Monitoring of airborne biological particles in outdoor atmosphere. Part 2: Metagenomics applied to urban environments

Andrés Núñez,1 Guillermo Amo de Paz,2 Alberto Rastrojo,3 Ana M. García,1 Antonio Alcamí,3 A. Montserrat Gutiérrez-Bustillo,2 Diego A. Moreno1*

1Higher Technical School of Industrial Engineering, Technical University of Madrid, Madrid, Spain. 2Department of Plant Biology II, Faculty of Pharmacy, Complutense University of Madrid, Madrid, Spain. 3Center of Molecular Biology Severo Ochoa, CSIC-UAM, Madrid, Spain

Received 20 March 2016 · Accepted 20 April 2016

Summary. The air we breathe contains microscopic biological particles such as viruses, bacteria, fungi and pollen, some of them with relevant clinic importance. These organisms and/or their propagules have been traditionally studied by different disciplines and diverse methodologies like culture and microscopy. These techniques require time, expertise and also have some important biases. As a consequence, our knowledge on the total diversity and the relationships between the different biological entities present in the air is far from being complete. Currently, metagenomics and next-generation sequencing (NGS) may resolve this shortage of information and have been recently applied to metropolitan areas. Although the procedures and methods are not totally standardized yet, the first studies from urban air samples confirm the previous results obtained by culture and microscopy regarding abundance and variation of these biological particles. However, DNA-sequence analyses call into question some preceding ideas and also provide new interesting insights into diversity and their spatial distribution inside the cities. Here, we review the procedures, results and perspectives of the recent works that apply NGS to study the main biological particles present in the air of urban environments. [Int Microbiol 19(2): 69-80 (2016)]

Keywords: airborne biological particles · metagenomics · next-generation sequencing (NGS) · air biomonitoring · urban aerobiology

Introduction

Worldwide population is coarsely concentrated in urban environments where people are exposed to allergens and pathogens transported by the air like pollen, fungi, bacteria and viruses. Pollen and fungal spores may come from distant natural locations surrounding the metropolitan areas, while airborne active pathogenic bacteria and viruses come likely from closer sources inside the cities. We scarcely know the biodiversity present in urban areas and, strikingly, a reliable automatic system has not been developed yet in order to monitor the bioaerosols present in the air, partly because of the low concentration of the biological particles and sample collection difficulties [23]. Moreover, the different airborne biological particles are usually studied by different disciplines independently and, hence, we have a lack of information concerning the existence of relationships about their relative abundance or fluctuations to each other.

During the last years, metagenomics have increased significantly our knowledge on the biodiversity in every environment compared with conventional methods. Current technologies in sequencing are not restricted to bacteria, so any biological sample with a complex mixture of organisms can be analyzed. Thus, next-generation sequencing (NGS) offers an interesting alternative to study the metropolitan
atmosphere and uncover the real diversity present in the air. Metagenomics and the technologies related have simplified the steps required to characterize a particular environment like the urban air (Fig. 1). Here, from sample collection to taxa characterization, we review the current knowledge about airborne biodiversity in urban environments from DNA sequencing-based studies.

**Sampling procedures in NGS studies applied to urban atmosphere**

Traditional studies in microbiology for airborne bacteria and fungi are usually culture-based methods, whose bias concerning to biodiversity is widely known. Although several commercial devices are available, they are usually designed to fulfill air control legislation indoors, with fixed times and/or volumes to sample. Sometimes it is the collection surface that determines these factors, like agar media, which tent to dehydration. Unfortunately, and according to our estimations, higher sampling volumes and times than those preselected in these devices may be required to study some biological particles present in the outdoor atmosphere by Next-Generation Sequencing (NGS) [55].

In addition to vacuum filtration used in outdoor urban and also non-urban areas [8,9,18,51,70], most metagenomic studies conducted in metropolitan environments employ Particulate Matter (PM) collectors, as those operating in air quality monitoring stations [7,10,14,25,56] (Table 1). The particles are harvested in fiber filters at flows >200 l/min and the DNA is subsequently extracted. Some samplers employed typically in aerobiological studies for visual fungi and pollen identification (Hirst-type spore trap, Fig. 2) have been also recently tested in NGS studies by some authors including us [27,43]. The particles are collected on an adhesive tape and the airflow rate is significantly lower than the formers (ca. 10 l/min), but closer to human inhalation rate.

Thus, solid surfaces seem preferable to collect airborne biological particles when DNA-sequencing technologies are applied in urban environments. Only a few works have employed liquid collection [20,74], likely because water-based buffers tend to evaporate quickly, restricting the airflow rate and sampling time. However, new devices as the Spin-Cyclon used by Yooseph et al. [77] look an interesting option, with an airflow rate of 450 l/min. Fahlgren et al. [17] performed a study comparing the results obtained from three different devices (a modified impactor, a liquid impinger, and a Teflon membrane filter), concluding that there are no significant differences on the bacterial diversity and dominant species based on 16S rRNA sequences analyses. However, Hoisington et al. [33] conducted a similar survey indoors employing and comparing diverse air samplers, finding significant disparities in the OTUs of bacteria and fungi detected by each device. Thus, more studies analyzing the concordance between the results of different sampler-types for NGS analyses and also for different biological particles should be performed to confirm and standardize the sampling methodology.
In addition to the variety in sampling devices, no consensus exists about the volume of air to analyze. Strikingly, works applying NGS to urban areas have employed air samples that range from 2.5 to 5000 m$^3$ (see Table 1). Fahlgren et al. [18] for bacteria, and Fierer et al. [20] for both bacteria and fungi, were able to perform microbial identification by cloning the DNA collected in air samples of only ≤3 m$^3$. Less than 10 m$^3$ are necessary to conduct high-throughput sequencing according to the data provided by Kraaijeveld et al. [43] and Woo et al. [75], who employed different samplers and analyzed airborne eukaryotic organisms. In regards to viruses, Whon and colleagues [74] were able to carry out a metagenomic study with 54 m$^3$ of air in the near-surface atmosphere. On the contrary, Yooseph et al. [77], despite the fact that they sampled large volumes of outdoor air (5900 m$^3$ at the 22nd floor air, 101 m above street level in Midtown Manhattan, New York City) did not obtain enough DNA, highlighting the influence of the altitude on airborne biological particles abundance.

Diverse sampling times have been also selected for each author, without any correlation with the final volume of air collected (see Table 1). Consequently, each work analyzes specimens from different devices, with different volume collected and different times sampled, what may lead the results between studies to differ, especially regarding the minor representatives of the airborne communities.

As we previously reviewed [55], biological particles abundance and diversity vary significantly across the year and they are also susceptible of quick changes (in days or even hours). Therefore, sampling volume and time are two variables to consider in order to obtain a representative sample of the “airbiota” in a certain area, and it may be necessary to modify these parameters throughout the time and depending on the purposes of the study.

**DNA extraction**

In addition to the different methodology for sampling, the DNA extraction procedure is something to take into account. Quality and high concentration are desirable for NGS analysis but difficult to obtain when samples come from poorly inhabited environments like the air. Some studies comparing different methods suggest that there are remarkable differences regarding yield and purity, which could bias the further analysis [30], although it has been also exposed that DNA yield does not always correlate with differences in microbial diversity after the analysis [78]. The debate is even more controversial regarding DNA extraction for viral metagenomic studies because of the difficulty of the procedures for purifying these organisms [5,72].

While some traditional methods (phenol:chloroform extraction or CTAB procedure) are still in use to study airborne microorganisms [18,51,63,74], the tendency is to incorporate commercial DNA purification kits when NGS technology is applied [4,7,9,10,68,70,76] (see Table 1). Nowadays, commercial kits provide similar quality than handling methods, saving time and preventing manipulation issues. A common factor is to include mechanical disruption such as a bead-beating step in addition to chemical disruption [78]. Accordingly, Jiang et al. [36] have recently described a protocol for purifying DNA from particulate matter to identify airborne microorganisms by metagenomics, which includes the use of a commercial kit for the DNA extraction with some modifications.

Nonetheless, one remarkable conclusion from Hart et al.
[30] is that the method for DNA extraction must be amended for each case. This is particularly important when we require the characterization of different types of organisms (with their physical and chemical particularities) coming from a complex habitat like urban air.

Next-Generation Sequencing (NGS)

Clinical and environmental samples containing a complex combination of organisms (viruses, archaea, bacteria, fungi, animals and/or plants) can be currently analyzed at once thanks to metagenomics, saving time and avoiding the previous bias of culturing. Traditional studies in aerobiology can also take advantage of this new branch of genomics to identify fungi and plants that are hardly assigned to low taxonomic levels (genus or species). Moreover, this novel methodology is particularly interesting for those unculturable organisms or obligate-intracellular parasites such as viruses or zoonotic pathogens from the genera Rickettsia, Coxiella or Chlamydia.

Although some recent DNA-based identification studies still apply traditional molecular stages (cloning, Sanger sequencing of RFLPs, etc.), current high-throughput DNA sequencing technologies permit to skip these steps. 454 pyrosequencing (Life Science, Roche) was the pioneer platform for the so-called next-generation sequencing (NGS). While it is still in use, more recent platforms as Illumina (MiSeq/HiSeq), Ion Torrent/Proton (ThermoFisher Scientific), MinION (Oxford Nanopore Technologies) or PacBio RS system (Pacific Biosciences) have quickly replaced the former, improving the cost, timing and reliability [26].

Two main approaches exist in environmental NGS studies:

- **Targeted Amplicon Sequencing (TAS).** This strategy implies the selection of a target region of the DNA to analyze. Any region can be chosen but it is usually one present in all the organisms to study, especially in biological diversity surveys (16S rRNA gene for bacteria or 18S rRNA gene/ITS for eukaryotic organisms). The technique requires a previous PCR amplification step by using specific primers for the preferred region. “Universal primers” have been described for each group of organisms (archaea and bacteria [2], fungi [67,73], eukarya [35], plants [13,34]), and the use of degenerated oligonucleotides reduces the bias for sequence polymorphisms. Unfortunately, there is always a fraction of sequences that are amplified with less efficiency, so some groups of organisms can be underestimated or undetected while others are magnified, sloping the relative abundance. Furthermore, this methodology is limited for viral identification because these organisms do not share any common marker gene (such as 16S rRNA in bacteria or 18S rRNA in eukaryotes), so a shotgun approach is the only way to uncover viral communities [64,74]. Despite these limitations, targeted sequencing is currently widely adopted for biological diversity characterization.

- **Shotgun metagenomics** (see Sharpton [69] for a full review). The entire DNA present in the sample is sequenced without specific-region enrichment, reducing the bias introduced in the other modality. Although bioinformatics analyses after sequencing are challenging, the results from this strategy include taxonomic biodiversity (mainly from ribosomal DNA fragments) and biological functions (from DNA coding sequences). This approach, frequently used for whole genome sequencing (WGS), is especially interesting for assembling genomes, functional metagenomics, viral identification and global biome characterization, providing also absolute abundance information.

In regards to urban environments, far from exclusive, both approaches are complementary, offering critical knowledge for results interpretation. However, because of the cost, the difficulties to obtain enough DNA from some areas and the complexity of the bioinformatics analyses for shotgun analyses, TAS is the favored strategy to study microbial communities in urban airborne spaces (see Table 1).

Specific marker genes for sequencing-based analyses

Although 16S rRNA gene (SSU) is widely accepted as a target for bacterial identification in metagenomic studies, one issue that remains controversial is the region to sequence to acquire the best representation. Ideally, the complete gene should be sequenced to obtain accurate taxonomic assignments. Since the diverse hypervariable regions present in SSU show different variability they provide different level of taxonomic discrimination [50]. Several studies involving direct cloning of the 16S rRNA gene from airborne microorganisms were published between 2008-2013 [17,18,20,58,60,61]. Universal primers 27F-1492R encompassing V1-V8 regions of the 16S rRNA [45] were very suitable for these analyses, providing
the first results extracting the microbial DNA directly from the collection surface. However, the current NGS tendency is to analyze the shortest length necessary to obtain good taxonomic identification, reducing the cost for sample and time for further bioinformatic analyses [26,42].

Accordingly, fragments comprising one or several regions within V1–V5 hypervariable span have been successfully used for both environmental and clinical studies [42,47,50,52], discarding V6–V9 regions for providing less resolution. Chakravorty et al. [12] and Salipante et al. [66] have demonstrated that the V1–V3 regions are more suitable for clinical studies, distinguishing important pathogenic bacterial species from the genera Staphylococcus, Mycobacterium, Streptococcus or Haemophilus. In recent published works using NGS, the amplification of fragments covering total or partial V1–V4 regions is favored to study airborne bacterial communities in urban environments (Table 1).

Although the results from any of these regions may be acceptable within the same study, a global consensus about the specific hypervariable regions to analyze is required for comparing among different studies in order to extract correct conclusions.

Similar to bacteria, ribosomal RNA is the favorite region for fungal taxa identification. Three regions: LSU (25S-28S), SSU (18S) and internal transcribed spacers (ITS1 and ITS2), are the most popular options for DNA sequence-based studies, being the latter (ITS2) proposed as universal barcode marker by the Fungal Barcoding Consortium [67]. Only a handful of studies performed in urban environments have analyzed fungal communities by NGS technologies (Table 1). Some authors have used LSU regions, domains D1 and/or D2 [56]; and some others the 18S and/or ITS fragments [14,70,75,76].

Again, this dichotomy in the selected region makes more difficult to compare among different studies and extract significant conclusions, since the results obtained from LSU and ITS may not match completely [49].

Defining a DNA barcode for plants is a challenging task and the debate is still on. Starting with the genomic DNA region ITS2 [13,73], latest reports are focused on plastid genes such as rbcL, matK, trnH-pbsA, trnL or a combination of two for obtaining an undisputable taxonomic identification at species level [6,34]. Although affordable for individual analyses, current NGS platforms cannot integrate two distant markers at the same time. Sequencing-based identification is an attractive approach for pollen identification since morphological characters are not always sensitive enough to distinguish genus or species by microscopy (the plant families Poaceae or Chenopodiaceae, for instance). Just a few works have been recently published evaluating NGS technologies for plant identification using trnL as gene marker [43], ITS2 [40], or rbcL [24,31], being the former the only one that analyzes urban airborne samples and generates promising results for using DNA sequence-based strategy as an alternative to traditional methods for pollen identification.

As pointed above, viruses do not possess common sequences to use as a marker to design universal primers, so the shotgun approach must be adopted. In addition, the small genome size of viruses is another limiting factor for recovering enough DNA to carry out the analyses. Thus, previous random amplification procedures as Multiple Displacement Amplification (MDA) [46], or Sequence-Independent Single-Primer Amplification (SISPA) [38], are usually performed to increase the quantity of DNA or RNA, respectively. However, both techniques have some important biases [46,65]: MDA amplifies more efficiently circular DNA molecules than linear ones; and the SISPA method has a biased amplification that depends on the sequence of the primer used. Unfortunately, the best way to avoid such biases is, so far, to obtain sufficient viral biomass to sequence the sample directly. As a consequence, only Whon and colleagues [74] have confronted a complete airborne study of the viral communities in different locations in Korea using shotgun metagenomics, while two others surveys conducted in urban areas detected some viral sequences but not as a main goal [4,11].

Bioinformatics

Open-source programs and web-based tools for bioinformatics processing have been strikingly promoted by NGS technology. A complete review about the analytical tools applied in DNA sequence-base studies was published by Kim and colleagues [41]. MEGAN, RDP Classifier, Mothur, QIIME or Muscle are some examples of software suites frequently used to process high-throughput sequencing data in addition to BLAST. Some of them require an open-access database to assign taxonomic affiliations to the sequences. Greengenes for bacteria, UNITE for fungi, and Silva for bacteria and eukarya, are the most popular curated databases for stand-alone workflow, all relying on the sequences and annotations submitted to GenBank and regularly updated.

After the bioinformatics processing, DNA sequences are assigned to taxonomic classification, providing the information necessary to characterize the diversity of each biological community in our samples.
<table>
<thead>
<tr>
<th>Ref.</th>
<th>Sampler</th>
<th>Flow rate</th>
<th>Time/Vol collected</th>
<th>Collection surface</th>
<th>Extraction method</th>
<th>Org.</th>
<th>Region sequenced</th>
<th>Approach</th>
<th>Seq. technol</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>[4]</td>
<td>Portable Sampling Unit (BioWatch Program)</td>
<td>ND</td>
<td>24 h x 7 days</td>
<td>Filters</td>
<td>UltraClean Soil DNA Isolation Kit&lt;sup&gt;1&lt;/sup&gt;</td>
<td>All</td>
<td>NA</td>
<td>Shotgun</td>
<td>Illu</td>
<td>NCBI</td>
</tr>
<tr>
<td>[7]</td>
<td>High-volume aerosol sampler&lt;sup&gt;b&lt;/sup&gt;</td>
<td>250 l/min</td>
<td>24 h x 10 days (360 m&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>Quartz fibre filter</td>
<td>Kit FastDNA Spin for Soil&lt;sup&gt;2&lt;/sup&gt;</td>
<td>B</td>
<td>16S: V5-V6</td>
<td>TAS</td>
<td>Illu</td>
<td>RDP</td>
</tr>
<tr>
<td>[10]</td>
<td>High-volume aerosol sampler&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1,130 l/min</td>
<td>3 days (5,000 m&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>Quartz fibre filter</td>
<td>PowerSoil DNA Isolation Kit&lt;sup&gt;3&lt;/sup&gt;</td>
<td>B</td>
<td>16S: V2</td>
<td>TAS</td>
<td>454</td>
<td>RDP</td>
</tr>
<tr>
<td>[14]</td>
<td>High-volume aerosol sampler&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1,130 l/min</td>
<td>72 h (4881 m&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>Quartz fibre filter</td>
<td>PowerSoil DNA Isolation Kit&lt;sup&gt;3&lt;/sup&gt;</td>
<td>F</td>
<td>ITS: ITS1-2</td>
<td>TAS</td>
<td>454</td>
<td>NCBI</td>
</tr>
<tr>
<td>[21]</td>
<td>Low volume gravimetric sampler</td>
<td>38.33 l/min</td>
<td>24 h (55 m&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>PTFE filter</td>
<td>FastDNA Spin for Soil Kit&lt;sup&gt;2&lt;/sup&gt;</td>
<td>B</td>
<td>16S: V3</td>
<td>TAS</td>
<td>454</td>
<td>RDP</td>
</tr>
<tr>
<td>[25]</td>
<td>High-volume aerosol sampler&lt;sup&gt;b&lt;/sup&gt;</td>
<td>200 - 500 l/min</td>
<td>24 h (288 - 720 m&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>Quartz fibre filter</td>
<td>FastDNA Spin for Soil Kit&lt;sup&gt;2&lt;/sup&gt;</td>
<td>B</td>
<td>16S: V5-V6</td>
<td>TAS</td>
<td>Illu</td>
<td>RDP</td>
</tr>
<tr>
<td>[43]</td>
<td>Burkard volumetric spore trap</td>
<td>10 l/min</td>
<td>24 h (7.2 m&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>Adhesive tape</td>
<td>QIAamp DNA Mini Kit&lt;sup&gt;3&lt;/sup&gt;</td>
<td>P</td>
<td>plastid: TrnL</td>
<td>TAS</td>
<td>Ion</td>
<td>NCBI</td>
</tr>
<tr>
<td>[56]</td>
<td>High-volume aerosol sampler&lt;sup&gt;b&lt;/sup&gt;</td>
<td>30 l/min</td>
<td>24 h (43 m&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>0.2 μm Polyester membrane filters</td>
<td>UltraClean Soil DNA Isolation Kit&lt;sup&gt;3&lt;/sup&gt;</td>
<td>F</td>
<td>28S: D1-D2</td>
<td>TAS</td>
<td>454</td>
<td>RDP</td>
</tr>
<tr>
<td>[63]</td>
<td>8-stage non-viable impactor</td>
<td>28.3 l/min</td>
<td>4 days (163 m&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>Uncoated PCTE filters</td>
<td>phenol/chloroform protocol</td>
<td>B</td>
<td>16S: V3-V4</td>
<td>TAS</td>
<td>454</td>
<td>Greengenes, RDP</td>
</tr>
<tr>
<td>[70]</td>
<td>Vacuum filtration</td>
<td>24 l/min</td>
<td>10 h (28.8 m&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>0.45 μm cellulose ester filters</td>
<td>FastDNA Spin for Soil Kit&lt;sup&gt;2&lt;/sup&gt;</td>
<td>B</td>
<td>16S: V1-V3</td>
<td>TAS</td>
<td>454</td>
<td>EzFungi, UNITE</td>
</tr>
<tr>
<td>[74]</td>
<td>Impinger</td>
<td>19 l/min</td>
<td>48 h (54 m&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>PBS</td>
<td>Chloroform, MiniElute Virus spin Kit&lt;sup&gt;3&lt;/sup&gt;</td>
<td>All</td>
<td>NA</td>
<td>Shotgun</td>
<td>454</td>
<td>CAMERA, NCBI</td>
</tr>
<tr>
<td>[75]</td>
<td>Two-stage bioaerosol cyclone sampler</td>
<td>3.5 l/min</td>
<td>12 h (2.5 m&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>Filter</td>
<td>DNeasy Plant Mini Kit&lt;sup&gt;3&lt;/sup&gt;</td>
<td>B</td>
<td>16S: V3-V4</td>
<td>TAS</td>
<td>454</td>
<td>RDP</td>
</tr>
<tr>
<td>[76]</td>
<td>Two-stage non-viable Andersen sampler</td>
<td>28.3 l/min</td>
<td>4 weeks (1,141 m&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>Glass fiber filter</td>
<td>PowerMax Soil DNA Isolation Kit&lt;sup&gt;1&lt;/sup&gt;</td>
<td>F</td>
<td>ITS: ITS1-2</td>
<td>TAS</td>
<td>454</td>
<td>NCBI</td>
</tr>
<tr>
<td>[77]</td>
<td>SpinCon HPAS 450-10A Wet Cyclone Portable Air Sampler</td>
<td>450 l/min</td>
<td>7 days (5,900 m&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>PBS</td>
<td>MO BIO PowerWater Kit&lt;sup&gt;1&lt;/sup&gt;</td>
<td>All</td>
<td>NA</td>
<td>Shotgun</td>
<td>454</td>
<td>NCBI</td>
</tr>
</tbody>
</table>

<sup>1</sup>: MoBio Laboratories; <sup>2</sup>: MP Biomedicals; <sup>3</sup>: Qiagen; NA: Not applicable; ND: Not described; TAS: Targeted amplicon-sequencing; NCBI: National Center for Biotechnology Information; RDP: Ribosomal Database Project. Organisms: B (Bacteria); F (Fungi); P (Plants); E (Eukaryotes). Sequencing Technology: Illu (Illumina); 454 (454-Roche); Ion (ion Torrent). <sup>a</sup>: Estimated when not specified; <sup>b</sup>: Particulate matter collector
Lessons from DNA-sequencing studies in urban spaces

Abundance of microorganisms in the air. It is well characterized that some biological entities present in the air suffer important seasonal fluctuations as the noticeable variation in abundance and composition of the pollen grains throughout the year [28].

Likewise, culture-dependent studies performed in urban areas suggest that fungi and bacteria are, in general, more abundant in summer than in winter [19,37,54]. Recent culture-independent studies are in concordance with these results. Woo et al. [75] described that there is a peak of DNA concentration in the air in urban environments during summer season (June to September) independent on rural or urban environments. Consistently, some other authors have used different approaches as DAPI [10], shotgun sequencing [4] or qPCR [7] to show that airborne bacteria abundance is higher in summer, only outnumbered in spring by the group Plant-Fungi [4]. Moreover, both Dannemiller et al. [14] and Fröhlich-Nowoisky et al. [22] also agree on the fact that fungal richness is higher in spring, being reduced in winter.

However, Fierer et al. [20] showed that some airborne organisms are submitted to more rapid variations in metropolitan areas. In this study, they observed that the abundance of bacteria and fungi changes significantly within a period of 10 days, based on cloned sequences. Similar results were obtained by Shin et al. [70] and Oh et al. [56] using NGS. Furthermore, some culture-based studies have reported that daily precipitations can affect fungal abundance during the following days [32], and even diurnal oscillations for bacteria have been described [19,48,79]. Thus, some divergences between studies may be obtained depending on environmental factors as the weather or the time of day when the sampling is performed.

Relative abundance of different taxa. Contrary to culture-based studies in which spore-forming Gram-positive bacteria (e.g. Bacillus, Micrococcus) are usually the most abundant group identified outdoors [44,48,79], NGS has proved this is not necessarily real, showing an unexpected diversity of Gram-negative bacteria [18,25,75,77]. A general conclusion (independent on shotgun or TAS approaches) is that the phyla Actinobacteria and Proteobacteria are the most abundant in urban outdoor environments, followed by Firmicutes and Bacteroidetes [7,18,21,25,70,75].

Interestingly, controversial results using NGS have been found within the fungal kingdom. It is unclear which group is more abundant in metropolitan spaces. Several authors have reported that the abundance of Ascomycetes in metropolitan spaces is greater than Basidiomycetes, and they are at all seasons [14,56,75]. The opposite is shown by other authors like Shin et al. [70] and Pashley et al. [57], supported by morphological identification, and these contradictory results are independent on methodology (cloning, TAS, shotgun) or the DNA region considered (LSU, ITS or 18S). However, as underlined above, some works suggest that significant changes in fungal abundance may occur within short periods of time (days or even hours). Since most of NGS analyses only register a few isolated days, both conclusions may be plausible. To shed some light on the matter, some morphology and culture-based surveys performed systematically along the year concluded independently that Ascomycetes spores are the main group in urban areas, being Cladosporium the most abundant fungal spore [15,16].

Additionally, it is widely accepted that Dothideomycetes (in which Cladosporium is included) is the most represented taxa in Ascomycetes, as Agaricomycetes is within the Basidiomycetes group [14,70,76]. Moreover, a positive correlation has been observed between relative humidity and Basidiomycetes spores. Analyzing the LSU region of the fungi present in urban atmosphere, Pashley and colleagues [57] found an increment of these microorganisms during wet days, what partly could explain the differences among studies.

Regarding airborne viruses, Cao et al. [11] and Be et al. [4] have been able to detect viral sequences from urban airborne samples, mainly bacteriophages and some human-related viruses as herpesviruses and Adenovirus C. Additionally, most of the viruses identified by Whon and coworkers [74] in the city of Korea were related to geminivirus, circovirus, microvirus, nanovirus and bacteriophages (Caudovirales).

The influence of meteorological factors is a challenge to address in urban environments employing NGS technologies. While traditional studies agree temperature and wind positively correlate with an increment of bacterial abundance according to counting and culture-based results [29,53], some NGS studies found no correlation with meteorological parameters. Shin et al. [70] did not find a direct correspondence between the abundance of microbial taxa and temperature or relative humidity during the study in childcare facilities (indoors and outdoors), neither did Bowers et al. [9] in bacterial composition from different land-use sites. Seasonal differences in microbial communities have been also detected [10,18,25,74], however, several parameters change altogether throughout the year so it is difficult to analyze the influence of each one separately.

The attempts to detect pathogenic organisms and
allergens using high-throughput sequencing technologies deserve a special remark. During the study conducted by Woo et al. [75] using TAS approach, they were capable of identifying members of the genera Legionella, Salmonella, Staphylococcus (potentially dangerous), in addition to Clostridium perfringens and Escherichia coli O157:H7 sequences. The same tactic was followed by Shin et al. [70] to detect fungal allergens such as Cladosporium and Alternaria both indoors and outdoors. However, the best strategy may be shotgun metagenomics, employed by Be et al. [4] and Cao et al. [11]. Thus, sequences from Streptococcus pneumoniae, Klebsiella pneumoniae, Staphylococcus epidermidis and fungi Alternaria, Cladosporium, and Aspergillus fumigatus can be detected in the same analysis and additional information about absolute abundance is provided.

**Differential distribution of microorganisms in urban spaces.** One interesting subject is whether particular places in the city have singular microbial communities and its correlation with anthropologic activities. Prussin et al. [62] studied virus-like and bacteria-like particles (VLPs and BLPs) concentration indoors and outdoors at different facilities, showing that the air composition outdoors (with higher VLPs counts) is the main source influencing the concentration of the indoor particles, both VLPs and BLPs. Similarly, Amend and coworkers [1] studied the fungal presence at different sites and countries, analyzing the ITS sequences extracted from dust samples. Besides a latitude connection, they concluded that fungal communities indoors are highly influenced by the outdoor environment, not finding differences among nearby buildings with diverse human activities. Some comparative studies performed at schools and childcare facilities support the idea that fungal populations are similar indoors and outdoors, while bacterial diversity differs due to human occupancy [63,70]. Thus, amplicon-sequencing results have shown common bacterial communities in both environments (indoors and outdoors) with species from Rhodobacteria and the genera Sphingomonas and Pseudomonas, but an increase in the relative abundance of human skin-associated bacteria indoors: Staphylococcus, Corynebacteria or Propionibacteria [18,63,70].

It has been also proposed that areas with high traffic density or sewage pollution have much higher concentrations of airborne bacteria [19,29], suggesting a strong influence of the antropogenic activities as a source of particular airborne biological communities. Accordingly, a metagenomic study conducted by Yooseph and colleagues [77] confirmed a common bacterial population in urban atmosphere with remarked abundance of some genera depending on the use of the building: Klebsiella and Bordetella in hospitals or marine-related Planctomyces, Pirellula and Synechococcus in piers. Alike, viral communities in the air of Korea can be distinguished based on the land use according to the study of Whon et al. [74].

Barberán and colleagues [3] performed a large-scale study collecting samples over a thousand houses in USA and evaluated dust-associated microorganisms by 16S and ITS analysis. Although they found that microbial communities compositions were highly variable across the United States (explained by climatic and soil variables), their results pointed out that microbial communities in urban air tend to homogenization, with less variability compared with rural areas (in accordance with Bowers et al. [9]). Additionally, Prussin et al. [62] observed that the concentration of VLPs and BLPs was similar between different city locations, supporting that the air is, in fact, a homogeneous fluid distributing microorganisms through all the spaces in the city and providing a common base of microbial diversity.

Nonetheless, despite its singularities as a microbial biome, urban environments are susceptible to external incomes from sporadic events like dust storms. Several studies have analyzed the biological diversity associated to these particles coming from long-distance locations as those conducted by Katra et al. [39] and Maki et al. [51]. Though performed in different regions, they agree the richness of DNA sequences in the air belong to prokaryotes and eukaryotes during these events is significantly increased with uncommon taxa that usually correlates with the soil and vegetation of the dust origin.

Taken all these studies as a whole, they clearly highlight the complexity of the urban airborne dynamics and the requirement to normalize the study procedures to reach a better comprehension of microbial communities in metropolitan environments. In perspective, these results remark the critical importance of the sampling organization (time, volume, synchronized sampling, etc.) and additional annotation regarding meteorological factors, human activity, microbial sources, etc. in order to obtain conclusive and comparable results.

**Conclusions and final remarks**

Recent studies have shown unexpected roles of the biological particles in the air, proving how important is to get a better knowledge of this habitat for environmental and clinical
reasons. The first challenge to face during the characterization of these airborne particles using NGS technologies is the low abundance of them and, as a result, low amount of DNA. Two approaches can be adopted to solve this problem: (i) to collect large volumes of air in a short time by using high volume cyclone samplers-type; (ii) to sample at lower flows but longer times. Assuming that the abundance and diversity of bioaerosols can shift very fast, the second option seems more appropriate to obtain a more representative sample. Even so, the first approach can be suitable for detecting specific pathogens or allergens in a particular site. Additionally, sampling at different weeks in a season can provide significant information to get a better characterization of the airborne biological diversity in a particular location. Either way, sampling methodology and devices should be adapted to the aims of the study.

Until shotgun metagenomics becomes economically more affordable, TAS is still a good proxy despite the loss of functional genes information and absolute abundance. V1-V4 regions within the 16S rRNA gene for bacteria and ITS2 for fungal propagules and spores are the most favored regions for identification and taxonomic assignation in airborne studies (see Table 1). In regards to pollen/plant identification, the discussion about the DNA barcode is still open. Up to now, Kraaijeveld and colleagues [43], using trnL as gene marker, have published encouraging results from metropolitan environments comparing with morphological determination. We have also tested the 18S rRNA gene to perform similar assays [27], although poor resolution at genus or species level can be obtained from this marker. More recently, we have evaluated the resolution of ITS2, confirming that this region is more suitable as a genomic marker and overcome the trouble with 18S gene (unpublished data).

Currently, Illumina is the preferred platform for high-throughput DNA sequencing in most researches. However, technical procedures are evolving and advancing strikingly fast. Ion Torrent, PacBio RS system or MinION are very promising platforms but they still need to prove their value in the environmental field.

The studies under reviewed suggest that microbial diversity of urban environments holds singular features, as a particular biome. Fungal, viral and majorly bacteria communities are under the influence of human presence and the building use [63,70,74,77]. Both culture-dependent and -independent studies indicate that the spring and summer seasons correlate with higher abundance and richness of microorganisms in the air and, as a result, an increase in DNA concentration. However, some apparent contradictions can be found, especially when the abundance of organisms from different taxa are examined. The abundance of some biological entities in the air like bacteria or fungi can shift quickly [20,48,79]. Since most NGS studies are usually restricted to a few discrete days or even a few hours, some disagreements may be expected. Consequently, the application of combined techniques (sequence-based and traditional) to solve the discrepancies is the best strategy until sequencing becomes complete trustworthy, and standard procedures are established. In accordance, NGS yields an overwhelming amount of data, which is processed essentially by computers. Reviewing the results to confirm they are plausible is essential and a multidisciplinary collaboration among microbiologists, botanists and bioinformaticians is highly recommended to curate the NGS outcomes.

On the other hand, culture-independent studies have proved to be useful to detect allergens and airborne pathogens [4,11,70,75]. However, one intrinsic weakness of DNA sequencing methods is that it cannot be distinguished between alive and dead, or complete and fragmented biological particles. For pollen and fungal allergens, even fragments can induce clinical symptoms when a threshold is overpassed [71]. In contrast, pathogenic fungi, viruses, bacteria or resistant spores usually need to be metabolically active to induce any disease. In both cases, any approach for quantification may be quite helpful. The results obtained from TAS permit to calculate relative abundance but we must take into consideration that it is based on several prior PCR amplifications. Although the number of cycles of these reactions is kept at minimum to keep the proportional ratio, the conclusions on relative abundance must be carefully considered. Moreover, the values of abundance can be biased for the number of copies of the selected DNA region. Several fungal genera (e.g. Alternaria, Leptosphaeria, Stemphylium, Pleospora) produce multicellular spores, so they contain several copies of genomic DNA, magnifying their representation. A similar effect exists in pollen grains from polyploid species when sequences from genomic DNA are chosen (SSU, LSU or ITS). Likewise, plastid sequences are controversial because of the number of chloroplasts that can be found in pollen grains or if they are even present in the pollen of all species [6,24,43]. As a result, DNA-based scores compared with morphology surveys might differ, so novel and unexpected results must be supported by other methods.

Some attempts to infer absolute abundance from the sequencing outcome have been proposed by some authors [14,59], and recently the RDP Classifier have implemented an algorithm with a gene-copy-number adjustment for bacteria.
and fungi to make the analyses more quantitative. Yet, more studies evaluating the effectiveness of these approaches are needed.

Overall, high-throughput sequencing is an outstanding technology with an extraordianry potential to come. Although some adjustments are needed to apply this methodology to metropolitan environments, its qualities for easy identification of any type of organism at once, flexibility to adapt to specific goals and its potential for pathogens and allergens identification make NGS a promising tool for real-time bioaerosols monitoring and revealing the air genome.

Acknowledgements. This study was funded by the Community of Madrid, Spain, under the AIRBIOTA-CM Program (S2013/MAE-2874).

Competing interests. None declared.

References

27. Gutiérrez-Bustillo AM, Sáenz C, Aránguez E, Ordóñez JM (2001) Polen...
atmosférico en la Comunidad de Madrid. Documento Técnico de Salud Pública No. 70. Consejería de Sanidad, Comunidad de Madrid, 204 pp. [In Spanish]


