Challenges and Opportunities of Airborne Metagenomics

Hayedeh Behzad[1], Takashi Gojobori[1,2,*], and Katsuhiko Mineta[1,3]

[1]Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

[2]Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

[3]Computer, Electrical, and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

*Corresponding author: E-mail: takashi.gojobori@kaust.edu.sa.

Abstract

Recent metagenomic studies of environments, such as marine and soil, have significantly enhanced our understanding of the diverse microbial communities living in these habitats and their essential roles in sustaining vast ecosystems. The increase in the number of publications related to soil and marine metagenomics is in sharp contrast to those of air, yet airborne microbes are thought to have significant impacts on many aspects of our lives from their potential roles in atmospheric events such as cloud formation, precipitation, and atmospheric chemistry to their major impact on human health. In this review, we will discuss the current progress in airborne metagenomics, with a special focus on exploring the challenges and opportunities of undertaking such studies. The main challenges of conducting metagenomic studies of airborne microbes are as follows: 1) Low density of microorganisms in the air, 2) efficient retrieval of microorganisms from the air, 3) variability in airborne microbial community composition, 4) the lack of standardized protocols and methodologies, and 5) DNA sequencing and bioinformatics-related challenges. Overcoming these challenges could provide the groundwork for comprehensive analysis of airborne microbes and their potential impact on the atmosphere, global climate, and our health. Metagenomic studies offer a unique opportunity to examine viral and bacterial diversity in the air and monitor their spread locally or across the globe, including threats from pathogenic microorganisms. Airborne metagenomic studies could also lead to discoveries of novel genes and metabolic pathways relevant to meteorological and industrial applications, environmental bioremediation, and biogeochemical cycles.

Key words: airborne microorganisms, culture-independent studies, microbial diversity, metagenomics, 16S rRNA sequencing, metabolic potential.

Introduction

The impact of airborne microbes and their overall contributions to global ecosystem is disproportionally understudied, even though microorganisms are ubiquitously present in the air. Airborne microbes were considered passive dwellers moving with the wind; however, several studies strongly suggest that atmospheric microbes are metabolically active. Microorganisms metabolize the organic matter in cloud water and potentially contribute to the biogeochemical cycles of earth (Amato et al. 2005, 2007; Hill et al. 2007). Some authors suggest that microbes affect atmospheric chemistry and the precipitation cycles of earth (Delort et al. 2010; Morris et al. 2011) implicating airborne microbes in atmospheric processes, such as ice nucleation and cloud formation. The increase in cloud formation may potentially be responsible for earth's climate changes (Rosenfeld et al. 2014). Beyond these studies, relatively little data exist on the potential impact of airborne microbes upon our planet.

The current understanding of airborne microorganisms comes from culture-based studies; however, the majority of environmental microbes cannot be cultured in this way. Moreover, as pure cultures of microorganisms contain a single type of microbes, culture-based approaches miss the opportunity to study the interactions among different microbes and their surrounding environment; therefore, they cannot uncover the genomic variances and biological functions associated with such interactions. The main advantage of culture-dependent studies is the ability to grow microorganisms in bulk and study them at the both molecular and

cellular levels. Their main disadvantage is that they disproportionally favor some microbes over others, providing an inaccurate representation of the microbial community as a whole.

In recent years, our knowledge of airborne microbial diversity and their potential metabolic impact has increased significantly through molecular analysis of microbial genetic material through culture-independent studies of environmental samples. Two common approaches to investigate microbial diversity and/or their metabolic potential in the environment are 1) the polymerase chain reaction (PCR)-based rRNA (16S and 18S) gene sequencing approach for assessment of microbial diversity in which a single rRNA gene is used as a phylogenetic marker to compare relatedness between microorganisms; and 2) the whole-genome shotgun metagenomics approach for assessment of microbial diversity and function, in which the entire microbial genome is sheared into smaller fragments, sequenced, and reconstructed to assess their diversity and metabolic potential (fig. 1).

In this article, we use the term metagenomics in its original sense to mean the application of whole-genome shotgun sequencing to environmental samples. With the advent of next-generation sequencing (NGS), shotgun metagenomic studies of various environments, including soil (Fierer et al. 2012; Sangwan et al. 2012), marine (La Cono et al. 2011), and human biome (Methe et al. 2012), have expanded significantly as evidenced by the large data sets available for these ecosystems. In contrast, progress in air metagenomic research has been relatively slow (fig. 2). Here, we will explore current

Fig. 1.—Flow chart of the process involved in shotgun metagenomic sequencing of airborne microbes from collection of air samples to analysis of its microbial content.

progress in airborne metagenomics with a particular emphasis on the challenges associated with these studies and their potential opportunities.

Airborne Microorganisms Play Active Roles in the Atmosphere; Exploration of This Seemingly Untouched Frontier Has Great Potential

A number of studies conducted in culture suggest that airborne microorganisms play active roles in atmospheric events and/or are metabolically active. For example, airborne bacteria were shown to metabolize glucose (Dimmick et al. 1975). More recently, Hill et al. (2007) used the uptake of tetrazolium dye by bacteria as an indicator of metabolic activity in cells and showed that 76% of the bacteria in cloud water were active. High concentrations of inorganic and dissolved organic nitrogen were found in cloud waters that contained nitrifying bacteria, suggesting that the bacteria in cloud water play a role in the cycling of organic nitrogen in the atmosphere (Hill et al. 2007). Amato et al. (2005) used the increase in ATP concentration in cloud water as an indicator of cell metabolic activity and showed that the bacteria in cloud water were growing exponentially and metabolically active. The bacteria isolated from cloud water metabolized the main carboxylic compounds (formate, acetate, lactate, and succinate) in cloud water; the end products of these metabolic activities were commonly found in cloud water, suggesting that these bacteria were actively involved in transformation of organic compounds in the atmosphere (Amato et al. 2007; Vaitilingom et al. 2010). Durand et al. (2006) identified a bacterial strain capable of biodegrading the herbicide mesotrione. Through biodegradation of organic matters in the atmosphere, airborne microorganisms could play a role in the bioremediation of environmental pollutions.

Airborne bacteria could also affect atmospheric conditions by engaging in ice-nucleating activities, cloud formation, or precipitation (Bauer et al. 2003; Mohler et al. 2008; Morris et al. 2011). Ice-nucleating proteins were identified on the outer membrane of plant associated Gram-negative bacteria such as *Pseudomonas syringae*, *Pseudomonas fluorescens*, and *Erwinia herbicola* (Deininger et al. 1988; Kozloff et al. 1991; Turner et al. 1991). Ice-nucleation active bacteria are present in the atmosphere (Constantinidou et al. 1990; Bowers et al. 2009). Some strains of fungi including *Fusarium avenaceum* and related genera also possess ice-nucleating abilities (Pouleur et al. 1992; Hasegawa et al. 1994). Biological ice-nuclei can catalyze ice formation at temperatures as high as $-2\,^\circ$C, higher than the temperature required for ice-nucleation by nonbiological particles (Mohler et al. 2007). Knowledge of the components involved in ice-nucleating activities of airborne microbes is paramount to understanding their roles in atmospheric events.

Most assays examining the metabolic activities of airborne microbes were conducted under laboratory conditions, prone
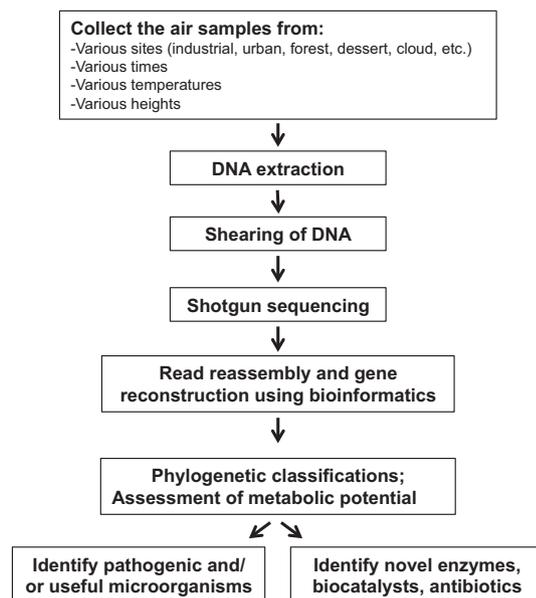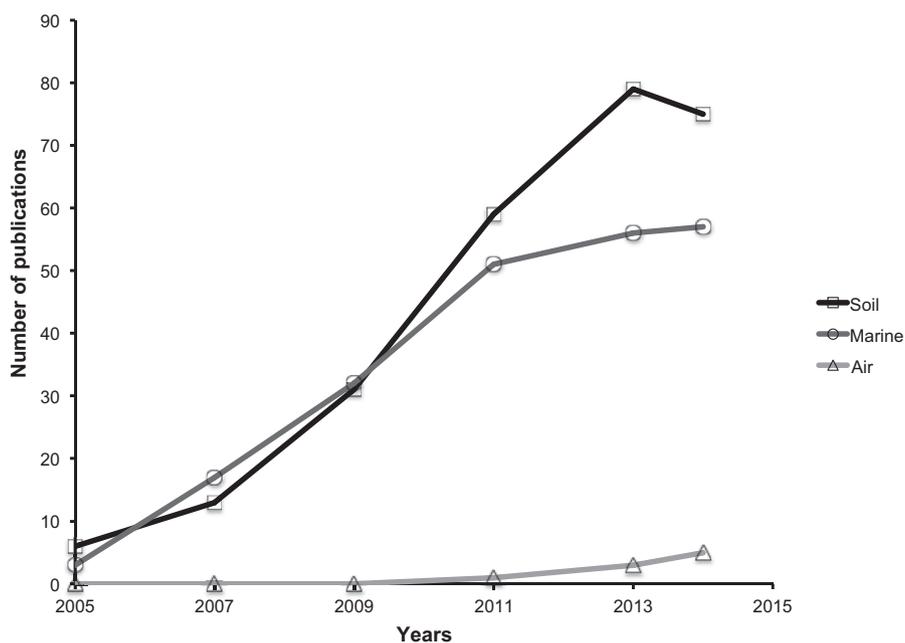
Fig. 2.—Relative number of publications related to metagenomic studies of soil, marine, and air. The graph was generated with the data obtained from Scopus using the search words: "soil metagenomic," "marine metagenomic," and "air/airborne metagenomic," respectively.

to bias as the exact atmosphere environment is not replicated. These studies were also mainly conducted on samples taken from cloud water. Cloud-borne microbes are not necessarily the same microbes that thrive in dry or humid air; therefore, they do not paint a full picture of airborne microbial ecosystems. However, these discoveries underscore the potential importance of airborne microbes and the need to characterize them in their entirety. As culture-based studies represent only a minority of the microbial population, they are not up to this task. Culture-independent studies are required.

Exploring Diversity of Microorganisms in the Atmosphere through Culture-Independent Analysis of Their rRNA Gene

It is estimated that in a cubic meter of air there can be hundreds of thousands of individual microbes (Burrows et al. 2009), with a diversity of taxa comparable to what has been found in soil (Franzetti et al. 2011). "Who" are these microbes and "what" do they do? In recent years, through use of culture-independent approaches and in particular with PCR-based applications such as rRNA gene sequencing, we have managed to answer the first question about who these microbes are through their phylogenetic diversity (table 1). However, the answer to the second question regarding the function of airborne microbes remains largely incomplete.

The PCR-based rRNA gene sequencing studies successfully demonstrated that air can harbor vastly diverse microbes. The composition of these microbes continuously changes in response to meteorological, regional (coastal, urban, and

forest), seasonal, and diurnal patterns (Gandolfi et al. 2013). Gram-positive as well as Gram-negative spore and nonspore forming bacteria are both present in the air. The most common bacterial groups identified were Protobacteria, Firmicutes, Bacteriodetes, Actinobacteria, Acidobacteria, and Cyanobacteria (table 1), some of which contain gene sequences classically found in bacteria dwelling in soil, water, or plants (Maron et al. 2005; Brodie et al. 2007). Known ice nucleating plant-associated bacteria including *P. syringae, P. fluorescens, E. herbicolas*, and *Sphingomonas* spp were also present in the atmosphere (Bowers et al. 2009). A portion of microorganisms in the air was found to be potentially pathogenic to plants, animals, and humans (Brodie et al. 2007). Various fungal genera including those of *Cladosporium, Aspergillus*, and *Penicillium* also dwell in the air, many of which are potential allergens (Oh et al. 2014). Although various studies identified plant, animal, and human pathogens in the air, a systematic approach to monitoring airborne pathogens and examining their impact on global ecosystems is lacking.

Research on the source of airborne microorganisms reveals that although the majority of these microbes originate from local sources (soil, plants, and marine), some are unique to the local atmosphere. These microbes could originate from distant sources transported by strong winds, sandstorms, or hurricane. A recent study by An et al. (2014) using amplicon-based 16S rRNA gene sequencing demonstrated that the atmospheric sand-associated bacterial composition was modified during sandstorm events in China and South Korea. In particular, they found an increase in the level of putative

GBE

Table 1

Microorganisms (bacteria, fungi, and viruses) Detected in the Atmosphere Using Culture-Independent Methodologies

Bacteria	Fungi/Plant	Viruses	Method of Identification	Location	Reference
Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria	NA	NA	16S rRNA microarray	Mountain (USA)	Smith et al. (2013)
Firmicutes, Proteobacteria	Ascomycota, Streptophyta	Pseudomonas phage F10	Metagenomic sequencing (NGS)	Urban area (USA)	Be et al. (2014)
Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria	NA	NA	16S rRNA sequencing (NGS)	Mushroom cultivation facility (South Korea)	Chun et al. (2012)
Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria	NA	NA	16S rRNA sequencing (Sanger)/metagenomic sequencing	shopping center (Singapore)	Tringe et al. (2008)
Acidobacteria, Actinobacteria, Bacteroidetes, Chloroflexi, Cyanobacteria, Firmicutes, Nitrospira, Protobacteria, Verrucomicrobiae	NA	NA	16S rRNA microarray	Urban area (USA)	Brodie et al. (2007)
Acidobacteria, Actinobacteria, Bacteroidetes, Firmicutes, Protobacteria	NA	NA	16S rRNA sequencing (NGS)	Upper Troposphere (international)	Deleon-Rodriguez et al. (2013)
Acidobacteria, Actinobacteria, Cyanobacteria, Fermicutes, Flavobacteria, Gemmatimonadetes, Planctomycetacia, Proteobacteria, Verrucomicrobia, Sphingobacteria	NA	NA	16S rRNA sequencing (Sanger)	Rural area (France)	Maron et al. (2005)
NA	Ascomycota, Basidiomycota, Chytridiomycota	NA	LSU rRNA sequencing (NGS)	Urban area (South Korea)	Oh et al. (2014)
Acidobacteria, Actinobacteria, Bacteroidetes, Cyanobacteria, Firmicutes, Protobacteria	Ascomycota, Basidiomycota, Zygomycota, Streptophyta	NA	16S rRNA sequencing (Sanger)/16S rRNA bar-coded pyrosequencing	Mountain (USA)	Bowers et al. (2009)
Actinobacteria, Bacteriodetes, Chloroflexi, Firmicutes, Proteobacteria	Ascomycota	Pseudomonas Phage F116, Human adenovirus C, Enterobacteria Phage P1	Metagenomic sequencing (NGS)	Urban area (China)	Cao et al. (2014)
NA	NA	Geminivirues-related viruses, Circoviruses, Nanoviruses, Microphage-related viruses	Metagenomic sequencing (NGS)	Forest, urban, coastal area (South Korea)	Whon et al. (2012)
Actinobacteria, Bacteriodetes, Firmicutes, Proteobacteria	NA	NA	16S rRNA sequencing (NGS)	Sand storm (China and South Korea)	An et al. (2014)

Note.—We used phylum name to represent the organisms except for viruses. NA, not applicable.

human bacterial pathogens in samples obtained during sandstorms (An et al. 2014). Microbes from sandstorm events, such as the ones originating from Gobi desert, can be transported very large distances, even reaching across the Pacific Ocean to North America (Bishop et al. 2002). This has significant implications for the dispersal of pathogenic microbes around the globe. Deleon-Rodrigues et al. (2013) show that hurricanes, similar to sandstorms, dramatically change atmospheric bacterial composition, even at high altitudes (10 km above the sea level), suggesting that hurricanes aerosolize and transport bacterial cells over long distances (DeLeon-Rodriguez et al. 2013).

Diverse and viable communities of bacteria reside high in the troposphere despite the extreme cold, high UV irradiation, and thin air, suggesting adaptive mechanisms. Surprisingly, viable bacteria form up to 20% of the total number of particles found in the troposphere (DeLeon-Rodriguez et al. 2013). Some atmospheric regions have extreme conditions, but microorganisms already live under even harsher environments with extremes of pH, temperatures, and radiation (Pikuta et al. 2007). The roles of high-altitude dwelling bacteria are not well understood. It is likely that these microbes affect meteorological events, such as cloud formation, precipitation, or atmospheric chemistry.

The PCR-based rRNA gene sequencing approach has provided a glimpse into the phylogenetic diversity of microorganisms in the atmosphere. However, this approach has many limitations, as follows: 1) It is challenging to identify novel microorganisms with this approach; the universal rRNA primers used to amplify the rRNA genes are designed based on preexisting rRNA sequences in the gene database, and these primers might not amplify new genes (Rajendhran and Gunasekaran 2011); 2) characterization of microorganisms based on a single rRNA gene is difficult due to potential horizontal gene transfer of the rRNA gene between distally related microorganisms (Acinas et al. 2004); 3) due to a high degree of similarity in the rRNA gene among related species and strains, this approach is limited to characterizing microbes mostly at higher taxonomic levels, such as family or genus where the difference in the rRNA gene is more significant, but not so at lower taxonomic levels, such as species or strains where more similarities in the gene are observed (Bowers et al. 2009; DeLeon-Rodriguez et al. 2013); 4) amplicon-based methods in general result in bias due to preferential amplification and generation of artifacts and erroneous products introduced during PCR steps (Hong et al. 2009; Wylie et al. 2012; Logares et al. 2014; Sharpton 2014); 5) As this approach is based on the analysis of a single rRNA gene, it cannot provide information on the function of other microbial genes and their metabolic potential; and 6) it cannot be used to study viruses because they lack common phylogenetic markers such as the rRNA gene.

Whole-genome shotgun metagenomics, on the other hand, provides a more powerful approach to direct sequencing of microbial genomes, and to the assessment of their diversity, and metabolic potential (Sharpton 2014). This approach is relatively new but has revolutionized environmental microbiology by providing new insights into the diversity and metabolic potential of microbes in their natural habitat, such as marine, soil, and human biome. Due to its unique properties, air could be another promising frontier for discovery of unexplored microbial ecosystems with unique genes, biomolecules, and novel enzymes waiting to be discovered.

Current Progress in Airborne Metagenomics and Its Potential Opportunities

Although progress in air metagenomics has been slow relative to other environments, the little progress that has been made is promising. In a recent study, Cao et al. (2014) used metagenomics to analyze the microbial composition of air samples during a severe smog event in China and successfully identified various airborne microorganisms including double-stranded DNA viruses. Their work identified the sequences of several respiratory pathogens and allergens in the air, and showed that their relative abundance increased with increased air pollution. Yooseph et al. (2013) established a metagenomic framework for studying the microbial composition of air samples from several indoor and outdoor environments, and successfully identified highly diverse microbial communities with genes involved in metabolism, transport, translation, and signal transduction. Tringe et al. (2008) studied the airborne metagenome in an indoor environment and identified genes for potentially adaptive mechanisms involved in resistance to desiccation and oxidative damage. Whon et al. (2012) investigated viral diversity in the near-surface atmosphere over three distinct land locations (residential, forest, and industrial), and showed the presence of diverse airborne viruses including those that infect plants and animals. They also revealed seasonal changes in the abundance of airborne viral communities with higher viral abundance toward winter and lower abundance toward spring. The increase in viral abundance during winter could explain higher rates of influenza viral transmission during cold and dry season (Lowen et al. 2007; Shaman and Kohn 2009). With continuous progress and advances in the field, metagenomic approaches could become convenient tools for monitoring airborne viral, bacterial, and fungal pathogens and studying their global distribution patterns.

There are also strong indications, based on culture-dependent studies, that airborne microorganisms can transform organic matters in clouds (Durand et al. 2006; Amato et al. 2007), or contribute to atmospheric events such as ice nucleation and cloud formation (Gurian-Sherman and Lindow 1993; Mohler et al. 2007, 2008; DeLeon-Rodriguez et al. 2013), but the extent of microbial involvement and the mechanisms involved are unclear. Metagenomic studies can help identify the airborne microbes, and their genes and metabolic pathways that potentially contribute to these events. These

studies have great potential for novel and advanced discoveries with applications in atmospheric transformation, meteorological applications, environmental bioremediation, and health. Among the challenges of metagenomic studies are unraveling this vast potential and finding ways to incorporate these finding into novel applications.

Current airborne metagenomic research demonstrates, in principle, that it can identify and characterize microorganisms in the air. However, considerable advances in the field are required to enable sufficient characterization of the entire airborne microbiome.

Challenges in Conducting Airborne Metagenomics

The Low Density of Microorganisms in the Air

Microorganisms including bacteria, viruses, and fungi are scattered throughout the atmosphere. The bacterial concentration in air is estimated to range between 10^4 and 10^6 microbes/m^3 (Lighthart et al. 2000) much lower than those of marine and soil environments, where 1 g of topsoil is believed to contain approximately 10^8–10^9 prokaryotes, and 1 ml of marine water is believed to contain approximately 10^4–10^7 microorganisms (Whitman et al. 1998).

A number of studies investigated the concentration of airborne microorganisms, but it is difficult to relate their data because they used different methodologies. Traditional methods use culture-based techniques, which measure the concentration of microorganisms as colony-forming units per cubic meter of air (CFU/m^3). As the vast majority of environmental microbes cannot be cultured, culture-based methods likely underestimate microbial concentrations (Li and Huang 2006). In recent years, microorganisms have been primarily quantified directly from the environment using techniques such as quantitative PCR (qPCR), epifluorescence microscopy, and flow cytometry. These techniques provide different estimates of microbial density in air, sometimes by a few orders of magnitude (Gandolfi et al. 2013). Airborne bacterial concentrations, for example, were reported over the following ranges: Approximately 0–10^4 CFU/m^3 using culture-based methods (Fahlgren et al. 2010; Ravva et al. 2012; Haas et al. 2013), approximately 10^3–10^6 cells/m^3 by epifluorescence microscopy (Bowers et al. 2009, 2011; Vaitilingom et al. 2012), and approximately 10^1–10^6 cells/m^3 by qPCR (Lee et al. 2010; Cho and Hwang 2011; Bertolini et al. 2013). Unfortunately, without base-line comparative studies, these numbers can only be understood in isolation. A precise comparison between various quantification techniques is only possible when samples are collected from the same location using the same collection device. This is important as microbial concentration varies significantly depending on the location where the air samples were taken and the type of air samplers used. Current air sampling devices often do not capture microorganisms at maximum efficiency. It is likely that the abundance of microorganisms in the air is greater than what has

been reported. If that is the case, the bottleneck lies with current technology.

Whole-genome shotgun sequencing requires large quantities of the starting genetic material, which entails collecting thousands of liters of air. In practice, capturing large quantities of microorganisms from air is challenging. Improving the collection efficiency of air sampling devices would help mitigate this. To circumvent the problem of low DNA yield from air, some metagenomic protocols incorporate an additional step of amplifying the environmental DNA prior to sequencing using methods such as Multiple Displacement Amplification (Lasken 2009; Yooseph et al. 2013). These amplification steps can produce micrograms of DNA from femtograms of the starting genetic material (Lasken 2009), ensuring that sufficient amounts of DNA are available for sequencing. However, amplification can generate primer mismatches, chimeras, and sequence biases—the degree of which depends upon the quantity and quality of the starting genetic material, and the number of required amplification cycles.

Metagenomic studies require high quantities of DNA, which could potentially be obtained through efficient collection of microorganisms from the air, and the development of unbiased methodologies to enhance DNA yield following extraction.

Efficient Collection of Microorganisms from the Air

Due to the lower density of microorganisms in air, it is paramount that air-sampling devices collect microorganisms from the air with high efficiency. Typical air samplers draw air containing various particles into an airflow nozzle, which then directs the air toward a collection surface. The sampling efficiency of such devices is affected by a combination of factors including intake air velocity, the shape and diameter of the airflow nozzles, the distance between the nozzle and the collection surface, the properties of the collection surface, the particle cut-off diameter, and collection times (Nevalainen et al. 1992; Whyte et al. 2007). The sampling efficiency is also influenced by the inertial properties of microorganisms, related to their size and density (Nevalainen et al. 1992).

Two air sampling strategies are commonly used to collect microorganisms from air for molecular analysis: Filtration-based air samplers work by drawing air, either through pumps or vacuum lines, through filters of varying pore sizes, shapes, and composition (Lundholm 1982; Brodie et al. 2007; Bowers et al. 2009; Fahlgren et al. 2010, 2011). These samplers are the most commonly used devices for aerosol collection. They are relatively easy to use and less expensive, but have shortcomings. The low porosity of the filters can dramatically reduce the airflow rates, increasing sample collection time. Long collection times on dried filters lead to desiccation of the collected microbes. Furthermore, these devices are not suitable for quantification-type studies due to difficulty in

retrieving intact microorganisms from filters (Fahlgren et al. 2011).

Liquid impinger-based samplers work by drawing high-flow-rate, high-velocity air through a nozzle into a liquid collection medium. The particles in the air stream are impacted on the collection plates, wetted, and retained in the collection medium (Burge and Solomon 1987; Fierer et al. 2008; Fahlgren et al. 2011). Due to their high air-flow rate, they can collect larger biomass in shorter time, relative to filter-based devices. As the microbes are collected in liquid medium, their viability is retained. These devices are typically designed for relatively short sample collection times, which is required to prevent evaporation of the liquid collection medium and loss of cell viability (Fahlgren et al. 2011). This could prevent collection of sufficient biomass for studies requiring large concentrations of nucleic acids. Modifications, such as replenishing liquids, could lengthen sample collection times and overcome this problem.

Specific types of studies necessitate the use of particular types of air samplers. For example, culture-based studies require air samplers that can maintain cell viability during the collection period. Metagenomic studies of the entire microbial communities require devices that can provide sufficient microbial biomass representing the sampled community. Studies examining specific types of microorganisms (viruses vs. bacteria vs. fungi) need devices that preferentially and efficiently collect microbes based on size. Once the capturing strategy is decided, that is, using filtration versus impingement, then the efficiency of the various devices using the same strategy must be thoroughly investigated before embarking on a large-scale study.

Variability in Airborne Microbial Community Composition

The composition of airborne microbes continuously changes due to meteorological, spatial, and temporal patterns (Brodie et al. 2007; Frohlich-Nowoisky et al. 2009; Bowers et al. 2011; Franzetti et al. 2011; Bertolini et al. 2013; Gandolfi et al. 2013). Bowers et al. (2011) investigated the spatial variability in airborne microbial composition in the near surface atmosphere, and showed that bacterial composition significantly varied depending on the land-use type, suggesting that bacteria from the local land sources contributed to this variability. Bertolini et al. (2013) demonstrated temporal variability in airborne bacterial community composition in an urban atmosphere, showing that bacterial composition varied significantly depending on the season. Even within the same season, bacterial composition was significantly different between consecutive days.

This inherent variability makes studies of airborne microbial composition challenging. It is difficult, for example, to study the possible contribution of environmental pollution or anthropogenic factors to microbial dynamics, if a background level cannot be established, due to variability (Brodie et al.

2007). Without the reproducible background level, monitoring of environmental pathogens becomes complicated. The continuous change in composition of microbes in the atmosphere also makes it difficult to compare across different studies as the outcome of these studies could vary significantly depending on location, time of day, season, altitude, and other environmental factors.

Lack of Standardized Methodologies in Sample Collection and Processing

The procedures and protocols used to obtain nucleic acids from the air play a significant role on the outcome of metagenomic studies. As metagenomic sequencing is conducted on the nucleic acids extracted directly from the environment, the quantity and quality of these nucleic acids determine the outcome of these studies. If sufficient DNA is not obtained, the metagenomic data will not sufficiently represent the sampled community. Similarly, if the composition of the extracted DNA does not sufficiently represent that of the sampled community, neither will the metagenomic data.

Air samplers can affect the outcome of metagenomic studies, as their collection efficiencies can differ significantly. Although many studies have investigated sampling efficiencies of various types of air samplers (Fabian et al. 2009; Griffin et al. 2011; Li 2011; Dybwad et al. 2014; Hoisington et al. 2014), no consensus exists on standardized sampling approaches across various studies. Li (2011) compared performances of different air samplers using qPCR and fingerprinting by denaturing gradient gel electrophoresis. The authors showed that although impactor-based samplers provided higher efficiency for culturable bacteria in terms of their count and diversity, impinger- and filter-based devices both demonstrated highest efficiency for total bacteria (Li 2011). Fahlgren et al. (2011) used 16S rRNA gene sequencing to examine whether different air samplers, including an impinger-, an impactor-, and a filter-based sampler, affected outcomes. They showed that all three air-samplers provided similar representations of higher abundant bacterial species and their overall composition in the atmosphere; however, the bacterial composition and diversity were considerably different for the lower abundant species depending on the sampler (Fahlgren et al. 2011). Hoisington et al. (2014) used pyrosequencing to examine bacterial and fungal diversity in indoor air using four different air samplers, and showed that only 14% of bacterial operational taxonomic units (OTUs) and 44% of the fungal OTUs were similar in samples collected by the four different bioaerosol samplers. These studies stress the use of standardized air-sampling techniques and devices to generate data that sufficiently represent the sampled community, and to compare between different studies.

The protocols used to extract nucleic acids from the collected microbes also significantly affect the outcome of studies (Morgan et al. 2010; Momozawa et al. 2011; Yuan et al.

2012). However, no consensus exists across laboratories to use standardized protocols for extraction. Microbes have varying degrees of susceptibility to the chemicals used during DNA extraction procedures. For example, Gram-positive bacteria with thicker peptidoglycan layers in their cell walls are less susceptible to certain chemicals or treatments compared with Gram-negative bacteria. In other words, certain DNA extraction procedures might preferentially lyse specific types of cells but be biased against others. This could result in DNA samples that are not representative of the microbial communities in the environment (Yuan et al. 2012). Yuan et al. (2012) evaluated six commonly used DNA extraction methods on a mock microbial community that contained 11 different human-associated bacteria, and showed that each of the six methods provided significantly different 16S rRNA gene sequencing profiles of the mock microbial community. These profiles were also different than those of the expected mock community profile. In other words, none of the methods provided an accurate representation of the mock microbial community. The optional steps that are added to the DNA extraction protocols to help with cell lysis also contribute to additional variability. For example, protocols that use bead beating (Brodie et al. 2007; Yamamoto et al. 2012) to mechanically disrupt cells appear to provide better representations of microbial community profiles compared with those that do not (Yuan et al. 2012).

In summary, studies that use different approaches and protocols generate different outcomes concerning microbial quantity, diversity, and composition. Establishment of standardized strategies and methodologies across studies can eliminate experimental variability, and make comparison of data between studies feasible.

DNA Sequencing and Bioinformatics Related Challenges

With rapid advances in the NGS technologies, metagenomics has become one of the fastest growing disciplines in the field of environmental microbiology. In spite of well-deserved success, metagenomics still faces many challenges that need to be addressed. We will briefly review some of the relevant challenges. For more detailed descriptions, refer to the following subject reviews: Wooley et al. (2010), Teeling and Glockner (2012), and Thomas et al. (2012).

Although metagenomics has a potential to identify and characterize the entire microbial communities and their genome from the environment, the identification and characterization of potentially millions of genes in a soup of nucleic acids remains a daunting task. In contrast to cultured-based studies where the genomic contents of single known microorganisms are examined, environmental samples typically contain diverse communities of unknown species of high and low abundance whose genomes need to be sorted and identified. Theoretically, metagenomic data should be able to provide an exact representation of both high and low abundant

microorganisms in the environment; however, the higher abundant species are generally overrepresented in metagenomic data, which leads to bias. This problem can be remediated, somewhat, through deep sequencing of DNA and by increasing the depth of coverage (reads per nucleotide) to confirm that every nucleotide in the sample is sequenced at least once or preferably numerous times to ensure that low abundant microorganisms have a representation in the environmental metagenome. Deep sequencing, however, requires high quantities of starting genetic material, which is challenging to obtain from an environment such as the air with relatively low concentrations of microorganisms.

As current DNA sequencing technologies are incapable of sequencing long intact genomes, environmental DNA needs to be sheared into small fragments before metagenomic sequencing. The millions of small reads that are subsequently generated need to be reassembled de novo using advanced bioinformatics tools and software. The reassembly of these reads into contiguous fragments (contigs) is a substantial computational challenge for metagenomics. Although great progress has been made in the development of bioinformatics software and tools, the complete reconstruction of the entire genome of microorganisms in environmental samples, in particular those from complex environments, remains difficult. The read lengths obtained from the current NGS platforms typically range between 75 and 1,000 bp, depending on the platform used (Henson et al. 2012). When the read lengths are short and the depth of coverage is shallow, large gaps are introduced in the assembled contigs. Depending on the length and number of these gaps, accurate assembly of the contigs, and thus reconstruction of the entire gene sequences, can become cumbersome or even impossible. Assembly of short reads becomes even more challenging when many long repeat sequences are present in the genes, which is usually the case for bacterial and archael genomes. It is often impossible to reconstruct the entire genome of microbial communities. These challenges are surmountable through development and refinement of high throughput gene sequencing technologies, and the establishment of methodologies used to sequence longer reads with greater depth and accuracy. For example, the third generation sequencing platform, Single Molecule Real Time Sequencing technology, is the latest innovation developed by Pacific Biosciences (PacBio), which has the potential to overcome the limitations of the current gene sequencing platforms by providing longer reads, up to 30 kb with the PacBio RS II platform, with an average read length of near 3,000 bp (Wu et al. 2014). Long overlapping reads, together with increased depth of coverage, make genome reconstruction less challenging.

One current issue with the analysis of metagenomic data is the limited availability of reference sequences in genome databases, particularly for airborne microorganisms. Very few databases for environmental microbes currently exist (for a review of metagenomic databases, see Thomas et al. 2012;

Kim et al. 2013; Sharpton 2014). This complicates identification of novel genes. The solution is more metagenomic study. Metagenomic studies lead to larger genomic databases, whereas larger databases make metagenomic analysis easier.

Another issue is the current state of the art with respect to computational software and tools. New tools that can identify genome-specific markers within metagenomic data, for example, could be useful in identifying related species or strains that share high sequence similarity (Tu et al. 2014). In general, the more complex a data set, the more advanced the computational software and analytical tools needed to make sense of the data. Tremendous progress has been made in the area and further advances in analytical software and tools will make metagenomics a more practical and desirable approach to studies of environmental microbes in general and those of air in particular. For a review on the list of available software and tools for metagenomics, please refer to Bahassi and Stambrook (2014), Kim et al. (2013), and Sharpton (2014).

With continued advances in gene sequencing platforms, more complete genomic databases, and advancements in bioinformatics, many of these constraints should resolve in the near future. These new capabilities would enable fuller understanding of the potential impact of airborne microorganisms on global ecosystems.

Conclusions

Metagenomics is a powerful approach to screening microbial diversity and metabolic capabilities in any environment. The approach has proven effective in characterizing the genomic diversity and metabolic potential of many environments, including soil, marine, and the human guts, and has unraveled the vast impact of microorganisms on these ecosystems. The air harbors vastly diverse microorganisms with a potentially diverse range of metabolic activities. This frontier of microbial discoveries has largely been unexplored due to a number of challenges, which are due to the lower density of microorganisms in the air; inefficient retrieval of microbes and their nucleic acids from the air; lack of standardized approaches and methodologies; and bioinformatics-related challenges of de novo genome reconstruction. Overcoming these challenges could potentially pave the way to discoveries of novel pathways and genes important in meteorological and industrial applications, and environmental bioremediation. Metagenomic studies can additionally facilitate monitoring of airborne microorganisms through identifying pathogenic microbes and their distribution patterns and involvement in disease outbreaks that impact plant, animal, and human health.

We propose a consortium of interested parties to establish uniform approaches and methodologies designed to improve the collection efficiency, protocol reproducibility, and subsequent comparative analysis of airborne microbes. Conferences or meetings specifically designed for airborne metagenomics would provide the necessary forums for interested parties to discuss potential solutions to the present challenges.

Acknowledgment

Literature Cited

Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF. 2004. Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. J Bacteriol. 186:2629–2635.

Amato P, et al. 2005. Microbial population in cloud water at the Puy de Dome: implications for the chemistry of clouds. Atmos Environ. 39: 4143–4153.

Amato P, et al. 2007. A fate for organic acids, formaldehyde and methanol in cloud water: their biotransformation by micro-organisms. Atmos Chem Phys. 7:4159–4169.

An S, Sin HH, DuBow MS. 2014. Modification of atmospheric sand-associated bacterial communities during Asian sandstorms in China and South Korea. Heredity 114:460–7.

Bahassi E, Stambrook PJ. 2014. Next-generation sequencing technologies: breaking the sound barrier of human genetics. Mutagenesis 29: 303–310.

Bauer H, et al. 2003. Airborne bacteria as cloud condensation nuclei. J Geophys Res Atmos. 108:4658.

Be NA, et al. 2014. Metagenomic Analysis of the Airborne Environment in Urban Spaces. Microbial Ecology 69:346–355.

Bertolini V, et al. 2013. Temporal variability and effect of environmental variables on airborne bacterial communities in an urban area of Northern Italy. Appl Microbiol Biotechnol. 97:6561–6570.

Bishop JKB, Davis RE, Sherman JT. 2002. Robotic observations of dust storm enhancement of carbon biomass in the North Pacific. Science 298:817–821.

Bowers RM, et al. 2009. Characterization of airborne microbial communities at a high-elevation site and their potential to act as atmospheric ice nuclei. Appl Environ Microbiol. 75:5121–5130.

Bowers RM, McLetchie S, Knight R, Fierer N. 2011. Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. ISME J. 5:601–612.

Brodie EL, et al. 2007. Urban aerosols harbor diverse and dynamic bacterial populations. Proc Natl Acad Sci U S A. 104:299–304.

Burge HA, Solomon WR. 1987. Sampling and analysis of biological aerosols. Atmos Environ. 21:451–456.

Burrows SM, Elbert W, Lawrence MG, Pöschl U. 2009. Bacteria in the global atmosphere—Part 1: Review and synthesis of literature data for different ecosystems. Atmos Chem Phys. 9:9263–9280.

Cao C, et al. 2014. Inhalable microorganisms in Beijing's PM2.5 and PM10 pollutants during a severe smog event. Environ Sci Technol. 48: 1499–1507.

Cho BC, Hwang CY. 2011. Prokaryotic abundance and 16S rRNA gene sequences detected in marine aerosols on the East Sea (Korea). FEMS Microbiol Ecol. 76:327–341.

Chun SC, et al. 2012. The microbial population in the air of cultivation facility of oyster mushrooms. Journal of Microbiology 50:1053–1057.

Constantinidou HA, Hirano SS, Baker LS, Upper CD. 1990. Atmospheric dispersal of ice nucleation-active bacteria—the role of rain. Phytopathology 80:934–937.

Deininger CA, Mueller GM, Wolber PK. 1988. Immunological characterization of ice nucleation proteins from *Pseudomonas syringae*, *Pseudomonas fluorescens*, and *Erwinia herbicola*. J Bacteriol. 170: 669–675.

DeLeon-Rodriguez N, et al. 2013. Microbiome of the upper troposphere: species composition and prevalence, effects of tropical storms, and atmospheric implications. Proc Natl Acad Sci U S A. 110:2575–2580.

Delort A-M, et al. 2010. A short overview of the microbial population in clouds: potential roles in atmospheric chemistry and nucleation processes. Atmos Res. 98:249–260.

Dimmick RL. 1975. Evidence for metabolic activity of airborne bacteria. J Aerosol Sci. 6:387–393.

Durand S, et al. 2006. First isolation and characterization of a bacterial strain that biotransforms the herbicide mesotrione. Lett Appl Microbiol. 43:222–228.

Dybwad M, Skogan G, Blatny JM. 2014. Comparative testing and evaluation of nine different air samplers: end-to-end sampling efficiencies as specific performance measurements for bioaerosol applications. Aerosol Sci Technol. 48:282–295.

Fabian P, McDevitt JJ, Houseman EA, Milton DK. 2009. Airborne influenza virus detection with four aerosol samplers using molecular and infectivity assays: considerations for a new infectious virus aerosol sampler. Indoor Air 19:433–441.

Fahlgren C, Bratbak G, Sandaa RA, Thyrhaug R, Zweifel UL. 2011. Diversity of airborne bacteria in samples collected using different devices for aerosol collection. Aerobiologia 27:107–120.

Fahlgren C, Hagstrom A, Nilsson D, Zweifel UL. 2010. Annual variations in the diversity, viability, and origin of airborne bacteria. Appl Environ Microbiol. 76:3015–3025.

Fierer N, et al. 2008. Short-term temporal variability in airborne bacterial and fungal populations. Appl Environ Microbiol. 74:200–207.

Fierer N, et al. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc Natl Acad Sci U S A. 109:21390–21395.

Franzetti A, Gandolfi I, Gaspari E, Ambrosini R, Bestetti G. 2011. Seasonal variability of bacteria in fine and coarse urban air particulate matter. Appl Microbiol Biotechnol. 90:745–753.

Frohlich-Nowoisky J, Pickersgill DA, Despres VR, Poschl U. 2009. High diversity of fungi in air particulate matter. Proc Natl Acad Sci U S A. 106:12814–12819.

Gandolfi I, Bertolini V, Ambrosini R, Bestetti G, Franzetti A. 2013. Unravelling the bacterial diversity in the atmosphere. Appl Microbiol Biotechnol. 97:4727–4736.

Griffin DW, et al. 2011. Observations on the use of membrane filtration and liquid impingement to collect airborne microorganisms in various atmospheric environments. Aerobiologia 27:25–35.

Gurian-Sherman D, Lindow SE. 1993. Bacterial ice nucleation: significance and molecular basis. FASEB J. 7:1338–1343.

Haas D, et al. 2013. The concentrations of culturable microorganisms in relation to particulate matter in urban air. Atmos Environ. 65:215–222.

Hasegawa Y, Ishihara Y, Tokuyama T. 1994. Characteristics of ice-nucleation activity in *Fusarium avenaceum* IFO-7158. Biosci Biotechnol Biochem. 58:2273–2274.

Henson J, Tischler G, Ning ZM. 2012. Next-generation sequencing and large genome assemblies. Pharmacogenomics 13:901–915.

Hill KA, et al. 2007. Processing of atmospheric nitrogen by clouds above a forest environment. J Geophys Res Atmos. 112:1–16.

Hoisington AJ, Maestre JP, King MD, Siegel JA, Kinney KA. 2014. Impact of sampler selection on the characterization of the indoor microbiome via high-throughput sequencing. Build Environ. 80:274–282.

Hong SH, Bunge J, Leslin C, Jeon S, Epstein SS. 2009. Polymerase chain reaction primers miss half of rRNA microbial diversity. ISME J. 3:1365–1373.

Kim M, et al. 2013. Analytical tools and databases for metagenomics in the next-generation sequencing era. Genomics Inform. 11:102–113.

Kozloff LM, Turner MA, Arellano F. 1991. Formation of bacterial-membrane ice-nucleating lipoglycoprotein complexes. J Bacteriol. 173:6528–6536.

La Cono V, et al. 2011. Unveiling microbial life in new deep-sea hypersaline Lake Thetis. Part I: Prokaryotes and environmental settings. Environ Microbiol. 13:2250–2268.

Lasken RS. 2009. Genomic DNA amplification by the multiple displacement amplification (MDA) method. Biochem Soc Trans. 37:450–453.

Lee SH, et al. 2010. Identification of airborne bacterial and fungal community structures in an urban area by T-RFLP analysis and quantitative real-time PCR. Sci Total Environ. 408:1349–1357.

Li CS, Huang TY. 2006. Fluorochrome in monitoring indoor bioaerosols. Aerosol Sci Technol. 40:237–241.

Li KJ. 2011. Molecular comparison of the sampling efficiency of four types of airborne bacterial samplers. Sci Total Environ. 409:5493–5498.

Lighthart B. 2000. Mini-review of the concentration variations found in the alfresco atmospheric bacterial populations. Aerobiologia 16:7–16.

Logares R, et al. 2014. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. Environ Microbiol. 16:2659–2671.

Lowen AC, Mubareka S, Steel J, Palese P. 2007. Influenza virus transmission is dependent on relative humidity and temperature. PLoS Pathog. 3:1470–1476.

Lundholm IM. 1982. Comparison of methods for quantitative determinations of airborne bacteria and evaluation of total viable counts. Appl Environ Microbiol. 44:179–183.

Maron PA, et al. 2005. Assessing genetic structure and diversity of airborne bacterial communities by DNA fingerprinting and 16S rDNA clone library. Atmos Environ. 39:3687–3695.

Methe BA, et al. 2012. A framework for human microbiome research. Nature 486:215–221.

Mohler O, DeMott PJ, Vali G, Levin Z. 2007. Microbiology and atmospheric processes: the role of biological particles in cloud physics. Biogeosciences 4:1059–1071.

Mohler O, et al. 2008. Heterogeneous ice nucleation activity of bacteria: new laboratory experiments at simulated cloud conditions. Biogeosciences 5:1425–1435.

Momozawa Y, Deffontaine V, Louis E, Medrano JF. 2011. Characterization of bacteria in biopsies of colon and stools by high throughput sequencing of the V2 region of bacterial 16S rRNA gene in human. PLoS One 6:e16952.

Morgan JL, Darling AE, Eisen JA. 2010. Metagenomic sequencing of an in vitro-simulated microbial community. PLoS One 5(4):e10209.

Morris CE, et al. 2011. Microbiology and atmospheric processes: research challenges concerning the impact of airborne micro-organisms on the atmosphere and climate. Biogeosciences 8:17–25.

Nevalainen A, Pastuszka J, Liebhaber F, Willeke K. 1992. Performance of bioaerosol samplers—collection characteristics and sampler design considerations. Atmos Environ. 26:531–540.

Oh SY, Fong JJ, Park MS, Chang L, Lim YW. 2014. Identifying airborne fungi in Seoul, Korea using metagenomics. J Microbiol. 52:465–472.

Pikuta EV, Hoover RB, Tang J. 2007. Microbial extremophiles at the limits of life. Crit Rev Microbiol. 33:183–209.

Pouleur S, Richard C, Martin JG, Antoun H. 1992. Ice nucleation activity in *Fusarium acuminatum* and *Fusarium avenaceum*. Appl Environ Microbiol. 58:2960–2964.

Rajendhran J, Gunasekaran P. 2011. Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. Microbiol Res. 166:99–110.

Ravva SV, Hernlem BJ, Sarreal CZ, Mandrell RE. 2012. Bacterial communities in urban aerosols collected with wetted-wall cyclonic samplers and seasonal fluctuations of live and culturable airborne bacteria. J Environ Monit. 14:473–481.

Rosenfeld D, Sherwood S, Wood R, Donner L. 2014. Climate effects of aerosol-cloud interactions. Science 343:379–380.

Sangwan N, et al. 2012. Comparative metagenomic analysis of soil microbial communities across three hexachlorocyclohexane contamination levels. PLoS One 7(9):e46219.

Shaman J, Kohn M. 2009. Absolute humidity modulates influenza survival, transmission, and seasonality. Proc Natl Acad Sci U S A. 106:3243–3248.

Sharpton TJ. 2014. An introduction to the analysis of shotgun metagenomic data. Front Plant Sci. 5:209.

Smith DJ, et al. 2013. Intercontinental Dispersal of Bacteria and Archaea by Transpacific Winds. Appl Environ Microbiol. 79:1134–1139.

Teeling H, Glockner FO. 2012. Current opportunities and challenges in microbial metagenome analysis-a bioinformatic perspective. Brief Bioinform. 13:728–742.

Thomas T, Gilbert J, Meyer F. 2012. Metagenomics—a guide from sampling to data analysis. Microb Inform Exp. 2:3.

Tringe SG, et al. 2008. The airborne metagenome in an indoor urban environment. PLoS One 3:e1862.

Tu QC, He ZL, Zhou JZ. 2014. Strain/species identification in metagenomes using genome-specific markers. Nucleic Acids Res. 42:e67.

Turner MA, Arellano F, Kozloff LM. 1991. Components of ice nucleation structures of bacteria. J Bacteriol. 173:6515–6527.

Vaitilingom M, et al. 2010. Contribution of microbial activity to carbon chemistry in clouds. Appl Environ Microbiol. 76:23–29.

Vaitilingom M, et al. 2012. Long-term features of cloud microbiology at the puy de Dome (France). Atmos Environ. 56:88–100.

Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A. 95:6578–6583.

Whon TW, et al. 2012. Metagenomic characterization of airborne viral DNA diversity in the near-surface atmosphere. J Virol. 86:8221–8231.

Whyte W, Green G, Albisu A. 2007. Collection efficiency and design of microbial air samplers. J Aerosol Sci. 38:97–110.

Wooley JC, Godzik A, Friedberg I. 2010. A primer on metagenomics. PLoS Comput Biol. 6(2):e1000667.

Wu Z, et al. 2014. A precise chloroplast genome of Nelumbo nucifera (Nelumbonaceae) evaluated with Sanger, Illumina MiSeq, and PacBio RS II sequencing platforms: insight into the plastid evolution of basal eudicots. BMC Plant Biol. 14:289.

Wylie KM, et al. 2012. Novel bacterial taxa in the human microbiome. PLoS One 6(2):e1000667.

Yamamoto N, et al. 2012. Particle-size distributions and seasonal diversity of allergenic and pathogenic fungi in outdoor air. ISME J. 6:1801–1811.

Yooseph S, et al. 2013. A metagenomic framework for the study of airborne microbial communities. PLoS One 8:e81862.

Yuan SQ, Cohen DB, Ravel J, Abdo Z, Forney LJ. 2012. Evaluation of methods for the extraction and purification of DNA from the human microbiome. PLoS One 7(3):e33865.

Associate editor: Bill Martin