date: December 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1081395/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Northeast India is considered as one of the major biodiversity hotspots in the world but the region is underexplored for their microbial biodiversity. Extensive characterization of biological aerosol (bioaerosol) samples collected from various locations of Northeast India was carried out for all the four seasons in a year. These were characterized in terms of particulate matters (inhalable, thoracic, and alveolic), their constituents (pollens, fungal spores, animal debris, and non-biological components), and finally the bacterial diversity was determined by DNA based metagenomic approach. The non-biological (non-viable) component of aerosols is found to vary from 30- 89% in pre-monsoon season which coexists with pollens (4-20%), animal debris (1-24%) and fungal spores (1-17%). The highest number of culturable microbial population in terms of CFU count was observed in the samples collected in pre-monsoon season (*i.e.*, 125.24-632.45 CFU/m³) and the lowest CFU was observed in monsoon season (*i.e.*, 20.83-319.0 CFU/m³). The metagenomic approach with the samples collected during pre-monsoon season showed a total of bacterial 184 OTUs (operational taxonomic units) with 28,028 abundance count comprising with 7 major phylum, 6 classes, 10 orders, 15 families, 13 genus, and 8 species of bacteria. The species level distribution clearly shows the presence of Gammaproteobacteria (52%) most abundantly followed by Bacilli (21%), Alphaproteobacteria (14%), Betaproteobacteria (5%, Flavobacteria (5%), and Sphingobacteria (3%). It is the first report from entire Northeast India to uncover bacterial diversity in aerosol samples through DNA based metagenomic approach.

Introduction

Bioaerosols are the suspension of airborne particles containing living organisms, dead debris or living components of biological origin with or without being attached to non-biological particles and the by-products of biological activities. About 20-40% of atmospheric aerosols are of biological origin (Jaenicke, 2005). These include a wide range of organic matter with differences in physical and chemical characteristics and surface properties (Georgakopoulos et al., 2009). Bioaerosols are central elements in the development, evolution, and dynamics of ecosystems and play a major role in the interactions between atmosphere, biosphere, climate, and public health. These influence the Earth's energy budget by both direct and indirect effects and thus influencing the climate and the hydrological cycle. For example, Pollen grains attract water at relative humidity well below 100% and thus might locally act as CCN influencing the cloud formation; Decaying vegetation (including associated bacteria) and marine plankton (Jaenicke 2005), Proteins (Scheel et al., 2016) and Fungi can act as efficient ice nuclei. Bioaerosols significantly influence both indoor and outdoor air quality. These are linked with spreading biological organisms and reproductive materials (Bundy et al. 2009; Perez-Padilla et al. 2009) and many different adverse health effects like infectious diseases, allergies, asthma, cancer, respiratory disorders in human (Kim et al 2017, Oh et al 2020). In the desert region, microbes present in dust pose a major public health issue.

Over the years, the field of aerobiology has explored the abundance, diversity, survival and transport of micro-organisms in the atmosphere. Microbes are probably the most successful and abundant life forms

on earth which can survive from the bottom of deep oceans up to the altitude of 50 km in the stratosphere (Sarma et al., 2020), which attracts the Astrobiologists. Variation of drift in airflow may be one of the most important factors that drive microbes from one location to another creating a great diversity in bioaerosols of various places on earth (Aalismail et al., 2019). Even though, bioaerosols provides extreme environment in terms of low nutrients, moisture content, and higher exposure rate to solar radiation *etc.* which may limit the viability of a large number of microbes, but a fraction of microbial populations are known to be capable of growing and proliferating under such condition (Burrows et al., 2009; Womack et al., 2010). Airborne bacteria represent a high risk to public health. They mostly cause severe breathing related diseases such as asthma, skin rashes, allergies and many more (Arancibia et al., 2002). Therefore, identification of various air borne microbes becomes the need of the hour for the prevention and control of various disease-causing microbes. Culture-based identification approach of microorganisms from bioaerosols have shown the presence of mostly Gram positive bacteria out of which *Streptococcus* sp., *Staphylococcus* sp., *Micropococcus* sp., *Bacillus* sp., and *Corynebacterium* sp. *etc.* are the most common all across the globe (Shin et al., 2015). On the other hand, Gram negative microbes such as *Acenatobacterium* sp., *Pseudomonas* sp., and *Moraxella* sp. *etc.* are also reported worldwide but are less frequent as compared to Gram positive microbes (Andersson et al., 1999). Other living microbial entities present in bioaerosol includes various fungal spores such as *Aspergillus* sp., *Penicillium* sp., *Cladosporium* sp., and *Alternaria* sp. *etc.* (Shin et al., 2015). Nevertheless, microbes are underexplored by aerobiologists (<99%) due to the limitations of conventional culturing methods of aerosol borne microbes; as growth medium sometime acts as recalcitrant and restrict their proper growth (Hugenholtz et al., 1998). Considering the above limitations, molecular identification techniques have become popular nowadays for their accuracy and culture independent approach (Shin et al., 2015). In very recent days, metagenome approach of bacterial identification becomes a new trend that enables researchers to identify microbial diversities based on DNA based techniques without culturing the microbes (Serrano-Silva and Calderon-Ezquerro 2018). Metagenome based analysis of microbial diversity involves direct isolation of DNA from microbes without culturing them and it provides an insight view of various microbes present in the sample using bioinformatics tools (Yo et al., 2017). However, the success of the process is also largely dependent on the instrument used for the collection of bioaerosol samples and the quantity and quality of the DNA extracted from the collected samples as extraction process demands very high level of precision and expertise.

Bioaerosols can be sampled by using a wide range of methods which involves the use of various traditional and modern devices to collect airborne microorganisms both culturable and non-culturable (Yoo et al. 2017). A particular bioaerosol sampler may not be fully sufficient to satisfy all the required experimental outcomes such as analyzing real time particle size distribution and complete microbial diversity analysis together at a time but the involvement of metagenome analysis of samples may provide the maximum possible insight into the wide range of microbial diversity in the samples in a cost effective and less labor intensive way. In view of this, metagenome analysis of bioaerosols have become popular all across the globe to get a better insight into the characterization of bioaerosols in terms of exploration of microbial diversity but are very limited till now. Researchers have analyzed bioaerosol

samples collected from indoor as well as outdoor environment (Shin et al., 2015), above sea (Aalismail et al., 2019), earth's lower atmosphere (Smith et al., 2018), livestock farms (Wang et al., 2019), urban spaces (Be et al., 2015), waste sorting plants (Degois et al., 2017) *etc.* Considering the Indian perspective, studies on airborne bacterial species through conventional approach is available for some of the regions such as Mumbai (Gangamma 2014), Delhi (Kumar et al., 2013), Bangalore (Nandini and Sivasaktivel 2014) *etc.* Further, systematic survey on pollen influence on human health was initiated in India long back at Calcutta in 1873 (Pathak et al., 2020). Later on, researchers have initiated exhaustive studies on airborne pollen types and their concentration. Pollen calendars prepared from systematic measurements of pollen at various places are very useful for clinicians as well as allergic patients to establish chronologic correlation between the concentration of pollen in atmosphere and seasonal allergic symptoms (Singh *et al.*, 1992). As an outcome of these studies, some important pollen and fungal allergens were identified, quantified and characterized for their allergenic properties from 18 different places in India.

Despite the fact that India is the world's seventh largest country in terms of total land area which covers rain-forests, deserts, grasslands, terrains *etc.* and northeastern part of the country is considered as one of the major biodiversity hotspots in the world with largely underexplored microflora and fauna, but only a handful of reports are available on the analysis of microbial diversity in bioaerosols applying metagenome approach (Chakrawarti et al., 2020). Hence, extensive research in this field is the need of the hour in order to expand the periphery of our knowledge on microbial diversities in bioaerosols. Additionally North East India is vulnerable to both primary and secondary aerosols of biogenic origin due to high vegetation cover (~66%). Biological sources are found to emit ~ 1150 teragrams of carbon per year globally in the form of VOCs. The majority of VOCs are produced by plants, the main compound being isoprene and the remaining is produced by animals, microbes and fungi (such as molds). But, research on bio-aerosols has not gained due importance so far in North East India, in the context of air pollution, climate impacts and health impacts in spite of the growing concern on various diseases caused by seasonal growth of primary bio-aerosols like bacteria, fungi, pollen, *etc.* Thus, the present work supplements to new knowledge in this regard on a regional basis over the North-East India.

Materials And Methods

Study Region and prevailing meteorology:

Bioaerosol samples were collected from nine discrete locations of Northeast India (**Figure 1**) in a campaign mode. The locations were selected based on urbanization, vegetation type and accessibility (Table **S1**). It is worthwhile to mention here that the tropical forests have been proposed as significant sources of bioaerosols (Jaenicke 2005). A full year (2018-2019) of campaign was conducted, covering all the seasons to explore the seasonality in abundance and type of bioaerosols present in the atmosphere and thus to obtain the regional mapping of bioaerosols. All the samples were collected within 6:00 hr to 18:00 hr. The climate of the study region is Subtropical humid in nature and a year is categorized into four seasons: pre-monsoon (March-May), monsoon (June-September), post monsoon (October-

November) and winter (December-February). A portable Automatic weather station (PORTLOG AWS, USA) was used for the simultaneous measurement of meteorological parameters such as Relative Humidity, Temperature, wind speed and direction etc. (Table S2).

Bioaerosol sampling and characterization:

A Burkard Pollen sampler was used for the collection of pollens in the designated locations with an intake air flow rate of 10 LPM (Núñez *et al.*, 2016). It contains a greased rotating drum at a speed of 1 rotation per 7 days in order to collect pollens and other aerosols through impaction. These were characterized under fluorescence microscope for identification based on their physical characteristics such as shape, size and other morphological features using acridine orange which stains the nucleic acids present in the samples.

A portable Bio-Aerosol Spectrometer (GRIMM Aerosol Technik Ainring GmbH & Co.KG., Germany, model 11.A) was used for real time continuous monitoring of airborne particles (non-biological and biological together) in terms of particle count distribution at 31 size intervals (Peters *et al.*, 2006), at a temporal resolution of 5 minutes and flow rate of 1.2 liters/min. The principle of operation of the spectrometer is available elsewhere (Pathak *et al.*, 2020). The particulate matter (PM) concentrations (in $\mu\text{g}/\text{m}^3$) measured by the instrument can be further categorized into inhalable matter ($\text{PM} \leq 32 \mu\text{m}$ for the bioaerosol spectrometer, $\leq 100 \mu\text{m}$ in general), thoracic ($\text{PM} \leq 10 \mu\text{m}$) and alveoli (alveolar, $\text{PM} \leq 4 \mu\text{m}$), that may be absorbed in the nasal tract; thorax and alveoli respectively of human body (Petavratzi *et al.*, 2005).

A greased glass slide was used as an impactor which collects aerosols in the form of very thin dusty spot which was used for culturing as well as direct extraction of DNA for metagenome analysis. Culturing of microbes from bioaerosol samples was done by overlay method (Aneja, 2010) by gently placing the slides on nutrient and potato dextrose agar followed by incubation at 37 °C for the isolation of bacteria and fungi respectively. The number of bacteria or fungi *i.e.*, the microbial concentration in the samples can be determined in terms of the colony forming unit (CFU) (Srivastava *et al.*, 2012; Lal *et al.*, 2013) which is defined as:

$$\text{Microbial concentration (CFU/m}^3\text{)} = \frac{\text{Number of colonies obtained after incubation}}{[\text{Flow rate (}\frac{\text{m}^3}{\text{min}}\text{)} \times \text{sampling duration (min)}]}$$

Health risk assessment of isolated microbes in terms of their multidrug resistance pattern and haemolytic ability:

The Kirby Bauer's disc diffusion method was used to screen out multidrug resistant microbes present in bioaerosol samples. The bacterial broth of each bacterial isolates was spread plated for overnight on Muller-Hinton agar (g/L: Beef extract-2.00, acid hydrolysate of Casein-17.50, starch-1.50, agar agar-17.00, at pH-7.4 $\pm$ 0.2 at 25°C). Commercially available standard discs impregnated with designated amount of narrow (*viz.*, Vancomycin-30 mcg and Erythromycin-15 mcg) and broad spectrum antibiotic (*viz.*,

Chloramphenicol-30 mcg, and Ampicillin-10 mcg) placed as per the prescribed guidelines of Clinical and Laboratory Standards Institute (USA) guidelines (Sweeney 2018). These were then incubated at 37° C for 24 h. The clear zones around the antibiotic discs were considered as sensitive zones.

Isolated microbes were individually inoculated on blood agar supplemented with commercially available defibrinated sheep blood followed by incubation at 37° C for 24 h to observe their haemolytic ability.

Molecular identification of bacteria present in bioaerosol samples:

Sample processing:

The environmental sample was mixed in Phosphate Buffered Saline solution in order to obtain a homogeneous representation of each organism. DNA was extracted from the suspension using DNeasy Microbial Kit (Qiagen Ltd, Strasse, Germany) following the manufacturer's instructions (Qiagen Ltd, Strasse, Germany).

16S rDNA gene sequencing:

Partial 16S rRNA gene sequences were amplified from extracted DNA using 16S gene specific primers for V3-V4 region as per Illumina 16S V3-V4 sequencing kit that also includes addition of adapter overhang nucleotide sequences to the partial 16S rRNA gene-specific amplicons. The amplicons were further processed by employing the 16S Metagenomic Sequencing Library Preparation Protocol (Part no. 15044223 Rev. B-Illumina). The PCR conditions used were 5 min at 95 °C, 35 cycles of 30 s at 94 °C, 30 s at 55 °C and 90 s at 72 °C, followed by 10 min at 72 °C. Amplifications were carried out using a Verity Thermocycler (Applied Biosystems). The integrity of the PCR amplicons was analysed by electrophoresis on a 2200 TapeStation Instrument (Agilent Technologies, USA).

MiSeq sequencing of 16S rRNA gene-based amplicons:

PCR products obtained following amplification of part of the 16S rRNA gene sequences were purified by a magnetic purification step involving Agencourt AMPure XP DNA purification beads (Beckman Coulter Genomics GmbH, Bernried, Germany) in order to remove primer dimers. The DNA concentration of the amplified sequence library was estimated through fluorimetric Qubit quantification system (Life Technologies). Amplicons were diluted to 4 nM and 5 µL of each diluted DNA amplicon sample was mixed to prepare the pooled final library. Sequencing was performed using an Illumina MiSeq sequencer with MiSeq Reagent Kit v3 chemicals.

Analysis of 16S rRNA microbial data:

The fastq files were processed using MG-RAST (Keegan et al., 2016). Paired-end reads were merged, and quality control implementation allowed the retention of sequences with a length between 400 to 450 bp, mean sequence quality score >30 and with truncation of a sequence at the first base if a low quality within a rolling 10-bp window was found. Sequences with mismatched forward and/or reverse primers

were omitted. 16S rRNA operational taxonomic units (OTUs) were defined at $\geq 97\%$ sequence homology using M5NR hits. All reads were classified to the lowest possible taxonomic rank using ribosomal RNA databases. Abundance profile for each OTU was computed from the expanded similarities.

All the 10 Samples were subjected to pre-processing of reads, de-replication, singleton removal, OTU clustering, and chimera filtering. Each OTU's annotation at species level was carried out using scientifically accepted tools and pipelines *i.e.*, QIIME2 release 2020.6 and Green gene Database (default for QIIME)(g-13-8-99-515-806-nb-classifier) respectively.

DADA2 plugin denoising step from QIIME2 converts the overlapping paired-end reads into consensus sequence of V3-V4 region. It finds the overlap for each pair and combines them into a single read. For quality control of the sequences, the DADA2 plugin in Qiime 2 was used to associate erroneous sequence reads with the true biological sequence from which they were derived, thus producing high-quality sequence variant data. Using DADA2, all reads were trimmed to 260 bp, based on the median quality score. In addition, chimeric sequences were detected and excluded from analyses. 16SrRNA Operational Taxonomic Units (OTUs) were picked from the Illumina reads using a closed-reference OTU picking protocol against the Greengenes database (<https://data.qiime2.org/2020.6/common/gg-13-8-99-515-806-nb-classifier.qza>). In the next step, taxonomy assignments were associated with OTUs based on the taxonomy associated with the Green genes reference sequence defining each OTU.

Data availability:

All 10 annotated gene sequences in FASTA format were deposited in NCBI GenBank database and the following accession number SAMN21545970 has been successfully received.

Results And Discussion

The meteorological conditions such as temperature (T), relative humidity (RH), wind speed (km/h) largely varied in all the sample collection sites during different seasons (Table **S2**). A strong seasonality is observed in the average concentration of bacteria in terms of colony forming unit in air, with highest value in pre-monsoon season (125.24-632.45 CFU/m³) and the lowest was observed during the monsoon (20.83- 319.0 CFU/m³) (**Figure 2**). The concentration of culturable bacterial population in air samples ranges from 45.5 - 645.84 CFU/m³ in the post monsoon season and 118.06-458.33 CFU/m³ in winter. During pre-monsoon, minimum bacterial concentration was observed in the sample collected in Guwahati, an urban location and the maximum was observed in Dibru-Saikhowa, a rainforest area with higher vegetation cover. Overall, spatially higher CFU values were observed in the highly vegetated locations: Shillong, Kaziranga, Dibru Saikhowa and Joypur. On the other hand, CFU values are lower in urban and semi urban locations (*i.e.*, Guwahati, Jorhat, Dimapur, Namsai and Dibrugarh) with less vegetation cover (**Figure 2**). However, in a unit volume of air, existence of these microbial species is minimal and only upto ~0.20% of CFU counts are found in PM10 counts in pre-monsoon season at Dibru-Saikhuwa followed by Joypur and Namsai. PM10 counts were maximum in Dimapur, an urban location

in winter reaching up to ~3 Millions and followed by post-monsoon and it lies within 1 million at other locations throughout all the seasons.

Particulate matter (PM), on the other hand is one of the major criteria of pollutants that affects air quality in urban and even rural areas in the world. Among the many pollutants, the impacts of particulate matter (PM) pollution on human health and ecology are the largest. Airborne fine particles smaller than $2.5\text{ }\mu\text{m}$ (PM_{2.5}) are responsible for haze formation, visibility reduction and also affect global radiation budget. Characterizing the emission sources, concentrations, transport patterns, and impacts is particularly difficult in developing countries, where data are scarce, emissions are high, and health impacts are often severe. The rapid industrialization in the developing countries of South and East Asia has significantly increased the demand on energy, which in turn is based on combustion of coal or gas and this leads to high level of anthropogenic PM emissions. The health impact assessment is done based on the concentrations of inhalable, thoracic and alveoli (**Figure 3**). The highest inhalable percentage of particulate matter value was observed for 'Guwahati (53%)' followed by 'Dimapur (52%)' and 'Dibrugarh (51%)' whereas alveolic percentage ranges from 13-32% which indicates very little possibility as health threat, 33-35% for thoracic. Over the highest vegetation covered location Joypur contribution of alveolar particles (89%) is also significant and minimum is observed in Dimapur, the least vegetated location. It can be inferred that the thoracic and alveolar particles over Joypur, are contributed significantly by bioaerosols as evident from highest CFU values and appreciable fraction of fungal spores, pollens and animal debris. Similar spatial variability during the post monsoon season only was reported earlier by Pathak et al., 2020.

Fluorescence microscopy analyses reveal large spatio-temporal heterogeneity in percentage distribution of both biological and non-biological aerosols in the North-East India (**Figure 4a-d**). The non-biological (non-viable) aerosols are found to vary from 30% (Shillong) - 76% (Jorhat) in monsoon season, 65% (Guwahati) - 89% (Jorhat) in post monsoon, 37% (Shillong) - 73% (Jorhat) in winter and 55% (Dibrugarh) - 89% (Jorhat) in pre-monsoon. Spatially, maximum percentage of non-biological aerosols was found in Jorhat, an urban location very close to the national highway. However, the minimum percentage of NBA was observed over Shillong during monsoon and winter while during post monsoon and pre-monsoon it is over Guwahati and Dibrugarh respectively. The aerosols viz. dust, silica particles etc. are categorized as non-biological aerosols, whose concentration is more in urban locations like Guwahati, Jorhat etc than in vegetated location like Kaziranga, Shillong *etc.* The predominant bioaerosol type found in the bioaerosol samples is pollens followed by fungal spores and animal debris. The existence of aerosols of biological origin in Northeast India is followed respectively by pollens (4-20%), animal debris (1-24%) and fungal spores (1-17%) in different seasons (**Figure 4a-d**). Pollen grain, a vital part of the flowering plant life cycle, is the most important biological component of ambient air. A pollen grain is a specialized structure that encloses the flowering plant male gametes whose biological function is to fertilize the female gametophyte. Pollens ($15\text{-}200\mu\text{m}$) are heavier than most bioaerosols. They settle down after heavy rains and high wind speeds. The average pollen count (in terms of percentage) is found to be higher in monsoon, closely followed by winter. Pollen count was the lowest during post monsoon. Spatially, observed pollen concentration was more in Shillong, a highly vegetated location, reaching upto 20% in

monsoon as well as in winter. This high pollen concentration may be due to high species diversity resulted from physiographical features, altitudinal variation, abundant rainfall, salubrious monsoonal climate and fertile soil of Shillong. Additionally, it may be due to several endemic plants exist in this locality.

Pollen size $\leq 10 \mu\text{m}$ contributed more significantly in monsoon (92.3%) than in winter (85%) in Shillong (Table **S3**). Shillong is followed by 40% contribution of total pollens at other vegetated locations Namsai in monsoon and Dibru-Chaikhuwa in winter. In these locations also pollens of size $\leq 10 \mu\text{m}$ dominate. On the other hand, the semi-urban location Jorhat receives least pollen burden in all seasons except in winter. There, pollens $\leq 10 \mu\text{m}$ varied from 75% in pre-monsoon to 97% in winter. Since, pollen concentration does not exhibit a distinct seasonality and flowering seasons varies from place to place depending upon type of vegetation, pollen identification is essential to explain the spatio-temporal variability elaborately. Animal debris contains insect parts, dead skin debris *etc.*, and possessing strong spatial and temporal variability. It is overdue in Kaziranga, the national park in Assam, followed by Joypur, the rainforest and Dibrugarh university campus. Almost similar spatial variability is observed for fungal spores, with more percentage contribution in monsoon and winter compared to rest of the seasons. Its maximum contribution is observed over Kaziranga in both the seasons followed by Joypur, Dimapur and Dibrugarh in monsoon and Joypur, Dimapur and Dibru-Chaikhuwa in winter.

The post-monsoon bioaerosol samples were subjected to metagenome analysis for the determination of overall bacterial biodiversity in all the sampling sites. The overall bacterial CFU count was found maximum during pre-monsoon season as compared to other seasons. Hence it was expected to observe maximum microbial biodiversity in the samples. Summarily, a total of 184 OTUs (bacterial operational taxonomic units) with 28,028 abundance counts was observed at a 99% nucleotide sequence identity across all 10 samples. Quality Control was performed using FastQC tool v0.11.7. Read quality was good with an average of more than 50,000 reads per sample and read length of 300bp. DADA2 plugin denoising step from QIIME2 converts the overlapping paired-end reads into consensus sequence of V3- V4 region. It finds the overlap for each pair and combines them into a single read. For quality control of the sequences, the DADA2 plugin in Qiime 2 was used to associate erroneous sequence reads with the true biological sequence from which they were derived, thus producing high-quality sequence variant data. Using DADA2, all reads were trimmed to 260 bp, based on the median quality score. In addition, chimeric sequences were detected and excluded from analyses. 16S rRNA Operational Taxonomic Units (OTUs) was picked from the Illumina reads using a closed-reference OTU picking protocol against the Greengenes database (<https://data.qiime2.org/2020.6/common/gg-13-8-99-515-806-nb-classifier.qza>). In the next step, taxonomy assignments were associated with OTUs based on the taxonomy associated with the Greengenes reference sequence defining each OTU.

High quality reads were taken for further analysis. The bar diagram in **Figure 5** shows the taxonomic distribution of the entire bacterial phylum in bioaerosol samples collected from all the designated locations. The maximum number of bacterial genera ($n=10$) was observed in the samples PM3, PM4, PM5, and PM7 of which Proteobacteria was found to be the most abundant phylum in all the samples.

On the other hand, sample PM6 exhibited the minimum bacterial diversity with a total of 7 different phyla (**Figure 5**). These were further extensively analyzed at their phylum, class, order, family, genus and species level. Overall, a total of 7 major phylum, 6 classes, 10 orders, 15 families, 13 genus, and 8 species of prokaryotes were identified in the biosensors samples (**Figure 6a-f**). The pie chart in **Figure 6a** shows the overall taxonomic distribution at phylum level with 71% of proteobacteria followed by firmicutes (21%), and bacteroidetes (8%). The species level distribution clearly shows the presence of Gammaproteobacteria as the most abundant one with a contribution of 52% followed by Bacilli (21%), Alphaproteobacteria (14%), Betaproteobacteria (5%), Flavobacteria (5%), and Sphingobacteria (3%) (**Figure 6f**). The Krona chart (**Figure 7**) summarily depicts the taxonomic profiling OUT of all the microbes analyzed and their abundance in the samples.

Molecular characterization shows presence of the following genus in the bioaerosol samples, viz., *Alkanindiges*, *Bacillus*, *Bordetella*, *Brevibacillus*, *Ignatzschineria*, *Kerstersia*, *Myroides*, *Nitrobacteria*, *Ochrobactrum*, *Paenibacillus*, *Paracaligenes*, *Pseudomonas*, and *Rhodopseudomonas* (**Figure 6e**). Even though, some of the species belongs to the genus *Alkanindiges* such as *Alkanindiges hongkongensis* known to cause breathing related infections such as asthma which pose a potential threat to human health (Fazlollahi et al., 2018; Abdel-Aziz et al., 2019) but, the current study shows the presence of *Alkanindiges illinoisensis* most abundantly in all the aerosol samples and is referred to as a potential biosurfactant producing alkane degrading bacteria with non-pathogenic activities (Ron et al., 2010; Bogan et al., 2003; Xu et al., 2018). *Nitrobacteria* and *Pseudomonads* are found as the 2nd and 3rd most abundant bacterial genus in the bioaerosol samples, out of which most of the species belongs to the genus *Nitrobacteria* are considered as non-pathogenic in nature and are widely used for augmenting the nitrogen level in soil for enhance agricultural yield (Vayenas et al., 2011; EPA, 2002). On the other hand, various species of *Pseudomonads* such as *Pseudomonas aeruginosa*, *P. maltophilia*, *P. fluorescens*, *P. putida*, *P. cepacia*, *P. stutzeri*, *P. maltophilia*, and *P. putrefaciens* etc. are considered as opportunistic pathogens which may cause not only respiratory diseases but also skin diseases, CNS infections, urinary tract infection etc. (Baron S. Galveston, 1996; Moradali et al., 2017). *Bacillus* was found to be the 4th most common bacterial genus in 6 bioaerosol samples. It is worth to mention that a large number of species belongs to the genus *Bacillus* such as *Bacillus subtilis*, *B. thuringiensis*, *B. cereus*, *B. anthracis* are known as spore forming bacteria with a possibility to disperse their spores through aerosols (Alhazmi 2015; Driscoll et al., 2007; Huszczyński et al., 2020). On the other hand, such species of *Bacillus* are well known as potential threat to public health causing respiratory, skin, CNS and gastrointestinal tract associated diseases (Alhazmi 2015; Driscoll et al., 2007; Huszczyński et al., 2020). Additionally, the identity of microbes at different taxon levels along with their abundance in individual bioaerosol samples are shown in **Figures 8a-e**, which gives a comprehensive snapshot of the data generated from individual samples. Even though, the DNA based metagenomic approach have shown the presence of possible pathogenic bacteria genus such as *Bacillus*, *Pseudomonas*, and *Alkanindiges* etc., but in contrast previous culture dependent approach of bioaerosol characterization for the screening of antibiotic resistant microbes showed no prevalence of such microbes (Pathak et al., 2020). It may also be noted that, acquired antibiotic resistance in the microbes present in environmental samples may pose a possible threat to

human health (Morehead and Scarbrough 2018; Naz et al. 2019). Prevalence of antibiotic resistant microbes in aerosol samples collected from various locations with massive anthropogenic activities in India and other countries are widely reported (Brooks et al. 2010; Yadav et al. 2015; Bragoszewska and Biedron 2018; Naz et al. 2019). However, the present work may be considered as the very first hand comprehensive report on the extensive characterization of biological aerosol samples of Northeast India in terms of inhalable, thoracic, and alveolic particles, determination of their constituents (pollens, fungal spores, animal debris, and non-biological components), and the microbial diversity through DNA based metagenomic approach in all the four seasons in the year. Metagenome approach to determine microbial diversity in aerosol samples has now become one of the very recent trends in aerobiology. But scientific findings are still limited and require extensive metagenomic characterization to have a broader view on the microbial biodiversity present in atmospheric aerosols.

Acknowledgements

The funding to carry out the multidisciplinary research work presented in this manuscript was provided by the Department of Science and Technology's Science and Engineering Research Board (DST-SERB), Govt. of India (Grant no. ECR/2016/00132) in Dibrugarh University. Authors gratefully acknowledge DST-SERB. Ankita Khatoniar is thankful to DST-SERB for providing her fellowship during the project tenure. Ankita Medhi and Sristisri Upadhaya are thankful to Miranda House, University of Delhi, Delhi and Dergaon Kamal Dowerah College, Dergaon respectively for providing them the infrastructure for research. Authors greatly acknowledges Mr Biologist LLC, Guwahati for their whole hearted co-operation in performing metagenomic analysis in Convergence Biocrop Pvt. Ltd, Bangalore, India.

Declarations

Authors declare that there is no conflict of interest.

Ethical Approval: Not approval

Consent to Participate: Not approval

Consent to Publish: Not approval

Authors Contributions:

Binita Pathak: Conceptualization, Methodology, Writing- Original draft, Project administration

Ankita Khataniar : Data curation, Investigation, Resources, Software, Formal analysis

Barlin Das: Visualization, Software, Formal analysis

Sristisri Upadhaya: Visualization, Writing - Review & Editing

Ankita Medhi: Software, Resources

Pradip Kumar Bhuyan: Writing - Review & Editing, Supervision

Alak Kumar Buragohain : Writing - Review & Editing, Supervision

Debajit Borah: Conceptualization, Methodology, Formal analysis, Writing- Original draft preparation, Project administration

Funding:

The funding to carry out the multidisciplinary research work presented in this manuscript was provided by the Department of Science and Technology's Science and Engineering Research Board (DST-SERB), Govt. of India (Grant no. ECR/2016/00132) in Dibrugarh University.

Competing Interests:

Authors declare that there is no competing interest.

Availability of data and materials:

All 10 annotated gene sequences in FASTA format were deposited in NCBI GenBank database and the following accession number SAMN21545970 has been successfully received.

References

- Aalismail A N, Ngugi K D, Díaz-Rúa R, Alam I, Cusack M and Duarte M C 2019 Functional metagenomic analysis of dust-associated microbiomes above the Red Sea; *Sci Rep.* 9 13741.
- Aalismail NA, Ngugi DK, Díaz-Rúa R, Alam I, Cusack M, and Duarte CM (2019): Functional metagenomic analysis of dust-associated microbiomes above the Red Sea; *Scientific Reports.* 9: 13741.
- Abdel-Aziz M, Vijverberg J H S, Neerincx H A, Kraneveld D A and Zee H A 2019 The crosstalk between microbiome and asthma; *Exploring associations and challenges Clin Exp Allergy* 49 1067–1086.
- Alhazmi A (2015): *Pseudomonas aeruginosa*- pathogenesis and pathogenic mechanisms; *International Journal of Biology*; 7(2): 44-67.
- Andersson AM, Weiss N, Rainey F, Salkinoja-Salonen MS (1999): Dust-borne bacteria in animal sheds, schools and children's day care centres. *J Appl Microbiol*; 86: 622-634.
- Aneja K R, Sharma C and Joshi R 2010 Fungal infection of the ear: A common problem in the north eastern part of Haryana; *International Journal of Pediatric Otorhinolaryngology* 74 604-607.
- Arancibia F, Bauer T T, Ewig S, Mensa J, Gonzalez J, Niederman M S and Torres A (2002): Community-acquired pneumonia due to gram-negative bacteria and *Pseudomonas aeruginosa*: Incidence, risk and prognosis; *Arch. Intern. Med.* 162(16) 1849-1858.
- Baron S, Fons M and Albrecht T 1996 Viral pathogenesis. *Medical microbiology.* Fourth edition; *University of Texas Medical Branch at Galveston* ch-27.

- Be N A, Thissen J B and Fofanov V Y 2015 Metagenomic Analysis of the Airborne Environment in Urban Spaces; *Microb Ecol.* 69 346–355.
- Bogan W B 2003 *Alkanindiges illinoisensis* gen. nov., sp. nov., an obligately hydrocarbonoclastic, aerobic squalane-degrading bacterium isolated from oilfield soils; *International Journal of Systematic and Evolutionary Microbiology* 53 1389–1395.
- Brągoszewska E and Biedroń I 2018 Indoor Air Quality and Potential Health Risk Impacts of Exposure to Antibiotic Resistant Bacteria in an Office Rooms in Southern Poland; *Int J Environ Res Public Health* 15(11) 2604. [https://doi: 10.3390/ijerph15112604](https://doi.org/10.3390/ijerph15112604).
- Brooks J P, McLaughlin M R, Scheffler B and Miles D M 2010 Microbial and antibiotic resistant constituents associated with biological aerosols and poultry litter within a commercial poultry house; *Sci. Total Environ.* 408 4770–4777.
- Bundy K W, Gent J F, Beckett W, Bracken M B, Belanger K, Triche E and Leaderer B P 2009 Household airborne *Penicillium* associated with peak expiratory flow variability in asthmatic children; *Ann Allergy Asthma Immunol.* 103 26-30.
- Burrows S M, Elbert W, Lawrence M G and Poschl U 2009a Bacteria in the global atmosphere Part-1: Review and synthesis of literature data for different ecosystems; *Atmos. Chem. Phys.* 9 9263-9280.
- Chakrawarti K M, Singh M, Yadav P V, Mukhopadhyay K 2019 Temporal Dynamics of Air Bacterial Communities in a University Health Centre Using Illumina MiSeq Sequencing; *Aerosol Air Qual. Res.* 20 966-980.
- Cytotoxicity in Human Bladder Cancer; *Cell* 181 1612–1625.
- Degois J, Clerc F, Simon X, Bontemps C, Leblond P and Duquenne P 2017 First Metagenomic Survey of the Microbial Diversity in Bioaerosols Emitted in Waste Sorting Plants; *Annals of Work Exposures and Health* 61 (9)1076-1086.
- Driscoll A J, Brody L S and Kollef H M 2007 The Epidemiology, Pathogenesis and Treatment of *Pseudomonas aeruginosa* Infections; *Drugs* 67 (3) 351-368.
- EPA 2002 nitrification report; *U.S. Environmental Protection Agency Office of Ground Water and Drinking Water Standards and Risk Management Division 1200 Pennsylvania Ave., NW Washington DC.*
- Fazlollahi M, Lee D T, Jade Andrade J, Oguntuyo K, Chun Y, Grishina G, Grishin A and Supinda Bunyavanich S 2018 The Nasal Microbiome in Asthma; *J Allergy Clin Immunol.* 142(3) 834–843.
- Gangamma S 2014 Characteristics of airborne bacteria in Mumbai urban environment; *Sci. Total Environ.* 488–489 70–74.
- Geogakopolus D G, Després V, Nowoisky J, psenner R, Ariya P A, Pósfali M, Ahern H E, Moffet B F and Hill T C J 2008 Microbiology and atmospheric processes: biological, physical and chemical characterization of aerosol particles; *Biogeosciences Discuss.* 5 1469–1510.
- <https://data.qiime2.org/2020.6/common/gg-13-8-99-515-806-nb-classifier.qza>

- Hugenholtz P, Goebel BM, Pace NR (1998): Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol*; 180: 4765-4774.
- Huszczyński, S M, Lam J S and Khursigara C M 2020 The Role of *Pseudomonas aeruginosa* Lipopolysaccharide in Bacterial Pathogenesis and Physiology; *Pathogens* 9 1.
- Jaenicke R 2005 Abundance of cellular material and proteins in the atmosphere; *Science* 308(5718) 73.
- Keegan K P, Glass E M and Meyer F 2016 MG-RAST, a metagenomics service for analysis of microbial community structure and function; *Methods Mol. Biol.* 1399 207–233.
- Kim K, Kabir E and Jahan S A 2017 Airborne bioaerosols and their impact on human health; *J. Environ. Sci.* 1001-0742.
- Kumar B, Gupta GP, Singh S and Kulshrestha, U C 2013. Study of abundance and characterization of culturable bioaerosol at Delhi, India; *Int. J. Environ. Eng. Manage.* 4 219–226.
- Lal H, Punia T, Ghosh B, Srivastava A and Jain V K 2013 Comparative study of bioaerosol during monsoon and postmonsoon seasons at four sensitive sites in Delhi region; *Int. J. Adv. Earth Environ. Sci.* 1(2) 1–7.
- Moradali M F Ghods S and Rehm B H A 2017 *Pseudomonas aeruginosa* Lifestyle: A Paradigm for Adaptation, Survival, and Persistence; *Frontiers in Cellular and Infection Microbiology* 7 2235-2988.
- Morehead S M and Scarbrough C 2018 Emergence of Global Antibiotic Resistance; *Prim Care Clin Office Pract.* 45 467–484.
- Nandini N and sivaskthiya S 2014 Microbiological Pollution of Air in Lal Bagh Botanical Gardens, Bangalore, Karnataka, India; *International Journal of Science and Research.* 3 648-652.
- Naz N, Nasim F H and Pasha T S 2019 Prevalence of antibiotic-resistant airborne bacteria along roadsides in Rahim Yar Khan, Pakistan; *Pol. J. Environ. Stud.* 28 1295–1303.
- Núñez A, de Paz A G, Alberto R, Ana G M, Antonio A, Gutiérrez Bustillo A, Montserrat M and Diego A 2016 Monitoring of airborne biological particles in outdoor atmosphere. Part 1: Importance, variability and ratios; *Int. Microbiol.* 19(1) 1–13.
- Oh Y D, Kwek S S and Raju S S 2020 Intratumoral CD4⁺ T Cells Mediate Anti-tumor
- Pathak B, Borah D, Katanier A, Bhyan P K, and Buragohain A K 2020 Characterization of bioaerosols in Northeast India in terms of culturable biological entities alongwith inhalable, thoracic and alveolar particles; *J. Earth Syst. Sci.* 129(1).
- Perez-Padilla et al 2009 Pneumonia and Respiratory Failure from Swine-Origin Influenza A (H1N1) in Mexico; *N Engl. J Med.* 361 680-9.
- Petavratzi E, Kingman S and Lowndes I 2005 Particulates from mining operations: A review of sources, effects and regulations; *Miner. Eng.* 18(12) 1183–1199.
- Peters, Thomas & Ott, Darrin & O'Shaughnessy, Patrick. (2006). Comparison of the GRIMM 1.108 and 1.109 Portable Aerosol Spectrometer to the TSI 3321 Aerodynamic Particle Sizer for Dry Particles.

The Annals of occupational hygiene. 50. 843-50. 10.1093/annhyg/mel067.

- Ron E and Rosenberg E 2010 Acinetobacter and Alkanindiges. In: Timmis K.N. (eds) Handbook of Hydrocarbon and Lipid Microbiology; *Springer, Berlin, Heidelberg*.
- Sarma P, Antunes A, Simões AF, and Sarma SD (2020): Earth's stratosphere and microbial life; *Curr Issues Mol Biol*. 38: 197-244.
- Scheel J F, Kunert A T, Kampf C J, Mauri S, Weidner T, Pöschl U and Nowoisky J F 2016 Identification & Characterization of Fungal Ice Nucleation Proteins; *Geophysical Research Abstracts* 18 2016-6865.
- Serrano-Silva N, and Calderon-Ezquerro MC (2018): Metagenomic survey of bacterial diversity in the atmosphere of Mexico City using different sampling methods; *Environmental Pollution* 235: 20-29.
- Shin S K, Kim J, Ha Sm, Oh H S, Chun J 2015 Metagenomic Insights into the Bioaerosols in the Indoor and Outdoor Environments of Childcare Facilities; *PLOS ONE* 10(5).
- Shin S-K, Kim J, Ha S-m, Oh H-S, Chun J, Sohn J, and Yi H (2015): Metagenomic insights into the bioaerosols in the indoor and outdoor environments of childcare facilities. *PLoS ONE* 10(5): e0126960.
- Singh AB and Kumar P (2004): Aerial Pollen Diversity in India and Their Clinical Significance in Allergic Diseases; *Indian Journal of Clinical Biochemistry* 19 (2) 190-20.
- Smith J D 2018 Airborne Bacteria in Earth's Lower Stratosphere Resemble Taxa Detected in the Troposphere: Results From a New NASA Aircraft Bioaerosol Collector (ABC); *Frontiers in Microbiology* 9 1752. <https://doi.org/10.3389/fmicb.2018.01752>.
- Srivastava A, Singh M and Jain V K 2012 Identification and characterization of size-segregated bioaerosols at Jawaharlal Nehru University, New Delhi; *Nat. Hazards* 60(2) 485–499.
- Sweeny M T, Lubbers B V, Schwarz S and Watts J L 2018 Applying definitions for multidrug resistance, extensive drug resistance and pandrug resistance to clinically significant livestock and companion animal bacterial pathogens; *J. Antimicrob Chemother* 73 1460–1463.
- Vayenas V D 2011 6.30 - Attached Growth Biological Systems in the Treatment of Potable Water and Wastewater; *Sciencedirect* 6 371-383.
- Wang Z, Li J, Qian L, Liu L, Qian J, Lu B and Guo Z 2019 Composition and Distribution Analysis of Bioaerosols Under Different Environmental Conditions; *J. Vis. Exp.* (143) e58795. doi:10.3791/58795.
- Womack A M, Bohannon B J M and Green J L 2010 Biodiversity and biogeography of the atmosphere; *Philos. Trans. R. Soc.* B365 3645-3653.
- Xu X, Liu W, Tian S, Wang W, Qi Q, Jiang P, Gao X, Li F, Li H and Yu H 2018 Petroleum Hydrocarbon-Degrading Bacteria for the Remediation of Oil Pollution Under Aerobic Conditions: A Perspective Analysis; *Front Microbiol.* 92885. doi: 10.3389/fmicb.2018.02885.
- Yadav J, Kumar A, Mahor P, Goel A K, Chaudhary H S, Yadava P K, Yadav H and Kumar P 2015 Distribution of airborne microbes and antibiotic susceptibility pattern of bacteria during Gwalior trade fair, central India; *J. Formos. Med. Assoc.* 114 639–646.

- Yoo K, Lee TK, Choi EJ, Yang J, Shukla SK, Hwang S, and Park J (2017): Molecular approaches for the detection and monitoring of microbial communities in bioaerosols: a review. J Environ Sci (China); 51: 234-47.

Figures

Figure 1

The NDVI map of Northeast India in showing nine discrete campaign locations.

Figure 2

Particulate (PM10) Counts (in Millions) versus colony formation unit (CFU/m3) at different study locations during four seasons.

Figure 3

Percentage share of inhalable, thoracic (PM10), alveolic particulate concentration.

Figure 4

Percentage Distribution pollens, fungal spores, animal debris and non-biological components over different sampling sites during the four seasons.

Figure 5

Taxonomic distribution of different genera in individual bioaerosol samples collected from various locations of Northeast India.

Figure 6

Overall distribution of microbes at (a) species, (b) genus, (c) family, (d) order, (e) class and (f) phylum level in the bioaerosol samples collected from various locations.

Figure 7

Krona chart depicting the taxonomic profiling.

Figure 8

Taxonomy profiling graph of all the bioaerosol samples at different taxon level.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [graphicsabstract.jpg](#)
- [SupplementaryTables.docx](#)