Clouds as atmospheric oases

Raphaelle Peguilhan

UNIVERSITÉ CLERMONT AUVERGNE
U.F.R de chimie

ÉCOLE DOCTORALE DES SCIENCES FONDAMENTALES

THÈSE

Présentée pour obtenir le grade de

DOCTEUR D'UNIVERSITÉ
Spécialité : Sciences de l'atmosphère

Par **Raphaëlle PÉGUILHAN**
Master Génomique et Environnement de l'Université Paris-Saclay

Les Nuages : Oasis de l'Atmosphère
Clouds as atmospheric oases

Défendue le 16 Septembre 2022 devant la commission d'examen :
Defended on September 16, 2022 before the review board:

Directeur de thèse/Thesis supervisor:

Pierre AMATO	Chargé de Recherche - Institut de Chimie de Clermont-Ferrand, CNRS/Université Clermont Auvergne (Clermont-Ferrand, France)

Présidente du jury/President of the jury:

Nathalie HURET	Professeur - Observatoire de Physique du Globe de Clermont-Ferrand, CNRS/Université Clermont Auvergne (Clermont-Ferrand, France)

Rapporteurs/Reporters:

Tina ŠANTL-TEMKIV	Assistant Professor - Department of Biology, Aarhus University (Aarhus, Denmark)
Ludwig JARDILLIER	Professeur - Institut Diversité Ecologie et Evolution du Vivant, CNRS/ Université Paris-Saclay (Gif-Sur-Yvette, France)

Examinateurs/Examiners:

Achim QUAISER	Maître de Conférences - Laboratoire Ecosystèmes Biodiversité Evolution, CNRS/Université de Rennes 1 (Rennes, France)
Françoise LUCAS	Professeur - Laboratoire Eau Environnement et Systèmes Urbains, Université Paris-Est Créteil (Créteil, France)

Invités/Guests:

François ENAULT	Maître de Conférences - Laboratoire Microorganismes : Génome Environnement, CNRS/Université Clermont Auvergne (Clermont-Ferrand, France)
Laurent DEGUILLAUME	Physicien - Laboratoire de Météorologie Physique, CNRS/Université Clermont Auvergne (Clermont-Ferrand, France)

Acknowledgment/Remerciements

First of all, I would like to thanks Tina Šantl-Temkiv, Lugwig Jardillier, Nathalie Huret, Françoise Lucas and Achim Quaiser for accepting to review my work.

I will continue in French for the following acknowledgement.

Tout d'abord, un grand merci à mon cher encadrant de thèse Pierre Amato. Tu t'es toujours montré disponible et compréhensif. Tu as su m'accompagner pendant ces 3 ans tout en me laissant la liberté nécessaire pour m'approprier ce beau sujet de thèse. Je te remercie aussi de m'avoir supporté même pendant mes moments les plus émotifs. Je pense que j'aurais difficilement pu imaginer meilleur encadrant de thèse, cela a grandement contribué à mon épanouissement dans mon travail et dans l'équipe. Nos discussions me manqueront !

Je souhaiterais remercier également Laurent Deguillaume et François Enault pour leur participation à mes comités de thèse informels et leur précieux conseils. Je voudrais exprimer toute ma gratitude plus particulièrement à Laurent pour son soutien lors des campagnes de prélèvement au puy de Dôme. Cela restera pour moi des moments inoubliables. Je voudrais également exprimer toute ma gratitude à François pour son soutien en bioinformatique et sa disponibilité malgré un emploi du temps chargé. Sa présence dans l'équipe puis lors des déjeuners aura bien égaillé mon passage au laboratoire !

Je tiens à remercier toutes les personnes avec qui j'ai pu collaborer lors de mon travail de thèse, telles que Bérénice Batut et Engy Nasr pour cette collaboration inattendue et très sympathique visant à améliorer le workflow bioinformatique que j'ai pu mettre en place lors de mon travail de thèse. J'espère qu'il permettra d'aider d'autres personnes comme moi avec leur jeu de données complexe. Je remercie également Pierre Peyret et Sophie Marre pour leur collaboration sur le projet de capture de gène et pour notre collaboration plus récente, et toujours en cours, visant à améliorer l'affiliation taxonomique de mes données de métagénome et métatranscriptome. Je voudrais également remercier Olivier Rué pour notre collaboration récente visant à améliorer l'affiliation taxonomique de mes données d'amplicon. Merci pour ces interactions très efficaces et constructives.

Je voudrais également remercier Pascal Renard et Angelica Bianco pour leur aide lors des campagne d'échantillonnage au puy de Dôme, ainsi que Jean-Luc Baray pour nos interactions et ses réponses efficaces et constructives envers nos nombreuses demandes concernant les données météorologiques du puy de Dôme. Un grand merci tout particulier à Pascal toujours très réactif et prêt à nous emmener échantillonner à toute heure !

J'aimerais exprimer ma gratitude à tous les doctorants, post-doctorants et CDD qui ont contribué à m'accueillir dans l'équipe et à l'atmosphère familiale du bureau malgré les hauts et les bas qu'il y a pu avoir : Clément Descarpentries, Christelle Ghaffar, Florent Rossi, Maxence Brissy, Saly Jaber, Léa

Gourbeyre, Hubert Casajus, Laurie Calarnou ; et plus récemment, Jon Vyskocil, Léo Paulat, Zeina Bourhane et Leslie Nunez. Un merci particulier à Florent Rossi qui m'a beaucoup aidé pour la collecte et le traitement des échantillons, et également pour sa bonne humeur et son côté toujours volontaire ! Je remercie aussi les stagiaires qui ont contribué à cette bonne ambiance : Thomas Charpentier et Antoine Lafont.

Je tiens également à remercier chaleureusement Muriel Joly et Céline Judon du service microbiologie pour leur soutien et toutes les conversations drôles et sympathiques qu'on a pu avoir !

Je remercie Barbara Ervens pour ces conseils lors de nos réunions.

Je remercie Lionel Nauton pour ces conseils et le support informatique.

Je remercie Clarisse Malet pour avoir accepté d'être ma marraine de thèse et pour nos discussions.

Je voudrais exprimer également toute ma reconnaissance à l'équipe BIOMETA, en particulier à Mounir Traikia, Boris Eyheraguibel, pour leur accueil durant ma thèse et pour tous les bons moments partagés.

Enfin, je remercie très chaleureusement ma famille, mon tendre et cher Thibaud et mes sœurs de cœur MC et Laurène pour leur soutien infaillible et tout le réconfort qu'ils ont pu m'apporter dans les moments les plus difficiles.

Contents

Abbreviations list

ABL: Atmospheric Boundary Layer (or PBL: Planetary Boundary Layer)

ATP: Adenosine Triphosphate

BLH: Boundary Layer Height

CCN: Cloud Condensation Nuclei

CO-PDD: Cézeaux-Aulnat-Opme-Puy de Dôme platforms

DNA: Deoxyribonucleic Acid

GO: Gene Ontology

HFRi: High-Flow-Rate impinger

IN: Ice Nuclei

MCE: Mixed Cellulose Esters

MDA: Multiple Displacement Amplification

MG: Metagenome/Metagenomics

MT: Metatranscriptome/Metatranscriptomics

NAP buffer: Nucleic Acid Preservation buffer

NGS: Next Generation Sequencing

OPM: Opme station

PBAP: Primary Biological Aerosol Particle

PCR: Polymerase Chain Reaction

PUY/PDD: Puy de Dôme Mountain station

QC: Quality Control

RNA: Ribonucleic acid

SBA: Secondary Biological Aerosol

SSU: Small Sub-Unit

TCA cycle: Tricarboxylic acids cycle

a.s.l.: above sea level

a.g.l.: above ground level

Laboratories:

ICCF: Institut de Chimie de Clermont-Ferrand UMR6296-CNRS-UCA / *Institute of Chemistry of Clermont-Ferrand*

LAMP: Laboratoire de Météorologie Physique UMR6016-CNRS-UCA / *Physical Meteorology Laboratory*

LMGE: Laboratoire Microorganismes : Génome et Environnement UMR6023-CNRS-UCA / *Microorganisms : Genome and Environment Laboratory*

OPGC: Observatoire de Physique du Globe de Clermont-Ferrand UMS833-CNRS-UCA / *Observatory of Earth Physics of Clermont-Ferrand*

General introduction

The outdoor atmosphere is a complex and highly dynamic environment where emitted sources of microorganism mix over short and long distances to form mosaic airborne communities. The average residence time of microorganisms in the atmosphere is one day to a week (Burrows et al., 2009a), demonstrating the variable and ephemeral nature of the atmospheric microbial assemblage in terms of biodiversity and richness. These airborne microorganisms are known to be present and viable in aerosols (i.e., dry or partially dry atmosphere) since the discovery of Louis Pasteur at the late 19th century (Pasteur et al., 1878). However, the *in situ* activity of aerosolized microorganisms was demonstrated much later with Krumins *et al.*, (2014), and the potential of clouds as microbial habitats was investigated in the 90s with Fuzzi et al. (1996) and then with Sattler et al. (2001). This lack of progress for decades in the field of aeromicrobiology can be explained by the complexity of the atmosphere as a subject of study, particularly clouds. Moreover, methods for analyzing general ecosystems activity are new to the field and still in development (e.g., metatranscriptomics). The subject has also lacked consideration for a long time, but recently the interest is increasing with the new concerns related to the environment, health, and climate.

Recent studies have focused on aerosols and active biodiversity with RNA-based analyses (Klein et al., 2016; Šantl-Temkiv et al., 2018; Womack et al., 2015). The total active communities in clouds was first studied in (Amato et al., 2017). However, the functional profile of airborne communities in aerosols is still unknown, and is in its beginning in clouds with a first exploration of their microbial functioning made by Amato *et al.* (2019) with the use of metatranscriptomics (MT) and metagenomics (MG). This study includes three cloudy events and is based on amplified metagenomes and metatranscriptomes by MDA. NGS sequencing introduces unavoidable quantification biases, and these biases are increased by amplification such as MDA (Yilmaz et al., 2010).

My thesis work is a continuation of the efforts initiated in Amato *et al.* (2019) and aims **to investigate whether clouds are specific microbial entities in the atmosphere, providing condensed water to cells, like oases in an atmospheric desert. To this end, cloudy situations were compared to other atmospheric compartments such as aerosols and precipitation in terms of biodiversity and functional profile.**

This thesis work was part of the project “Modeling biologically-driven processes in clouds” (MOBIDIC) which is related to the ANR plan “Make our planet great again” (MOPGA). This project aims including biological processes in atmospheric chemistry and physics models.

This manuscript is divided into four chapters, with first (**Chapter 1**) a state of the art of current knowledge on microorganisms in the outdoor atmosphere and clouds.

Then, the results are presented as three chapters:

Chapter 2 presents the singular sampling sites and the experimental procedure used during my thesis work. Details on the validation of the protocols and on the multiple negative and positives controls performed are presented. The results section is divided into two articles written for submission to scientific journals. The first one develops the validation of the experimental procedure for the analysis of nucleic acids on the atmosphere. Good sampling and data processing practice are also recalled. The second article focus on our first attempt to compare clouds to aerosols. Amplicon sequencing was used to investigate bacterial diversity in both atmospheric compartments. These two articles support the work presented in **Chapter 3** and **4**.

Chapter 3 describes our second sub-objective, which is to compare clouds and precipitation. Bacterial communities are also studied by amplicon sequencing, and efforts were made to sample to sample precipitation and the associated cloud in a coordinated manner to reveal direct links between the two compartments. This part of the results has been published and the chapter presents an introduction to the study, the published paper and a conclusion with additional results.

Chapter 4 directly pursues what was started in the study by Amato *et al.* (2019) with several improvements. Here, metagenomes and metatranscriptomes were obtained without prior amplification, limiting quantification bias. Clouds were compared to aerosols to highlight possible microbial functional specificities in the presence of condensed water. Thus, the chapter first contains an introduction to the methods used (metagenomics and metatranscriptomics) and their specific challenges, with a description of the bioinformatic workflow developed to process this very complex dataset. Then, a short introduction to this study is proposed, followed by the functional study written for submission, and a conclusion of this chapter.

Finally, a general conclusion with perspectives are proposed. An extended abstract in French is also present at the end of the manuscript

This thesis work involved the collection of environmental samples for *in situ* analysis of clouds, aerosols and rain, sample processing, and raw data analysis with a large part of bioinformatics processing. This was a multidisciplinary work implying knowledge of atmospheric physics and chemistry (especially clouds), and skills in field sampling, molecular biology and bioinformatics.

Multiples collaborations were undertaken with laboratories in France and Germany to improve the processing of the unique datasets obtained during this thesis project.

Chapter 1: Literature review

1. Structure of the atmosphere

1.1. The Earth's atmosphere

The Earth's atmosphere is composed of multiple atmospheric layers, the troposphere being that of lower altitude (0 to 10-20 km) in contact with surface ecosystems and notably characterized by a decrease of temperature with increasing altitude (American Meteorological Society - AMS, 2012). The troposphere includes the atmospheric boundary layer (ABL; or planetary boundary layer, PBL) characterized by vertical turbulence due to relief and temperature gradients. This atmospheric layer is strongly influenced by local emission sources (natural and anthropogenic). Its depth is variable in time and space (several meters to ~2 km) depending mostly on the diurnal cycle and surface temperature (Garratt, 1994). Thus, the boundary layer height (BLH) is greatest on hot summer days, while it is lowest on cold winter nights. The free troposphere, on the other hand, is mainly subject to horizontal air movements on a continental and planetary scale, the air masses encountering no obstacles. The free troposphere is also the atmospheric layer that houses the water cycle, with the formation of clouds and precipitation (Figure 1).

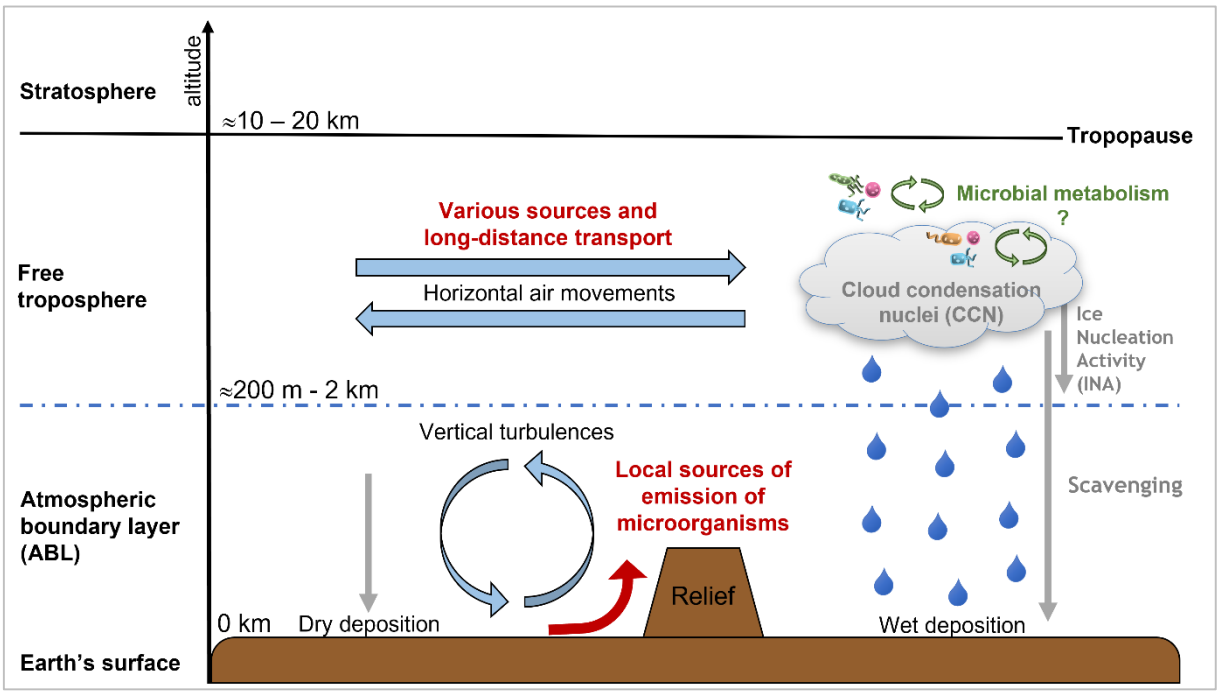


Figure 1: Schematic representation of the troposphere layers with the main atmospheric phenomena. Bleu arrows: major air movements; red arrow: aerosolization of microorganisms; grey arrows: atmospheric physical phenomena; green arrows: biological processes.

1.2. Bioaerosols

The term aerosol refers to every solid and/or liquid particles in suspension in the atmosphere (Fuzzi et al., 2006). Differentiation is done between primary aerosols, particles emitted directly into the atmosphere, and secondary aerosols which are generated in the atmosphere by condensation of gaseous precursors. Among primary aerosols, some have a biological origin and are called bioaerosols (or primary biological aerosol particles, PBAP). They contain dead or alive organisms, isolated or aggregated and their fragments and products: animal and plant debris, pollen, biofilm fragment, bacterial and fungal cells and spores, archaea, viruses, etc. (Després et al., 2012). PBAP size is highly variable with a range from 1 nm (viruses and cell's fragment) to 100 µm (pollen, plant debris) (Després et al., 2012; Fröhlich-Nowoisky et al., 2016).

PBAPs can impact cloud chemistry by diluting into the aqueous phase (Marinoni et al., 2004; Sellegri et al., 2003), or through microbial metabolisms (Khaled et al., 2021); they can also impact the physics of the atmosphere by serving as cloud condensation nuclei (CCN) or ice nuclei (IN) (Möhler et al., 2007; Zhang et al., 2021a) (**Figure 1**). The atmospheric transport of PBAPs plays a very important role in the spread of microorganisms and in the exchange of genetic material between geographically distant areas. PBAPs are thus, key elements in the development, evolution, and the dynamics of ecosystems (Fröhlich-Nowoisky et al., 2016).

Concerning secondary aerosols, atmospheric modeling has estimated that secondary biological aerosol (SBA) could be produced in cloud droplets by the multiplication of cells and impact the chemical processes of clouds in the same way as PBAPs (Ervens and Amato, 2020).

1.3. Clouds

1.3.1. Insoluble particles

Clouds are composed of condensed water droplets in suspension in the atmosphere formed around aerosols, referred to as cloud condensation nuclei (CCN). Not all aerosol particles are CCN-active, depending on their size and hygroscopicity (Petters and Kreidenweis, 2007). The larger a particle is, the greater its radius of curvature and surface area available for water condensation (cf. Koehler theory; Wex et al., 2008). The size distribution and chemical composition of these aerosols will thus be able to influence the processes of cloud formation and development (Asmi et al., 2012; Matthias-Maser et al., 2000). It is estimated that cloud droplets have a mean diameter of several micrometers (~3 to ~50 µm) (Miles et al., 2002). Atmospheric particles which can form CCN are: dust and volcanic dust, marine salt, soot, and other particles in suspension in the atmosphere, as well as PBAPs (Matthias-Maser et al., 2000). Approximately 25% of insoluble particles in clouds are biological and this proportion increases with the rate of anthropogenic sources in the air mass (Matthias-Maser

et al., 2000). The microorganisms, can serve as CCN and participate, to a certain extent, in the formation of cloud (Després et al., 2012; Fröhlich-Nowoisky et al., 2016; Hamilton and Lenton, 1998; Mikhailov et al., 2021). Moreover, certain bacteria are known to have the best IN activity among IN particles, causing water to freeze at -2°C for the bacterium *Pseudomonas syringae*. The IN activity could participate in the induction of precipitation (Morris et al., 2004; Sands D.C, 1982), and thus impact the physics of clouds.

1.3.2. Soluble material

Regarding soluble compounds in cloud water, the major organic and inorganic ions found (from several to thousands µM) are acetic, formic, succinic, malonic and oxalic acids, as well as Cl^-, NO_3^-, SO_4^{2-}, Na^+, NH_4^+, K^+, Mg^{2+} et Ca^{2+} ions (**Table 1**) (Collett et al., 1993; Renard et al., 2020). Ion concentrations are variable between and within clouds. The liquid water content (LWC) is one of the main factors influencing the concentration of the liquid phase, along with the origin of the chemical sources and their proportion (Marinoni et al., 2004). There are concentration gradients depending on the size of the droplet and the height in the cloud (Petrenchuk and Drozdova, 1966), but still little is known about this subject.

Table 1: Minimal, maximal and average concentrations of organic and inorganic ions (µM) and pH in cloud water in Central France area (data from Renard et al., 2020).

Organic ions (µM)								
	Acetate	Succinate	Malonate	Oxalate	Formate			
Min	0	0	0.08	0	0.25			
Max	71.24	33.02	3.5	377.24	109.63			
Median	7.07	0.48	0.41	1.59	8.53			
Inorganic ions (µM)								
	Cl^-	NO_3^-	SO_4^{2-}	Na^+	NH_4^+	K^+	Mg^{2+}	Ca^{2+}
Min	0.16	0.8	0.49	0.37	2.19	0	0	0
Max	409.52	516.51	247.35	678.56	531.13	159.42	47.92	602.14
Median	20.24	33.34	22.66	20.78	65.75	4.28	4.33	9.14
pH								
Min	3.1							
Max	7.6							
Median	5.59							

The chemical composition of the cloud droplets is also determined by chemical and physical processes such as the dynamics of cloud formation, the composition and concentration of aerosols that dissolve in the aqueous phase from CCN or impaction scavenging, the transfer of volatile species through the air/water interface during the lifetime of the cloud, and the chemical reactions that occur in the liquid phase (Marinoni et al., 2004). According to their chemical composition, a classification of clouds could be determined at Puy de Dôme mountain (1,465 m a.s.l.; France) using the concentration of the six main ions (Cl^-, NO_3^-, SO_4^{2-}, Na^+, NH_4^+ et Mg^{2+}), pH, and back-trajectory calculations of the air masses (Deguillaume et al., 2014; Renard et al., 2020). The three ions Cl^-, Na^+ and Mg^{2+} are markers of marine sources, while NO_3^-, SO_4^{2-} and NH_4^+ ions are markers of continental sources.

1.4. Precipitation

Precipitation is the end-of-life process of clouds (complete or partial) that have become too heavy in water content. Water droplets condense around CCN and fall back to the Earth's surface. Depending on temperature, precipitation falls either in the form of rain, or in the form of snow or hail. When falling, rain washes out the air column and collects airborne particles (Bourcier et al., 2012). This complex phenomenon is called scavenging and involves different physical processes such as Brownian diffusion, inertial impaction, and interception. The efficiency of scavenging depends on aerosol particle size and composition, as well as on rain drop size and rainfall intensity (Hou et al., 2017; Mircea et al., 2000; Sonwani and Kulshrestha, 2019; Willis and Tattelman, 1989). Precipitation is therefore cloud water that has passed through the air layer (Figure 1). The chemical and biological composition of rain have already been studied and investigated (Aho et al., 2019; André et al., 2007). Meteorological conditions (wind, temperature, …) and local emission sources are known to influence rain composition. However, much remains to be understood on the importance of the contribution of scavenging to the composition of rainfall, especially for the biological part. Models and calculations have been proposed to estimate the scavenging efficiency based on chemical composition of precipitation or on the size distribution and number of aerosol particles and raindrops (Bertrand et al., 2008; Blanco-Alegre et al., 2018; Mircea et al., 2000; Moore et al., 2020a). However, these estimations are not based on environmental samples or only looks at precipitation and aerosols, but not the original cloud. The initial composition of the rain (when it falls from the cloud) can therefore only be extrapolated. Wet deposition in one of the main pathways for the redeposition of microorganisms on surface ecosystems (Figure 1) (Aho et al., 2019; Morris et al., 2004; Woo and Yamamoto, 2020), it is therefore of interest to better characterize which part comes from long-distance sources (from clouds) or from local sources (scavenging from the air column) to better understand the potential influence of precipitation on ecosystems (lake, vegetation, crops) and on Human health. Phytopathogens are, among other

microbes, present in precipitation that redeposit them on plants and crops. Some are even suspected of inducing precipitation due to IN activity like Pseudomonas species (Morris et al., 2004).

2. Microorganisms in the atmosphere

2.1. Biodiversity

Nowadays, the main microbial diversity in the outdoor atmosphere starts to be well documented. For the domain of Bacteria, the most abundant phylum is Proteobacteria, notably the orders Pseudomonadales (genera *Pseudomonas, Psychrobacter* and *Acinetobacter*), Burkholderiales (*Massilia* and *Janthinobacterium*), Rhizobiales (*Methylobacterium*), Rhodospirillales (*Acetobacter*) and Sphingomonadales (*Sphingomonas*). Other phyla are commonly found in the air such as Actinobacteria of the orders Corynebacteriales (*Corynebacterium*), Actinomycetales (*Streptomyces*) and Micrococcales (*Arthrobacter*); Firmicutes of the orders Bacillales (*Bacillus* and *Staphylococcus*), Clostridiales (*Clostridium*) and Lactobacillales (*Streptococcus*); and Bacteroidetes of the order Sphingobacteriales (Amato et al., 2017; Bowers et al., 2013; Ruiz-Gil et al., 2020). The domain of Eukaryotes, among unicellular organisms, is mainly represented by Fungi (Ascomycota and Basidiomycota, often spore forming), with also some representatives of Stramenopiles and Alveolata (Amato et al., 2017). In Basidiomycota, Agaromycetes, Tremellomycetes and Microbotryomycetes classes are the most represented, while in Ascomycota it is the Sordariomycetes, Eurotiomycetes, Leotiomycetes, Dothideomycetes and Saccharomycetes classes (Amato et al., 2017; Els et al., 2019; Fröhlich-Nowoisky et al., 2009, 2012). Archaea and viral sequences are also observed but not much is known about their global representation in the atmosphere, as most of the studies do not focus on them (Amato et al., 2019; Jaing et al., 2020; Reche et al., 2018; Smith et al., 2013).

Although these major bacterial and fungal groups are found almost everywhere in the atmosphere, probably due to their high capacity for atmospheric dispersal, airborne microbial communities remain very diverse and variable depending on: altitude (Bryan et al., 2019; Drautz-Moses et al., 2022; Jaing et al., 2020; Prass et al., 2021), season (Bowers et al., 2013; Cáliz et al., 2018; Tignat-Perrier et al., 2020), meteorological conditions (Bertolini et al., 2013; Maron et al., 2006) and sources of emission (Bowers et al., 2011; Li et al., 2020). Which leads us to the next sections about the sources of airborne microorganisms (section 0), atmospheric transport (section 2.3) and factors influencing the biodiversity and the biomass of airborne microbial assemblages (section 2.4).

2.2.Natural and anthropogenic sources

2.2.1. Bacteria

It is estimated that there are ~10^2 to ~10^6 bacterial cells by m^3 of air in the atmosphere (Bauer et al., 2002; Burrows et al., 2009b), depending on the altitude and the distance from the emission source. The sources are multiple and have natural or anthropogenic nature. Common bacterial groups associated with source types are summarized in **Figure 2**.

Concerning natural origins, vegetation (leaf surface, or phyllosphere) is one of the most important sources of atmospheric bacteria, along with soils (Lindemann et al., 1982; Samaké et al., 2020). The number of bacteria on plant leaves is estimated at 10^8 cells/g, with a global population of microorganisms on leaves of approximately ~10^{26} cells (Lindow and Brandl, 2003), which represents an important bacterial reservoir on Earth. Many phyllosphere-inhabitant bacteria are found in the atmosphere such as *Pseudomonas*, *Sphingomonas* or *Massilia* (Aydogan et al., 2018; Lindemann and Upper, 1985; Lindemann et al., 1982; Rastogi et al., 2013). Burkholderiales, Rhizobiales and Sphingomonadales orders were also found dominant in pine forest and rural areas in Bowers *et al.* (2011, 2013). A recent study, Manirajan *et al.* (2018), has revealed the existence of a diverse microbiome associated with pollen. The dominant bacterial phyla are Proteobacteria (*Pseudomonas* and *Rosenbergialla*), Firmicutes (*Bacillus* and *Lactococcus*) and Actinobacteria (*Curtobacterium* and *Friedmannialla*). This study correlates also bacterial taxa with insects (*Rosenbergiella*) and wind pollination (*Methylobacterium*). Members of the genus *Methylobacterium* are often found in air samples (Amato et al., 2017; Samaké et al., 2020; Tignat-perrier et al., 2019) and have a wide variety of characteristics that likely make them well fitted for survival in the atmosphere (desiccation tolerance, nitrogen-fixing activity, biofilm formation, facultative methylotrophy and pigmentation) (Ruiz-Gil et al., 2020).

Soils are also an important potential natural sources of airborne microorganisms with a number of prokaryotic cells estimated to 4 x 10^7 cells/g in forest soils and to 2 x 10^9 cells/g in other type of soils (including desert and cultivated soils)(Burrows et al., 2009b; Whitman et al., 1998). Common bacterial phyla associated with soil are Firmicutes, Proteobacteria (Rhizobiales) and Actinobacteria (Bowers et al., 2011; Després et al., 2012).

Marine environments are the largest sources of aerosolized microorganisms after soils and vegetation, given their surface (Després et al., 2012; Ruiz-Gil et al., 2020). Sea spray aerosols are formed by waves breaking and bubble bursting, enabling the atmospheric transport of microorganisms from the sea to other nearby and distant environments (Dueker et al., 2011, 2012; Evans et al., 2019; Michaud et al., 2018). The presence of several bacterial taxa with a high aerosolization capacity has been demonstrated in coastal Pacific seawaters. These bacterial phyla were Proteobacteria

(Pseudomonadales, Rhizobiales, Alteromonadales and Vibrionales), Actinobacteria (Micrococcales and Corynebacteriales), Bacteroidetes (Flavobacteriales ans Saprospirales), Firmicutes (Bacillales and Lactobacillales) and Cyanobacteria (Synechococcales) (Graham et al., 2018; Michaud et al., 2018).

Another significant source is desert dust which is widely studied because of its global impact on the atmosphere with dust plumes and potential consequences for Human health. Dust come from mostly arid regions of the North Africa (Sahara and Sahel), South Africa, Asia (Gobi Desert), Australia, and South America (Griffin, 2007). During dust events, the airborne bacterial concentration can increase by one order of magnitude (Cha et al., 2016; Maki et al., 2017). The phyla Proteobacteria (Sphingomonadales and Burkholderiales), Firmicutes (Bacillales), Actinobacteria (Micrococcales and Corynebacteriales) and Bacteroidetes (Bacteroidales and Flavobacteriales) are the most frequent bacteria found associated with desert dust (Barberán et al., 2015; Griffin, 2007; Ruiz-Gil et al., 2020).

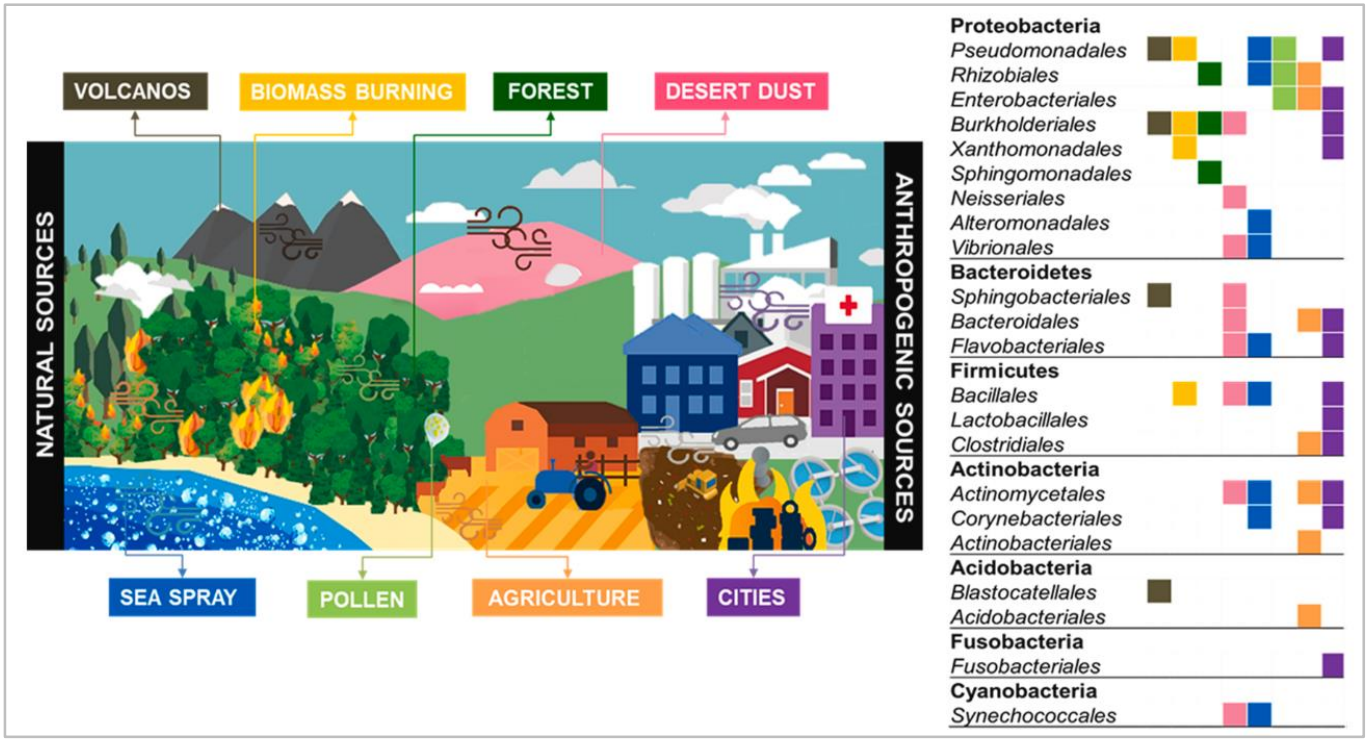


Figure 2: Schematic representation of the main bacterial sources with the most associated bacterial orders (from Ruiz-Gil et al. 2020).

Among anthropogenic sources, first there are aerosol emissions from urban activities (e.g, hospitals, houses, pet feces, construction, and transportation). Dominant airborne bacterial phyla associated with urban and suburban areas are Actinobacteria (Actinomycetales and Corynebacteriales), Firmicutes (Bacillales, Lactobacillales and mostly Clostridiales), Proteobacteria (Pseudomonadales, Enterobacteriales, Burkholderiales and Xanthomonadales) and Bacteroidetes (Bacteroidales and Flavobacteriales) (Bowers et al., 2011, 2013). Some of those bacteria are known to be potential human pathogens (genera *Bacteroides*, *Burkholderia*, *Enterococcus*, *Staphylococcus*,

Corynebacterium, *Streptococcus* and *Vibrio*) and their occurrence significantly increased with urbanization development for example in China, especially in samples from hospital areas (Li et al., 2019).

Second, agriculture activities (livestock and wastewater) can be sources of bioaerosols. Rural areas are dominated by the same bacterial phyla as in urban areas but in different proportions, with mostly the orders Rhizobiales, Enterobacteriales, Bacteroidales and Clostridiales (Bowers et al., 2011, 2013).

Third, waste treatment facilities (e.g., compost, landfill, and wastewater) can also generate bioaerosols and contain potential human and plant pathogens (Yang et al., 2018).

Finally, there are also wildfire and biomass burning (anthropogenic and natural source) that can significantly contribute to microbial inputs in the atmosphere. About 78% of the microorganisms in smoke are inferred to be viable and can increase by a four-fold cell atmospheric concentrations (Kobziar et al., 2022; Moore et al., 2020b). The main bacterial phylum found both in smoke and ambient air was Actinobacteria; and the phyla Bacteroidetes, Chloroflexi, Planctomycetes, Acidobacteria and Deltaproteobacteria were found more abundant in smoke than in ambient air. In Firmicutes, the two orders Bacillales and Clostridiales were more present in smoke (Kobziar et al., 2022).

2.2.2. Fungi

Concerning Fungi, concentrations in the outdoor atmosphere are estimated between 10^1 and 10^6 cells/m^3 of air (Bauer et al., 2002; Elbert et al., 2007; Tignat-perrier et al., 2019) depending again on the altitude and the proximity with the emission source. It is also estimated that fungal spores contribute to ~23% of the total PBAP in the atmosphere (Heald and Spracklen, 2009).

The main natural and anthropogenic sources are the same as for bacteria. Basidiomycota are dominant in all environments and are especially highly represented in continental areas, while Ascomycota are proportionally more represented in marine air (**Figure 3**). Fungi are overall less represented in coastal regions, which correlates with the fact that ocean is not a major source for fungal spores (Elbert et al., 2007; Fröhlich-Nowoisky et al., 2012). The genus *Clapospor*ium (Ascomycota) has been reported as highly dust-associated and is one of the most common allergens. Several members of this genus are also known as major plant pathogens (Barberán et al., 2015). More recently, *Cladosporium* has been reported as part of the core microbiome of pollens (Manirajan et al., 2018), with also the genus *Aureobasidium* (Ascomycota).

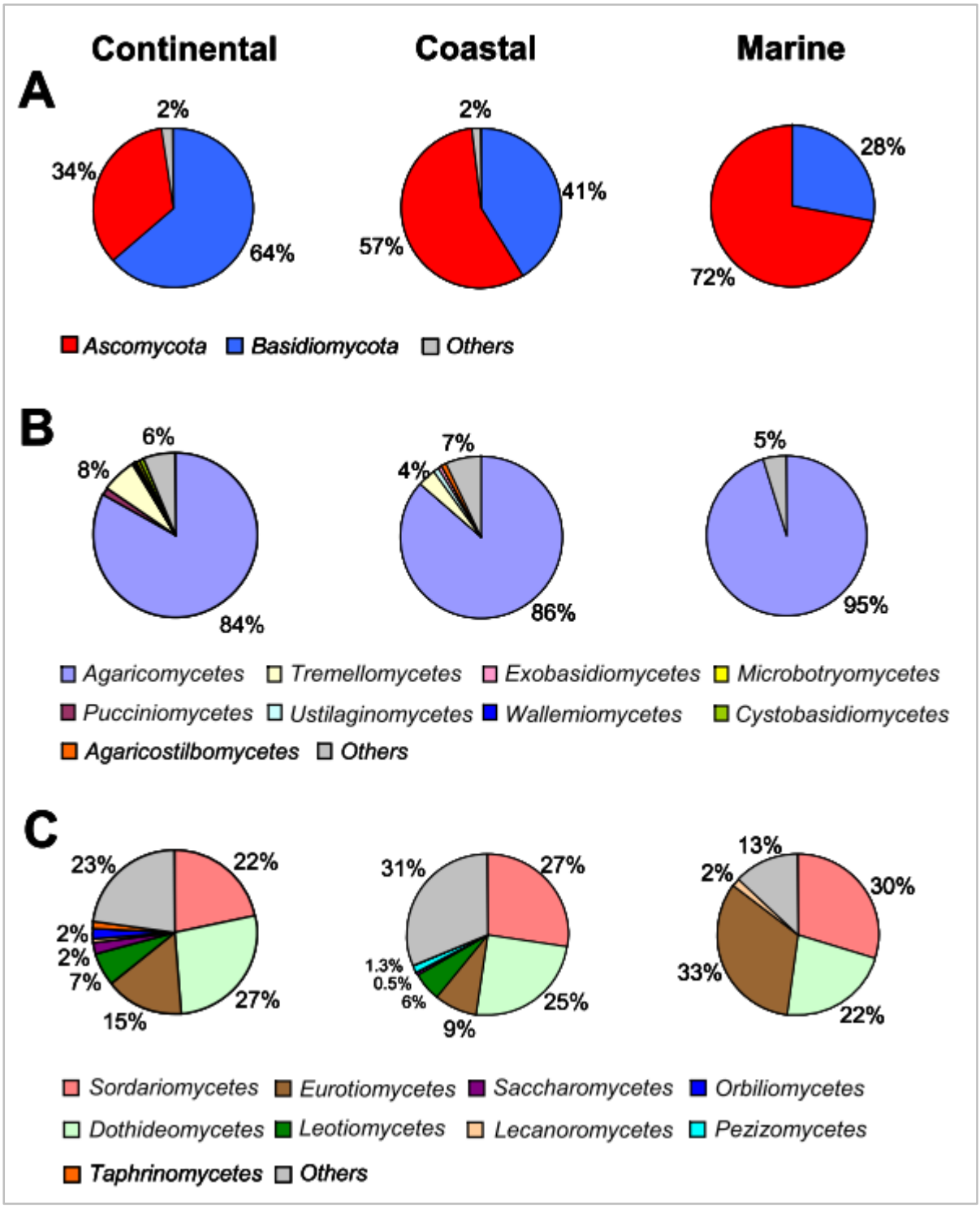


Figure 3: Species richness of airborne fungi; mean relative proportions of different phyla (A), different classes of Basidiomycota (B), and different classes of Ascomycota (C) in continental (Austria, Arizona, Brazil, Germany), coastal (China, Taiwan, Puerto Rico, UK), and marine (Pacific, Indian, Atlantic, Southern Ocean) air samples (Fröhlich-Nowoisky et al., 2012).

2.3.Short and long-distance transport

It is very important to better understand what drives the transport of microorganisms in the atmosphere because it allows them to spread over long distance and to reach a wide variety of ecosystems. This can have a significant role on the dynamics of local ecosystems with a contribution of new non-endemic species and an input of new genetic material (Fröhlich-Nowoisky et al., 2012; Womack et al., 2010).

Microorganisms are part of the small PBAPs (~1 µm for bacteria and ~1.5 to 20 µm for fungal spores; Elbert *et al.* 2007; Després *et al.* 2012) and are thus susceptible to be transported over long time and distances in the atmosphere. According to Burrows *et al.* (2009a), the particles of about 1 µm in diameter comprise aerosol with the longest atmospheric residence times. Indeed, the Greenfield gap is an interval gap in scavenging efficiency by rain for particles between 0.2 and 1 µm (Blanco-Alegre et al., 2018; Ladino et al., 2011; Radke et al., 1980), which contributes to decrease the wash-out efficiency for these particles. The atmospheric residence time for the other particles depends on their aggregation capacity for the smallest (0.06-0.2 µm) and on sedimentation for the larger ones (>1 µm). Burrows *et al.* (2009a) estimate a mean global atmospheric residence time for bacteria of 3.4 days for the CCN-active (bacteria active as cloud condensation nuclei; Bauer *et al.* 2003) and of 7.5 days for the CCN-inactive. Since most bacteria are considered CCN-active, their average residence time in the atmosphere is therefore closer to a few days (3.4 days) than a week. For fungi, it is estimated that the biggest fungal spores have a mean atmospheric residence time of 1 day (Elbert et al., 2007), and this is probably similar as for bacteria concerning the smallest spores. It is also important to notice that microorganisms can be in suspension in the atmosphere as single cells or aggregated with other microbial cells, bigger organic fragments (plant or insect fragments) and/or inorganic particles (Després et al., 2012). The average residence time of microorganisms in the atmosphere can therefore be reduced depending on the size of the aggregate.

In the lower troposphere (ABL), where atmospheric biomass is the highest (Els et al., 2019), the atmospheric transport is mainly determined by short-scale turbulent vertical air motions (**Figure 1**) (Fröhlich-Nowoisky et al., 2016). On the contrary, above the ABL in the free troposphere, air masses are driven by horizontal movements at planetary scale. Main air movements on Earth are determined by Hadley cells and the Earth's rotation (**Figure 4**). As bacteria have been found in the upper troposphere up to the stratosphere (DeLeon-Rodriguez et al., 2013; Jaing et al., 2020; Smith et al., 2018; Triadó-Margarit et al., 2019), they can thus integrate clouds (Amato et al., 2007; Sattler et al., 2001) and air masses of high altitude, and be transported at regional to continental scales (Barberán et al., 2015; Smith et al., 2013).

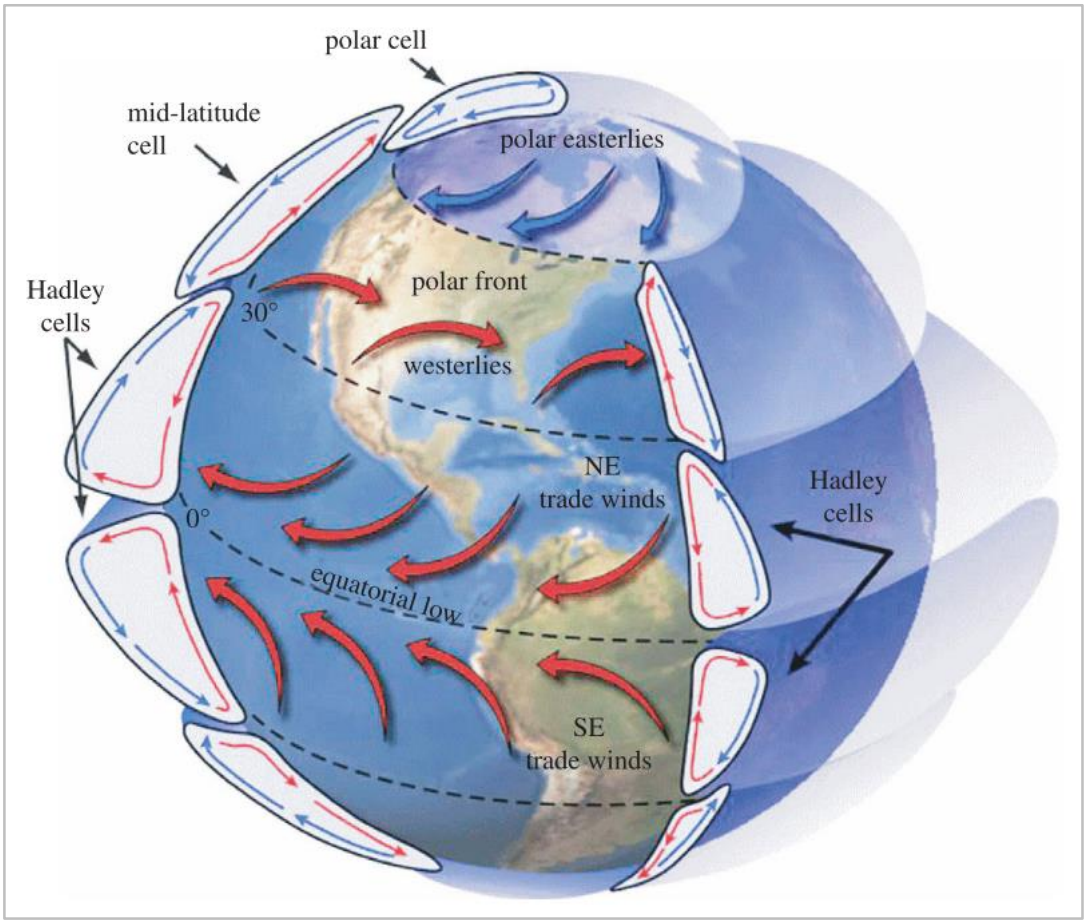


Figure 4: The six major air cells of the Earth's atmosphere (sources: Womack et al., 2010).

2.4. Factors influencing the biomass and diversity of microbial communities in the atmosphere

Atmospheric microbial communities are very variable, mostly because of the multiple mixed natural and anthropogenic sources that compose them, but not only. Seasonality, diurnal cycles, and meteorological conditions also influence their composition.

One of the main aspects is the fact that meteorological conditions and seasons will affect the contribution of natural and anthropogenic sources for the airborne microorganisms at a set location (Bowers et al., 2013). Indeed, we saw in section 0 that the two main sources of aerosolized bacteria and fungi were vegetation and soils. Seasonal vegetation provides different contributions to airborne communities. For example, airborne microbial communities in puy de Dôme area (France) comprised more plant-associated taxa during summer than winter, while in winter soil and dead material-associated microbes were dominant (Tignat-Perrier et al., 2020). Cáliz *et al.* (2018), also found a significant correlation airborne microbial communities and seasonal changes in Central Pyrenees Mountains (South-West Europe), but in this case this also corresponded to change in remote sources.

There were for example, more air masses originating from the Atlantic in winter and more Saharan dust intrusions during summer, which led to different microbial contributions.

Meteorological conditions can also influence atmospheric microbial communities within season, like air temperature, relative humidity, wind speed and precipitation. Air temperature can have a significant impact on both inter- and intra-seasonal variability and was correlated positively with bacterial biomass (Bowers et al., 2013; Gandolfi et al., 2013). Relative humidity and rain were negatively correlated with microbial biomass since moisture intensifies deposition by increasing particles size, and wet soil surfaces make aerosolization difficult (Gandolfi et al., 2015; Smets et al., 2016). Rain also washes-out microorganisms from the air and potentially preferentially specific taxa, depending on particle diameter (Moore et al., 2020a; Woo and Yamamoto, 2020). However, heavy rain can also generate bioaerosols when reaching soil surface, increasing the airborne microbial concentration (Huffman et al., 2013; Joung et al., 2017). Finally, wind speed has also been associated positively with concentration and diversity of airborne microorganisms. Indeed, wind can be an important enhancing factor of bioaerosol generation by sweeping soils and water surface, particularly contributing to the formation of sea spray (Bowers et al., 2013; Gandolfi et al., 2015; Graham et al., 2018; Ruiz-Gil et al., 2020).

Another aspect is the selection pressure on microorganisms for adaptation due to stress factors (e.g., oxidants, solar radiation, temperature) in the atmosphere, which lead to different microbial communities with diverse acclimatization (e.g. spore formation, pigmentation) (Smets et al., 2016). This part will be discussed with more details in section 3. Selection pressure can also be caused by other ecological factors, such as availability of certain substrate such as acetate, ethanol or formate (Amato et al., 2005; Krumins et al., 2014; Šantl-Temkiv et al., 2022).

3. Clouds as microbial habitats

A habitat in ecology is defined as an assembly of organisms together in interaction with their abiotic environment. It can also be defined as a place where all the environmental conditions an organism needs to survive are met: shelter, water, nutrient and space.

3.1.Stress factors and survival

Microorganisms in the atmosphere are exposed to stressful conditions like UV radiation, desiccation, temperature and chemical shifts, and the presence of reactive oxygen species (hydrogen peroxide)(Smets et al., 2016). In a cloudy environment, microorganisms are protected from desiccation by the presence of condensed water but must deal with additional stressors such as osmotic shock, freeze-thaw cycle, and chemical composition of the droplets (**Figure 5**) (Joly et al., 2015), as well as

with potentially higher exposure to light due to Mie scattering. It is a highly dynamic environment where temperatures and liquid water content can change within minutes (Šantl-Temkiv et al., 2022). Despite, these harsh conditions, some microorganisms remain viable and active (Amato et al., 2005; Joly et al., 2015; Sattler et al., 2001). Indeed, the atmosphere and clouds are not the most extreme environment on Earth and microorganisms. Bacteria and Achaea, particularly, have been found living in far more extreme environments (Rothschild and Mancinelli, 2001; Womack et al., 2010, and references below) such as acidophilic archaea that can growth at pH near 0 (Edwards et al., 2000; Schleper et al., 1995) or alkaliphilic microbes that can growth at pH ~11 (Jones et al., 1998); there is also psychrophilic bacteria living at temperature near 0 or below (Morita, 1975).

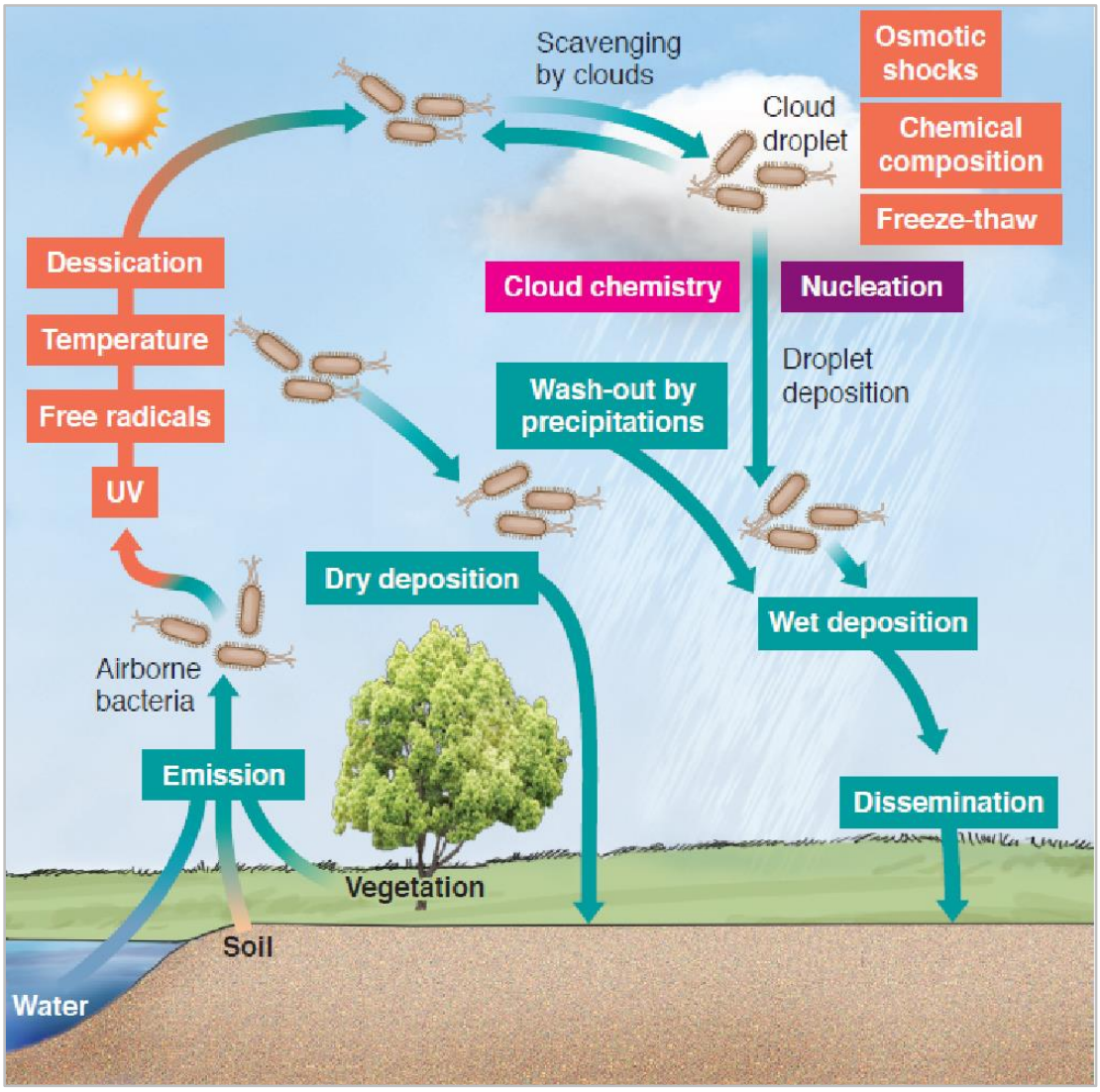


Figure 5: The microbial atmospheric cycle (Amato 2012, Clouds provide atmospheric oasis for microbes. Microbe Magazine 7, 3, 119-123).

Some adaptations to stress factors in clouds have been studied or hypothesized. First, it was suggested that bacterial cells can survive better when attached to substrates, by mitigating level of UV exposure when embedded within larger particles such as dust, pollen or water droplets (Lighthart, 1997; Lighthart and Shaffer, 1995, 1997). The better survival may also be due to the presence of substrate and nutrients available to microbial cells that allow them to maintain metabolic activity. Second, certain bacteria such as some Firmicutes have the capacity to enter a state of dormancy by forming resistance spores or undergo other cell wall modifications and slow down or stop their metabolic activities (Delort et al., 2010; Smets et al., 2016). It can improve their resistance to physical stresses, and thus favor their survival in the atmosphere (Romano et al., 2019). The spores of Fungi are reproductive and dissemination structures are also well adapted to harsh conditions, and these can survive in the atmosphere and in clouds (Després et al., 2012; Elbert et al., 2007). Third, phenotypic traits such as pigmentation can serve as protection mechanisms against UV radiation, free radicals, low osmotic pressure, and low temperature effects (Delort et al., 2010; Fong et al., 2001; Mueller et al., 2005; Sajjad et al., 2020). Many pigmented bacteria were observed among the viable cultivable organisms in cloud water (Amato et al., 2005), representing up to 60% of the cultivable fraction. Some pigmented fungi were also detected in the high atmosphere and in clouds (Imshenetsky et al., 1978; Joly et al., 2015).

3.2.Microbial activity

We know from decades that microorganisms are prone to survive in the atmosphere and in cloud-like environments (Fuzzi et al., 1996), and that they may be active in clouds (Amato et al., 2005, 2017; Sattler et al., 2001). However, little is still known about the active microorganisms in clouds and the functions expressed there. In this section, some of the few studies focusing on bacterial and fungal metabolic activity in the atmosphere and in clouds will be presented. We will focus first on the most recent laboratory experiments and then on global approaches that have been used to study active microbial communities.

In Vaïtilingom *et al.* (2010), laboratory experiments involving five microbial strains isolated from clouds (3 *Pseudomonas* strains, 1 *Sphingomonas* strain, and 1 *Dioszegia* strain) were carried out to determine whether these strains could biodegrade some abundant atmospheric carbon compounds as formate, acetate, lactate, or succinate. It appears that, at low temperatures representatives of low altitude clouds (5°C and 17°C) in a liquid solution that mimicked cloud water composition, microbial strains could significantly participate to the degradation of formate, acetate, and succinate. This is not surprising given that microorganisms are known to use these as substrates, and be possibly active at low temperatures at or below 0°C (Amato, 2013; Anesio et al., 2009). Other laboratory experiments involving natural cloud water and its endogenous microbial community in a custom bioreactor

(Vaïtilingom et al., 2013) have demonstrated that microorganisms remain active in the presence of UV light and •OH radicals, and affect the concentrations of H_2O_2 and major carbon compounds such as formaldehyde and carboxylic acids. These compounds such as formaldehyde or acetate, formate and phenol are interesting for studying the impact of biological processes in clouds, as it has been estimated in Khaled *et al*. (2021) that compounds with an intermediate solubility, and therefore present in the gaseous and aqueous phases of clouds, are better degraded by microorganisms. It has also been demonstrated that H_2O_2 modulates the energy metabolism of cloud microbiota (strong correlation between ATP and H_2O_2 concentrations), and thus, impacts the cloud chemistry, especially the biotransformation rates of carbon compound (Wirgot et al., 2017). This can consequently change the interactions between the cloud system and the global atmospheric chemistry. Substrate dependence was also studied for aerosolized bacteria. Krumins *et al.* (2014) used the bacterial strain *Sphingomonas aerolata*, isolated from aerosols, to measure its activity when aerosolized. Cells were incubated in rotating gas phase bioreactors with or without the presence of volatile carbon substrates (acetic acid and ethanol). According to RNA:DNA content ratios, the airborne bacteria exhibited significantly higher activity in the presence of substrates, indicating that bacteria (at least some of them) can be active and metabolize substrate even when aerosolized. Thus, if bacteria are active in aerosols, it is very likely that they are also active in an aqueous environment such as cloud water.

Methanotrophic airborne bacteria were investigated in Šantl-Temkiv *et al.* (2013), the oxidation of methane was measured in enriched air and rainwater samples and in cloud-like cultures. It reveals that methanotrophic bacteria are viable and active in both the dry and wet phases of the atmosphere. Moreover, airborne methanotrophs were able to oxidize methane at atmospheric concentration, even at low pH as can be found in cloud droplets and appear to be more competitive in environments with low nutrient concentrations and low biomass. Therefore, it is proposed in this study that cloud droplets provide a suitable environmental niche for their activity and growth in the atmosphere.

However, these described microbial activities were studied under control laboratory conditions and not *in situ.* This therefore does not represent all the environmental pressures and the real state of microbial communities in the atmosphere. Other studies have explored the global microbial activity in real *in situ* samples using RNA as an activity marker.

In Klein *et al.* (2016) active bacteria in the atmosphere were investigated in aerosols at high altitude (Mt. Bachelor, 2763 m a.s.l.; OR, USA) using rRNA and rRNA gene (rDNA) sequencing. The bacterial order Rhodospirillales (Proteobacteria) was the most potentially active, with also the orders Actinomycetales (Actinobacteria), Saprospirales and Cytophagales (Bacteroidetes). Rare taxa in the whole community (rDNA) were found much more potentially active (given rRNA:rDNA content ratio)

compared to abundant taxa (**Figure 6**). This observation was also done in other environments such as marine (Campbell et al., 2011; Hunt et al., 2013) and freshwater (Wilhelm et al., 2014) systems.

Airborne bacterial activity in the Artic was investigated in Šantl-Temkiv *et al.* (2018) using total 16S rRNA copy number. A high activity potential in aerosols was found for Rubrobacteridae, Cyanobacteria and Clostridiales. The subclass Rubrobacteridae (Actinobacteria phylum) is known for containing many desiccation resistant species (Barnard et al., 2013). On the contrary, a low activity potential was observed for Proteobacteria (given cDNA:DNA ratios). Nevertheless, the genus *Pseudomonas*, although not present in the active fraction of the bacterial community of the air samples, was enriched in the active communities of the rain samples in Šantl-Temkiv *et al.* (2018), suggesting a potential for high activity in cloud water as found in Amato *et al.* (2017).

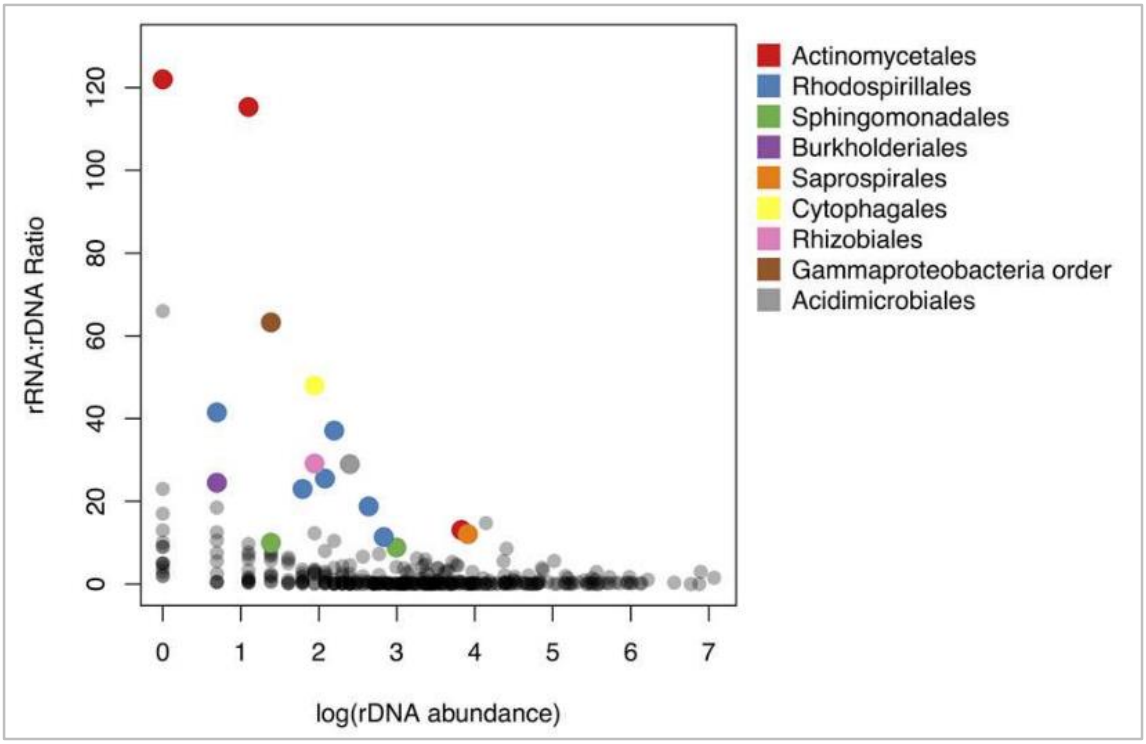


Figure 6: Relationship between rRNA:rDNA content ratio and abundance in the rDNA community. rRNA:rDNA content ratio is considered the potential activity rate. Colored points represent taxa significantly overrepresented in the active airborne community. Points are colored by taxonomic order. (From Klein et al., 2016).

Concerning fungi, Womack et al. (2015) investigated both total and active fungal community in aerosols above the Amazon rainforest, using total DNA and RNA sequencing (metagenomics and metatranscriptomics). Basidiomycota were found dominant in the whole fungal community but, the active part of the community was mainly represented by Ascomycota (Figure 7). Active Basidiomycota were mainly represented by the class Agaricomycetes. For Ascomycota it was the classes Sordariomycetes, Lecanoromycetes and Saccharomycetes. Sordariomycetes are known to be

endophytes, pathogens, and saprotrophs (Xylariales) (Zhang et al., 2007). Also, several genera with known IN capability were detected such as Agaricus, Amanita, Aspergillus, Boletus, Lepsita, Mortierella, Puccinia, Rhizopus and the lichen fungus Cladonia. The class Lecanoromycetes is mainly represented by lichen, and lichen fungi are also known to have an efficient IN activity (Kieft and Ahmadjian, 1989). Among others, species Physcia stellaris and Rinodina milvina were detected. These results with Ascomycota as the main active fungi are in agreement with the fact that Ascomycota have single-celled or filamentous vegetative growth forms that will be easily aerosolized given their small size, while Basidiomycota are larger and are aerosolized mainly under the form of metabolically inactive spores (Moore et al., 2011; Womack et al., 2015).

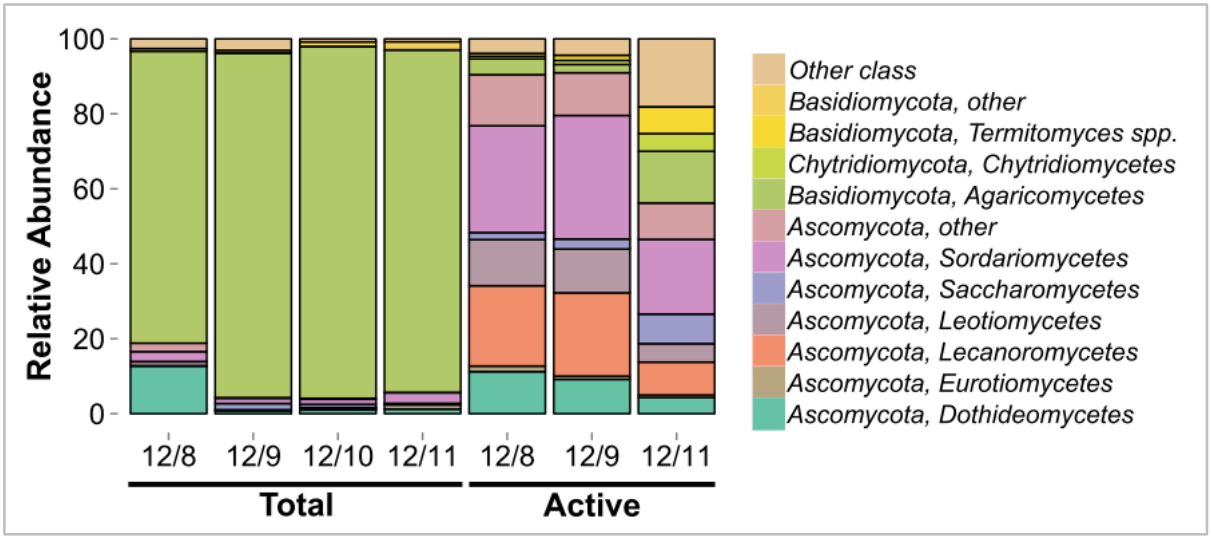


Figure 7: Airborne total and active fungal relative abundances above the Amazone rainforest. Bars are colored according to class-level taxonomic assignements. (From Womack *et al.* 2015).

The studies presented above, among a few others, focused on active bacterial and fungal communities in aerosols, and there are even fewer studies investigating global microbial activity in clouds. We will now focus on the only two studies examinating cloud's active microbial communities using total DNA and RNA sequencing.

First, in Amato *et al.* (2017) microbial communities in cloud water were fixed upon sampling and investigated by high-throughput sequencing of total DNA and RNA. The active part of the microbial community represented approximatly 26% of the richness observed for prokaryotes and 82% for eukaryotes. Regarding RNA:DNA ratios for bacteria, Alpha- and Gamma-Proteobacteria were clearly dominant in the active fraction (ratio > 1; i.e. potentially metabolically active taxa) with mainly genera associated to the phylosphere such as *Enhydrobacter*, *Acidiphilium*, *Sphingomonas*, *Pseudomonas* and *Methylobacterium* (**Figure 8-A**). There were also the bacterial phyla Deinococcus-Thermus (*Spirosoma* and *Deinococcus*), Actinobacteria (*Frigoribacterium* and *Curtobacterium*) and the sub-phylum Beta-Proteobacteria (*Janthinobacterium*).

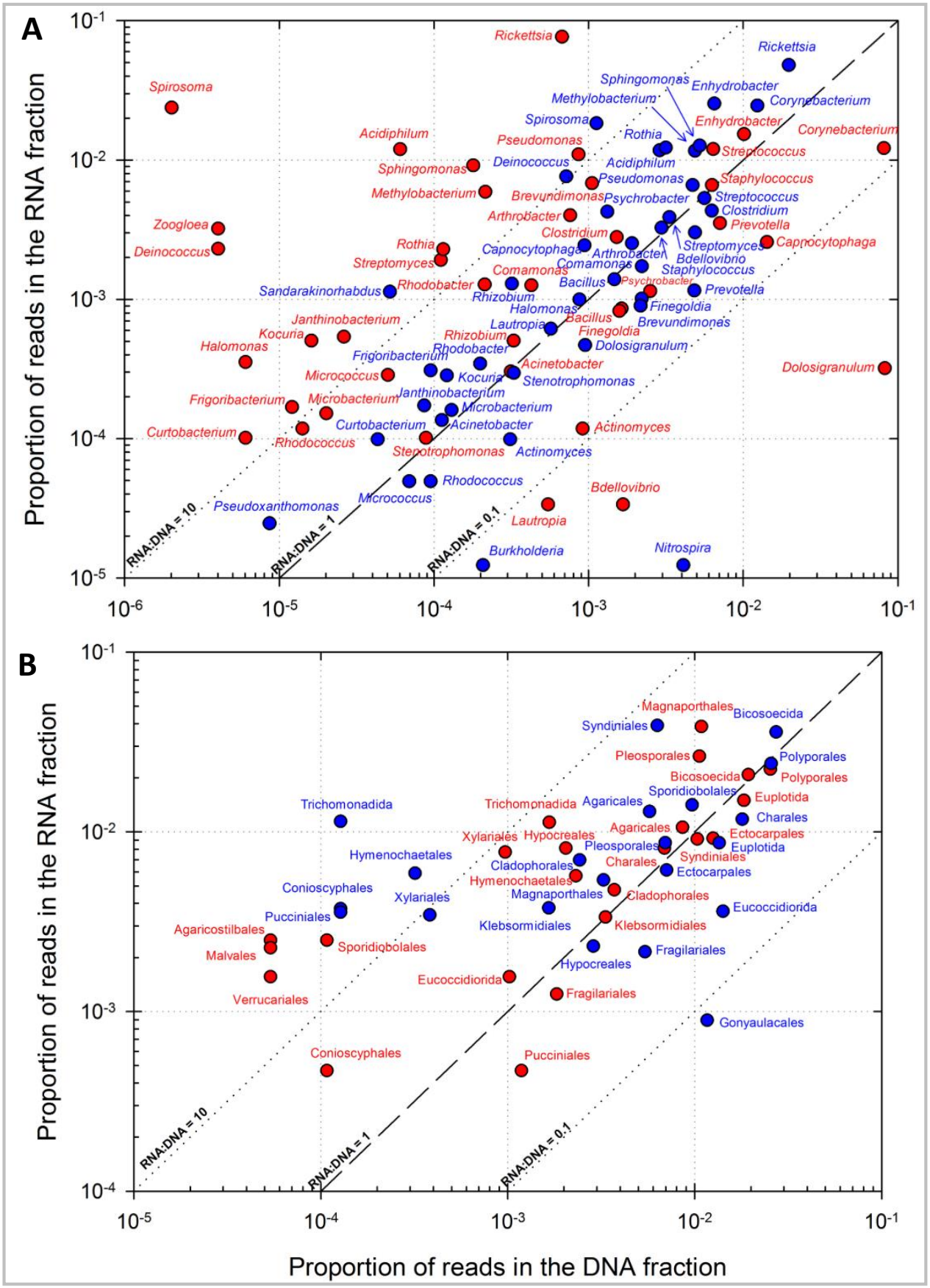


Figure 8: Representation of the major prokaryotic genera (A) and eukaryotic orders (B) in DNA and RNA datasets from clouds. Dashed and dotted lines depict RNA:DNA ratios of 0.1, 1 and 10. The top 20 genera/orders based on their average position rank over 3 cloud samples are shown, as well as some selected for high representation in RNA datasets (43/24 distinct genera/orders in total). Red dot: "Polluted" type cloud; Blue dot: "Continental" type cloud. Adapted from Amato *et al.* (2017)

For eukaryotes, fungi were the most active with the orders Magnaporthales (Sordariomycetes) and Pleosporales (Dothideomycetes) in Ascomycota and Polyporales (Agaricomycetes) and Sporidiobolales (Microbotryomycetes) in Basidiomycota (**Figure 8-B**). There were also some active Stramenopiles and Alveolata. In Amato et al., (2017), Ascomycota were not found more active than Basidiomycota in clouds, contrary to what was found in aerosols in Womack *et al.* (2015). This can be because of the different atmospheric compartment studied (cloud and aerosol) or because of differences in local influences (Amazon Forest vs Central France). Also, quantification were not absolutes but relatives.

We have described the main bacterial and fungal taxa active in aerosols and clouds, but still, nothing about their functional profile. Some functions hypotheses have been made in the studies previously described with regard to the active taxa present, but there are no data on this subject. The first global study of active microbial functions and metabolic pathways in clouds was in Amato et al. (2019) with metatranscriptomics and metagenomics. One of the main points in this publication was that, while eukaryotes (mainly fungi) were highly dominant in metagenomes (MG; so in the total community), prokaryotes (mostly bacteria) had a much higher relative contribution in metatranscriptomes compared to their contribution to MG (**Figure 9**). It suggests that the active microbiota of clouds is mostly composed of bacteria.

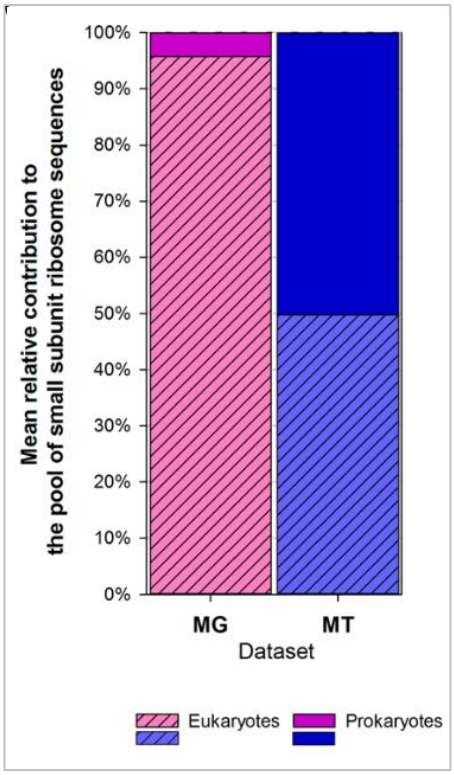


Figure 9: Mean relative contribution of eukaryotic (dashed) and prokaryotic (clear) taxa to the pool of identified SSU rRNA gene sequences in cloud's metagenomes (MG, pink) and metatransciptomes (MT, blue). Adapted from Amato *et al.* (2019).

Many biological functions reflecting a challenging environment were expressed based on metatranscriptomes: processes involved in the maintenance of homeostasis and in the response to oxidative stress, as well as free radical and oxidant (hydrogen peroxide) detoxification processes (**Figure 10-A**). These imply enzymes such as catalase (GO:0004096; GO = gene ontology), superoxide dismutase (GO:0004784), peroxiredoxin (GO:0051920) and peroxidase (GO:0004601), and the production of antioxidant compound (glutathione; GO:0006750). These supported previous observations by Vaïtilingom *et al.* (2013) and Wirgot *et al.* (2017) about the impact of biological activity toward H_20_2 and the modulations of energy metabolism in response to it in the cloud microbiome. ATP biosynthesis and metabolic processes were also highly expressed. Another interesting point is the presence of mRNAs related with siderophore synthesis and transport processes. Iron is known to have an impact on the concentrations and cycling of free radicals and oxidants such as H_2O_2. Siderophores are also important for microorganisms as iron is cofactor of many electron transfers processes and metalloenzymes. *Pseudomonas* species appeared to be the dominant siderophore-producers in clouds (Vinatier et al., 2016).

Still based on these metatranscriptomes, translation and transcriptional activities were maintained despite temperatures between 0 and 1°C during sampling, with, among others, the synthesis of glycine (GO:0006545) and glutamate (GO:0006537), both constituents of gluthathione, the main intracellular redox regulator. Some of the biological processes observed can potentially participate to cold acclimation in bacteria, such as glycine metabolism, lipid metabolism and transport, glycerol ether, steroid, phospholipid and unsaturated fatty acid metabolisms, and branched-chain fatty acids biosynthesis.

Finally, genes related with central metabolic pathways such as the tricarboxylic acids (TCA) and pentose phosphate cycles were expressed. The pentose phosphate shunt is known to be a major pathway involved in the regulation of cell redox homeostasis. Also, other carbon metabolic pathways such as glucose metabolic processes, polysaccharide synthesis and processes involving one-carbon compounds (GO:0006730) were overexpressed (**Figure 10-B**). All these processes attest of the activity of bacteria not just to survive, but perhaps also to thrive.

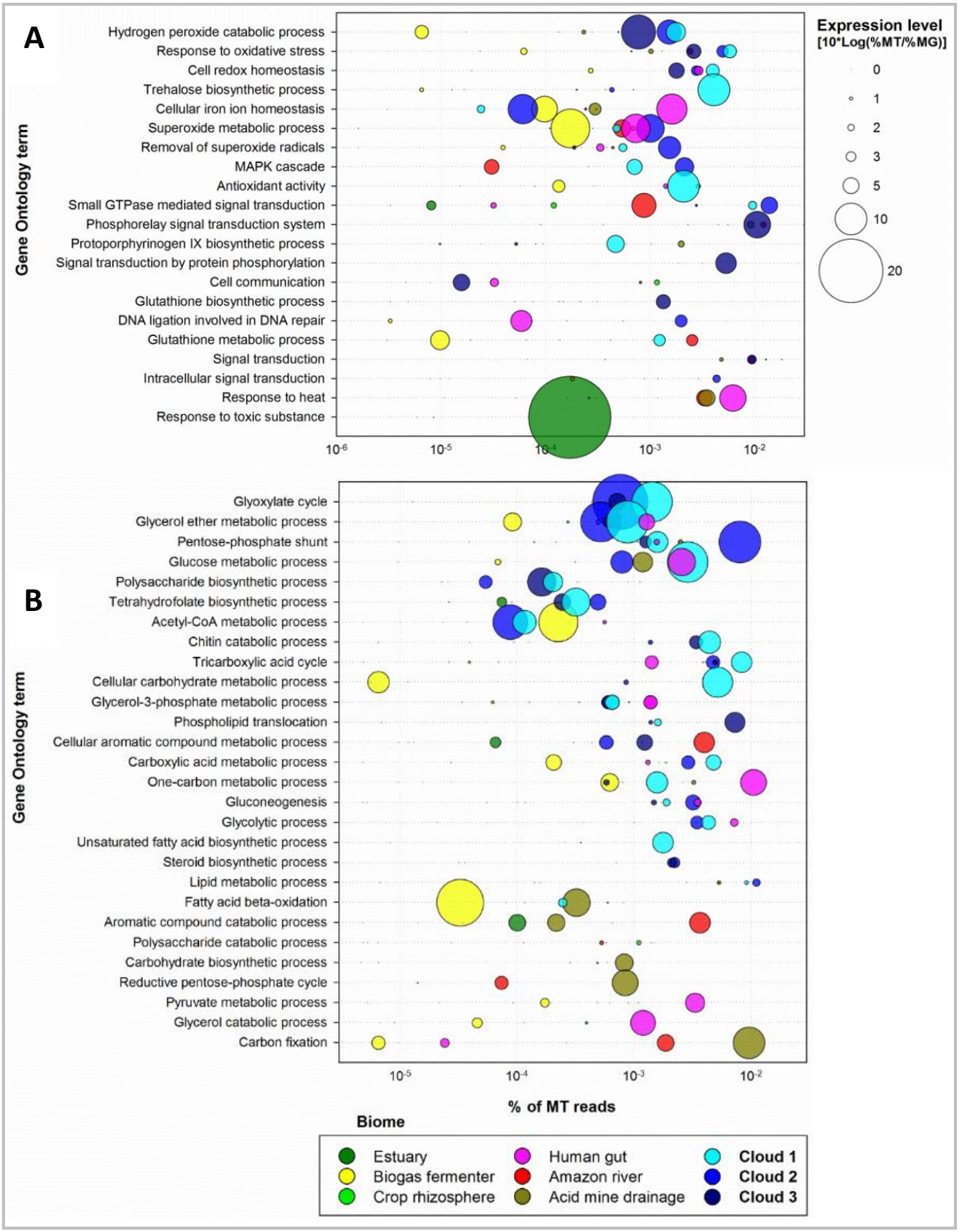


Figure 10: Biological processes related with stress response and signaling (A) and carbon metabolism (B) in cloud communities, compared with other environments, ordered by their summed expression level in clouds. The expression level (bubble size) depicts the relative importance of the corresponding GO term in the metatranscriptome (MT) dataset(s) after normalization to metagenome (MG); as expressed this is equal to 0 for similar representation in MT as in MG, and >0 for greater representation in MT, so only overexpressed functions are visualized. Adapted from Amato *et al.* (2019).

4. Problematics and challenges

4.1.Context and objectives

The outdoor atmosphere is a dynamic environment where multiple local and distant sources of airborne microorganism mix over short and long distances (**Figure 1** and **2**). Clouds and aerosols are likely to harbor variable, yet similar, microbial diversity as cloud droplets condense around aerosols. Microbial activity has also been reported in both atmospheric situations. However, unlike the dry atmosphere, clouds provide an aqueous environment for microbial cells, perhaps with protection from direct sunlight and better access to nutrients, potentially promoting the resumption of cell activity.

The problematics of this thesis work was therefore to investigate **if clouds could be specific biological entities in the atmosphere**. In this purpose, a comparative study of clouds with respect to other atmospheric compartments (i.e., precipitation and dry aerosols) was carried out in terms of microbial diversity and functional profile.

4.2.Challenges

Many challenges are associated with aeromicrobiology. First of all, cloud sampling is not an easy thing: a specific site at an altitude high enough to be embedded in clouds is needed. In addition, the quantities of sample collected are limited by meteorological conditions for clouds and precipitation. Second, the low biomass in the atmosphere (~10^4 cells m^{-3} of air), particularly at high altitude, makes it difficult to sample sufficient biomass in a short time (< 24 h to avoid mixing of air masses) to perform nucleic acid-based analyses. High-volume samplers are therefore needed. However, a High-flow rate can impact all the more cells viability and integrity when collected. This bring us to the third point, which is the preservation of cells integrity and of the *in situ* state of the sample during collection. These aspects are key elements to study the activity of microorganisms. The use of sampling by impingement (impaction on a liquid interface) allows collection in physiological liquid or a fixative, thus preserving the in situ state of airborne microorganisms (Šantl-Temkiv et al., 2020). As a last point on aeromicrobiology, the huge diversity in the atmosphere and the high variability between sampling event make comparative analyses more difficult. To circumvent this problem, it is necessary to collect multiple events to observe a trend, and ideally to do sampling replicates for each collection event.

Nucleic acid sequencing data analysis also has its share of challenges. Multiple quantification biases and amplification of contaminants (or chimera) are associated to NGS sequencing. These issues can however be monitored and restricted with good sampling and sample processing practice, the use of positive and negative controls and an appropriate bioinformatics processing (de Goffau et al., 2018). Finally, a last challenge is the lack of joint bioinformatics workflow for the analysis of more complex dataset from omics-sequencing studies (such as metatranscriptomics).

Chapter 2: Experimental procedures and challenges

1. Sampling sites

Sampling was performed using the instrumented atmospheric research stations Cézeaux-Aulnat-Opme-Puy de Dôme (CO-PDD) (Baray et al., 2020) (**Figure 1**). CO-PDD is internationally recognized as a global station in the GAW (Global Atmosphere Watch) network. Puy de Dôme (PUY) and Opme (OPM) stations were investigated in this thesis work.

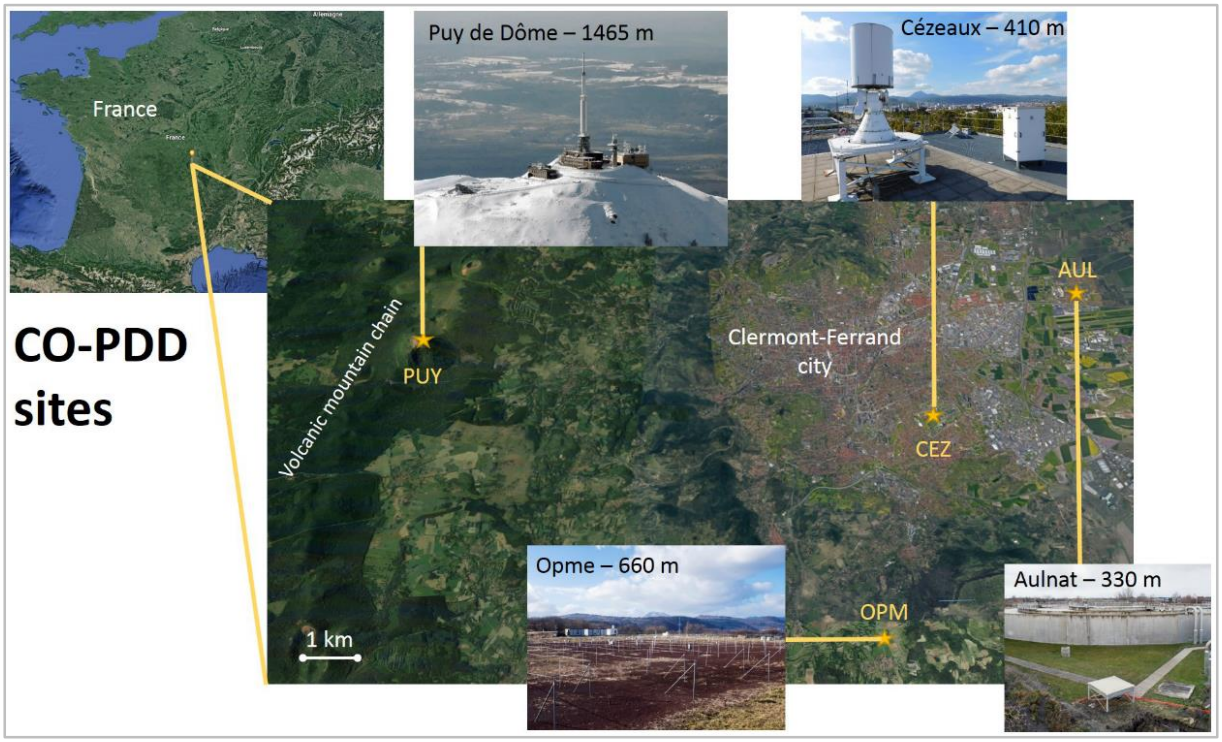


Figure 1 : Maps and photos of the CO-PDD sites (Baray et al., 2020). PUY and OPM stations were investigated in this study.

PUY station was used for cloud and aerosol sampling. This is a unique site in France for cloud sampling due to its particular topography and all the facilities of the station. PUY culminates at 1,465m altitude above sea level (a.s.l.) and is therefore in or out of the ABL depending on the season: PUY is in the free troposphere 50 % of the time in winter, and mainly in the ABL in summer (Baray et al., 2020). The possibility to sample in the free troposphere is of great interest to observe regional-scale influences due to long-range source transport with horizontal air movements. The station is on average embedded in clouds 30 % of the time and up to 60 % in winter (Baray et al., 2019). Given the very difficult and sometimes dangerous meteorological conditions at PUY summit in winter (freezing temperatures and strong wind), the two best seasons for cloud collection are therefore fall and spring (positives temperatures and low BLH) (**Figure 2**). It is possible to handle the samples under sterile conditions and to carry out biological analyzes at the sampling site within the fully equipped PUY station.

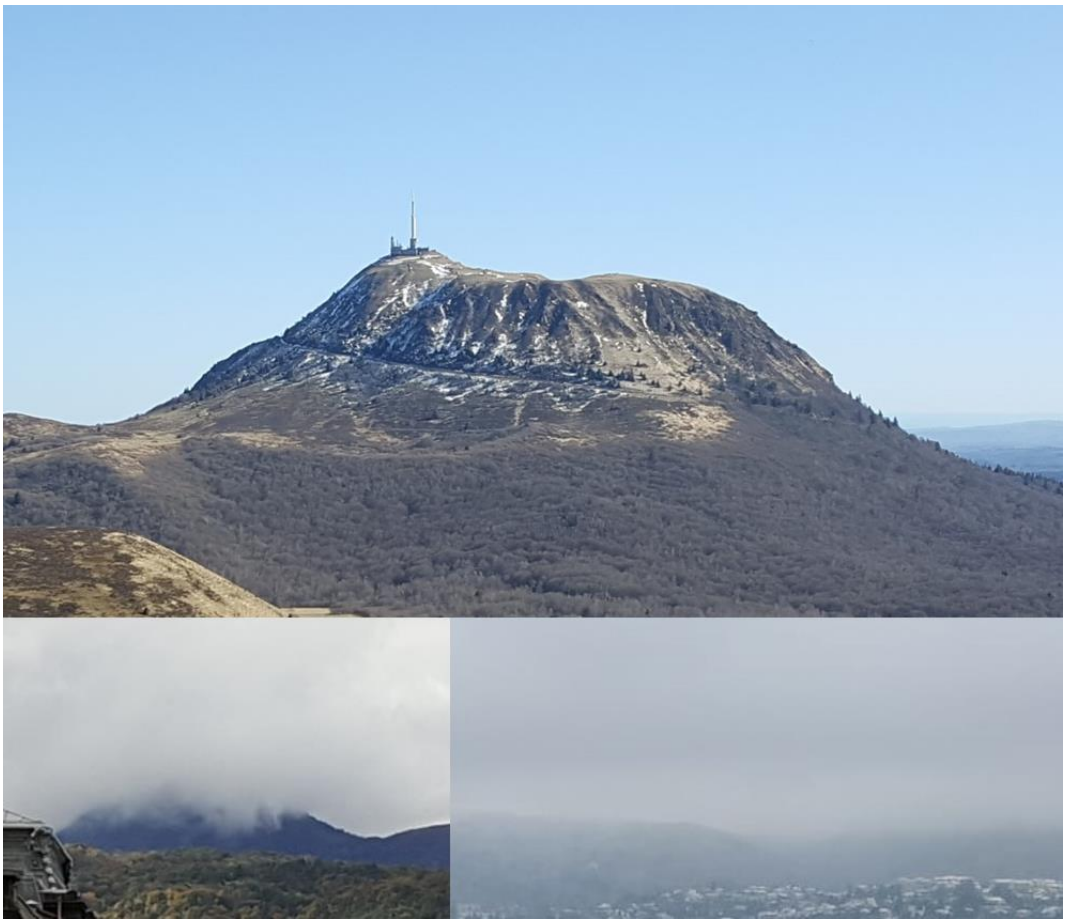


Figure 2: Photos of the puy de Dôme Mountain embedded or not in the cloud.

Rain samples were collected at Opme station (680 m a.s.l.), 12 km from the PUY station and with an altitude difference of 785 m (**Figure 1**). No experimentation room or sample processing facility is available at this site (essentially a field with meteorological measurement equipment), but it does provide access to a power source for the samplers and a closed field for sampling without passers-by around.

In addition, these stations provide access to continuous atmospheric measurement data such as wind speed and direction, temperature, or relative humidity (see **Annex 1** for more information on the equipment of each site).

2. Experimental procedure

The problematic of this thesis was to investigate whether clouds can be specific entities in the atmosphere in terms of biology. The objective was then to explore the potential microbiological specificities of clouds (i.e. biodiversity, functional profile) compared to other atmospheric compartments.

Environmental samples were collected during the first two years of this thesis (**Table 1**). In total, 10 clouds, 11 aerosols, and 7 rain were collected. For each of these samples, routine analyses were carried out (**Figure 3**): ATP quantification by bioluminescence, total cell count by flow cytometry, ion quantification by ion chromatography (Cl-, NO3-, SO42-, Na+, NH4+, K+, Mg2+ and Ca2+) and pH measurement. Meteorological conditions during sampling (temperature, wind speed and direction, and relative humidity) and backward trajectories of the air masses were provided by the LAMP.

Bacterial diversity was first investigated in clouds versus aerosols (**Chapter 2**) and in cloud versus rain (**Chapter 3**). Because most aerosols serve as cloud condensation nuclei (CCN), microbial diversity in aerosols is expected to be the primary contribution to cloud biodiversity. For rain, microbial diversity should be the combination of cloud biodiversity and microorganisms in the air column below. Indeed, precipitation scavenges aerosols, and thus biomass, from the air column as it falls (Radke et al., 1980). Bacterial diversity was studied by amplicon sequencing of the 16S ribosomal RNA gene (**Figure 3**). This approach allows the exploration of bacterial diversity in the environment and the identification of bacteria at family to genus level. However, it does not allow for absolute quantification and the data from sequencing must be treated as compositional datasets (Gloor et al., 2017). The bioinformatics was performed under the guidance and in collaboration of François Enault (see **section 4** for collaborations).

Next, metagenomics (MG) and metatranscriptomics (MT) approaches were used to study microbial diversity and functional profiling in aerosols and cloud events (**Chapter 4**). Direct sequencing was performed (i.e., without a prior amplification step) to limit quantification bias due to amplification. There are many challenges with these techniques, especially in the context of environmental studies: sample preservation during field sampling, low biomass of some ecosystems such as the atmosphere (Šantl-Temkiv et al., 2020), and recovery of nucleic acids (DNA and RNA) from very diverse and complex communities (Behzad et al., 2015). Moreover, bioinformatic processing of MG and MT data is not yet well defined, especially for MT. Some public bioinformatics workflows are available but there is no standard method. Therefore, a bioinformatics workflow was constructed according to our specific needs based on existing tools and, again, with the advice of François Enault. A collaboration was also initiated with the Galaxy team in Freiburg (Germany) to improve the bioinformatics workflow and make it freely accessible and usable on the Galaxy Europe platform.

Table 1: List of aerosol, cloud and rain samples collected during this thesis work and purpose. MG: metagenome; MT: metatranscriptome. Sample ID are written as following: date (yyyymmdd) and environmental type (CLOUD; AIR, i.e. aerosol; RAIN).

Sample ID	Sampling station	Amplicon 16S sequencing		MG and MT sequencing
		Cloud vs Air (Chapter 2)	Cloud vs Rain (Chapter 3)	Cloud vs Air (Chapter 4)
AEROSOLS				
20190712AIR	PUY	X		
20190918AIR	PUY	X		
20200206AIR	PUY	X		
20200518AIR	OPM	X		
20200610AIR	PUY	X		
20200707AIR	PUY	X		X
20200708AIR	PUY	X		X
20200709AIR	PUY	X		X
20200922AIR	PUY	X		X
20201118AIR	PUY	X		X
20201124AIR	PUY	X		X
Total		**11**	**-**	**6**
CLOUDS				
20190925CLOUD	PUY		X	
20191002CLOUD	PUY		X	X
20191022CLOUD	PUY	X	X	X
20200311CLOUD	PUY	X		X
20200717CLOUD	PUY	X		X
20201016CLOUD	PUY	X		X
20201028CLOUD	PUY	X		X
20201103CLOUD	PUY	X		X
20201110CLOUD	PUY	X		X
20201119CLOUD	PUY	X		X
Total		**8**	**3**	**9**
RAIN				
20191001RAIN	OPM		X	
20191015RAIN	OPM		X	
20191022RAIN	OPM		X	
20191023RAIN	OPM		X	
20191028RAIN	OPM		X	
20191031RAIN	OPM		X	
20191104RAIN	OPM		X	
Total		**-**	**7**	**-**

According to the multiple challenges associated with nucleic acid-based studies, verifications of our experimental procedure from sampling to sequencing were performed. The first paper developed in the following section includes the following aspects and supports the main results of this thesis work (**Figure 3**):

- Sampling and nucleic acid extraction procedure verifications
- Blanks and sequencing quality controls (i.e., mock communities).
- Three aerosol samples collected in replicate to verify the reproducibility between samplers during a single event.

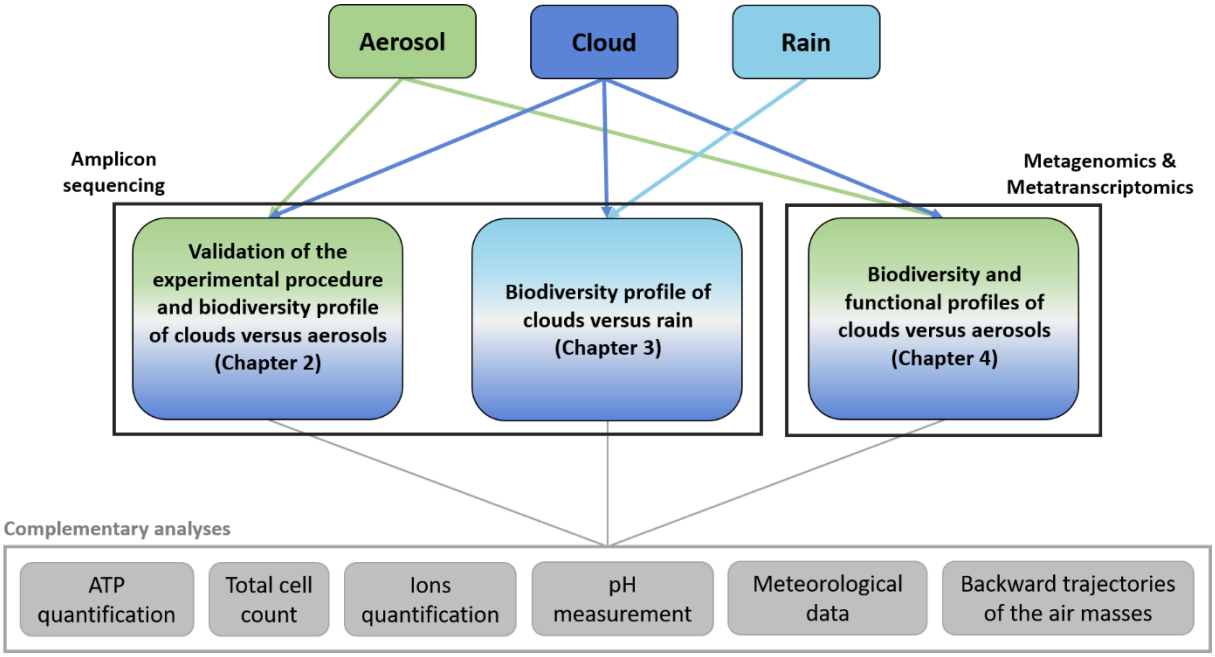


Figure 3: Schematic representation of the experimental procedure of this thesis work divided into three topics (respectively chapters 2, 3 and 4 of the manuscript). "Complementary analyses" are carried out for all the environmental samples and are part of each of the three topics.

3. Sampling challenges and controls

All the results of the controls and experiments presented here are described in **section 5**.

3.1. Challenges and choice of the sampler

Many challenges are associated with atmospheric sampling. First, given the low biomass in the atmosphere (~10^4 cells m^{-3} of air), samplers must have high flow-rate to collect enough biomass in a short time (less than 24 hours). A minimum of 5 ng of dry DNA is required for Illumina sequencing. We estimate, for example, that to obtain only 100 ng of total DNA from an air sample we had to sample for 7 hours with our sampler, based on the average DNA content per bacterial cell (~2.5×10^{-15} g DNA/cell) (Button and Robertson, 2001) and the average airborne cell concentration (~8×10^3 cells/m3 of air). Moreover, preservation of cell integrity and fixation of cells during sampling for several hours will be key elements to studying *in situ* microbial activity with RNA-based methods. For all these reasons, we chose to use impingement sampling. An impinger sampler is an active impactor (i.e., use a vacuum pump to create a specified airflow rate) where the particles impact a liquid surface. This sampling method is recommended to preserve cell integrity and can be used with nucleic acid preservative solution as the collection liquid. The ideal candidate was the High-flow-rate impinger (HFRi) (**Figure 4**), previously described and tested in biological assays in Šantl-Temkiv et al. (2017). The HFRi is a modified Kärcher DS5600 or DS6 vacuum cleaner that holds a 1.7 L volume of collection liquid and operates at a flow rate of 3,100 L min^{-1}. The large volume capacity of HFRi for collection liquid is as well an advantage because we need at least 150 mL for routine analysis only. Finally, this sampler can be used for both aerosol and cloud sampling, which limits potential collection bias. Controls and adjustments were made to improve the sampling method with the HFRi.

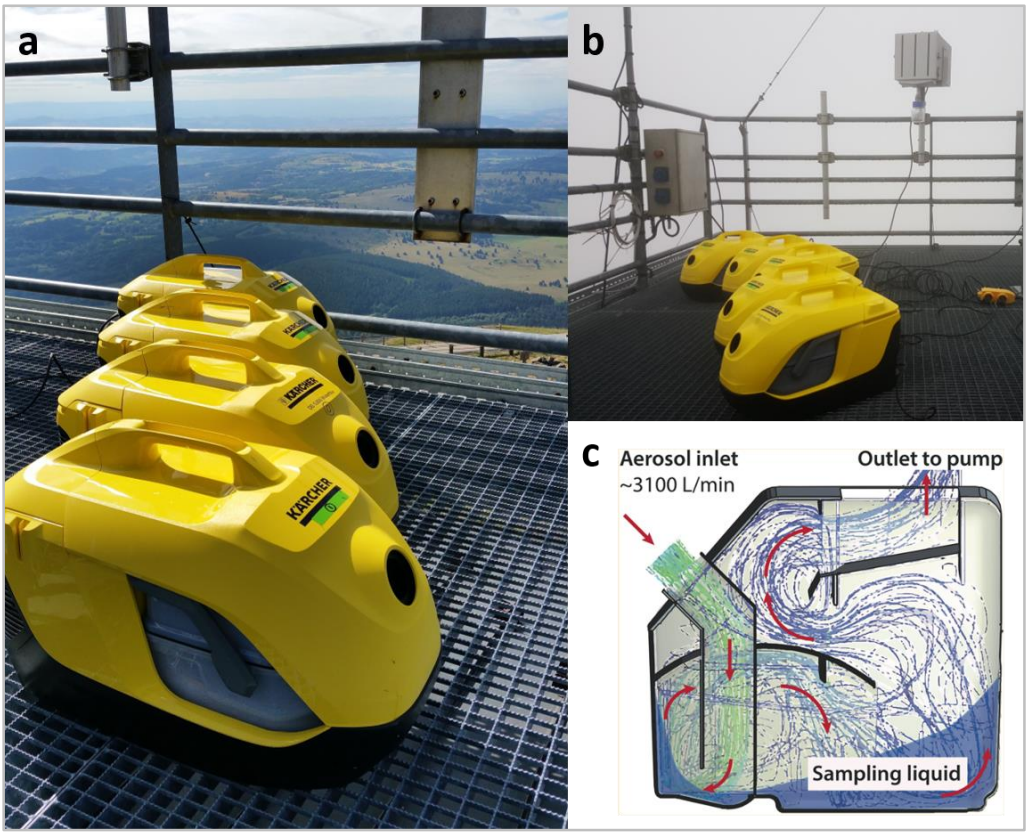


Figure 4: Photos of the High-Flow-Rate (HFR) impingers during aerosol (a) and cloud sampling (b) at Puy de Dôme (PUY) station, and schematic representation of the air flow in the impinger's tank (c; from Šantl-Temkiv et al., 2017).

3.2.Sampling controls and adjustments for HFRi

3.2.1. Sampling configuration for clouds and aerosols

The configuration for aerosol and cloud sampling is summarized in **Figure 5**. Details on controls and sampling preparation are described in the next sections.

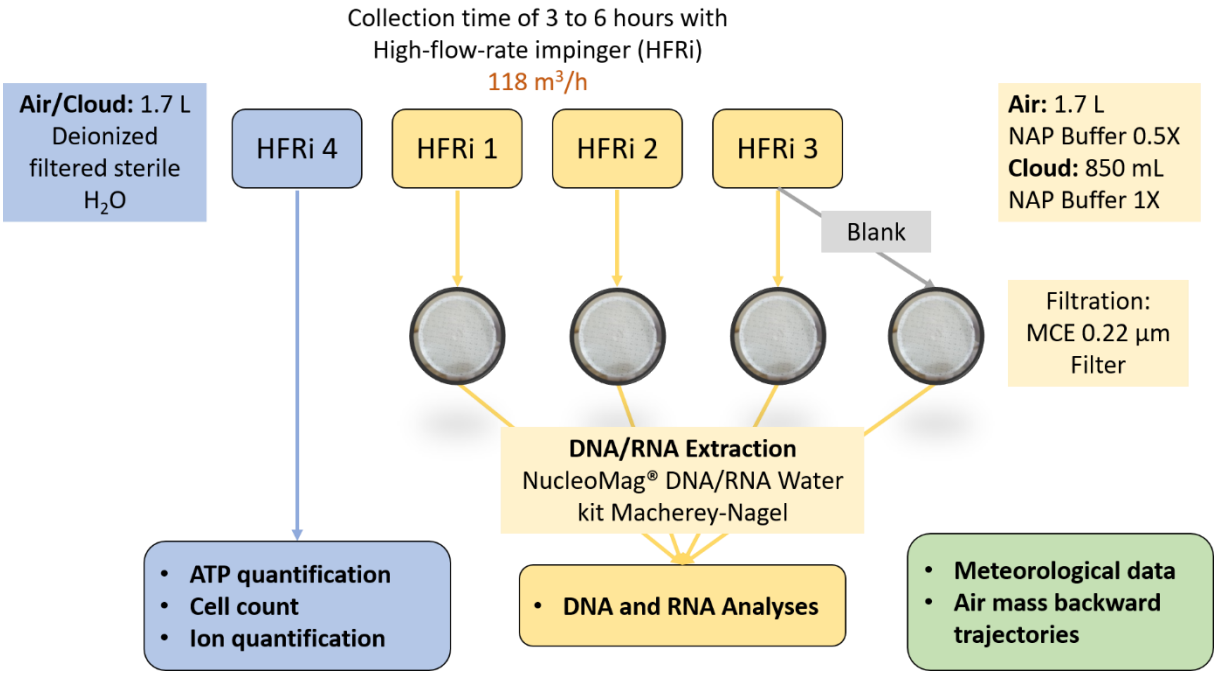


Figure 5: Schematic representation of the sampling configuration for cloud and aerosol sampling. Air: aerosol; HFRi: High Flow Rate impinger; NAP buffer: Nucleic Acid Preservation buffer; MCE: Mixed Cellulose Esters.

3.2.2. Nucleic acid preservation (NAP) buffer

The NAP buffer was needed to fix cells and prevent changes that would occur during sampling. Approximately 2.5 L of 1 X NAP buffer was required for one sampling event: 3 times 850 mL of 1 X NAP buffer for cloud sampling and 3 times 1,700 mL of 0.5 X NAP buffer for aerosol sampling (**Figure 5**). In the case of aerosol sampling, there was a loss of collection liquid through evaporation during sampling which was compensated by adding autoclaved ultrapure water every hour (we assumed that only H2O evaporates). For the clouds, the NAP buffer was diluted over time with the collection of cloud water during sampling. Therefore, only half a volume of 1X NAP buffer was used at the beginning of sampling. These differences between cloud and aerosol sampling could have induced some biases in the analysis.

Due to the large quantities required and the price of commercial products, NAP buffer was self-made according to the recipe of Camacho-Sanchez et al. (2013): 0.019 M of ethylenediaminetetra-acetic acid (EDTA), 0.018M of sodium citrate trisodium salt dehydrate, 3.8 M of ammonium sulfate, and H2SO4 to adjust the pH to 5.2. The NAP buffer was filtered on GF/F Glass Microfiber filters (47 mm

diameter, Whatman) to remove impurity particles, dispensed into 2 L bottles (sample-ready volume), and autoclaved.

3.2.3. Control of the contaminants

3.2.3.1. Decontamination of the HFRi tank

Sterilization of the sampler is an important question to ensure the sample is free of contamination. When possible, sampling material was autoclaved, but the HFRi tanks were made of polypropylene and could not be autoclaved. Therefore, the tanks were decontaminated as follows: rinsed with deionized H_2O, exposed to 2 liters of 70% ethanol in the tank for 10 min, and exposed to UV light for 10 min.

To validate the decontamination protocol, tree tanks were exposed to different conditions: 1) contamination and decontamination; 2) contamination and water rinse; 3) no contamination (**Figure 6**). HFRi tanks were intentionally contaminated with bacterial cells (*P. syringae* at 10^7 cell mL^{-1}) or ATP (~2.3 mg) in conditions 1 and 2. ATP was quantified by bioluminescence and the total cells were counted by flow cytometry at three sampling points: in the contaminated tank; in the clean water bottle before contact with the tank; and after 10 min of sampling. The results showed a return to blanks concentrations (in cell and ATP) in the impinger tanks after the standard decontamination protocol, and even with a simple water rinse. However, to be sure of the decontamination of the tank, we chose to keep the standard protocol. Detailed results are presented in **section 5**.

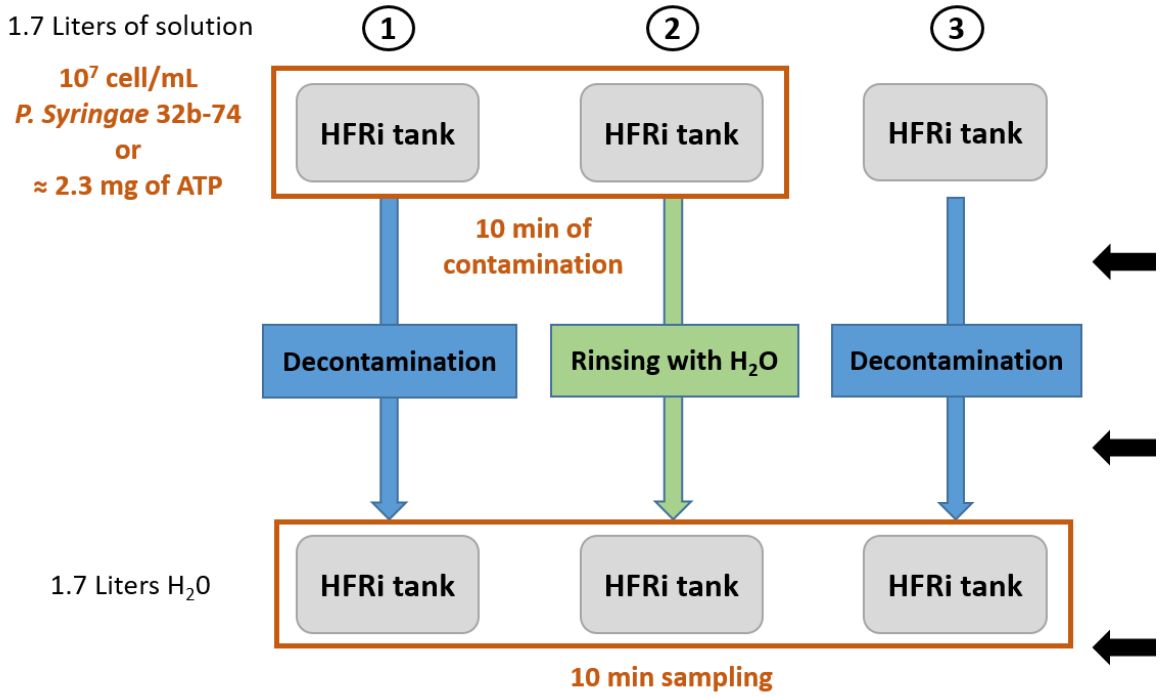


Figure 6: Schematic representation of the High-Flow-Rate impinger (HFRi) tank decontamination experiments.

3.2.3.2. Negative controls

Sampling blanks were performed before each sampling event. For nucleic acid analysis, the blank consisted in autoclaved NAP buffer left for 10 min in a HFRi tank. For other routine analyses, the blank was taken at t0 when water was poured into the tank.

Water blanks were also taken with clean autoclaved ultrapure water directly processed as environmental samples and sent to sequencing to monitor potential contaminants during sample processing.

3.2.4. Sampling duration

The amount of DNA recovered from samples was highly variable for aerosols and a longer sampling time was required to be sure of obtaining enough biological material for downstream analyses (**Figure 7**). For clouds, the amount of DNA was less variable between samples and a sampling time of 2 to 3 hours could be sufficient. However, unlike aerosols, DNA amounts did not increase with sampling time, which seems surprising. The efficiency of the HFRi in collecting clouds droplets (or evaporation) may be the cause, but we must also keep in mind that each cloud event was distinct, with different cloud density, droplet size and air mass origin. In addition, cloud sampling was conducted under generally windier meteorological conditions (than for aerosols) and strong wind may have interfered with the collection efficiency of the impinger.

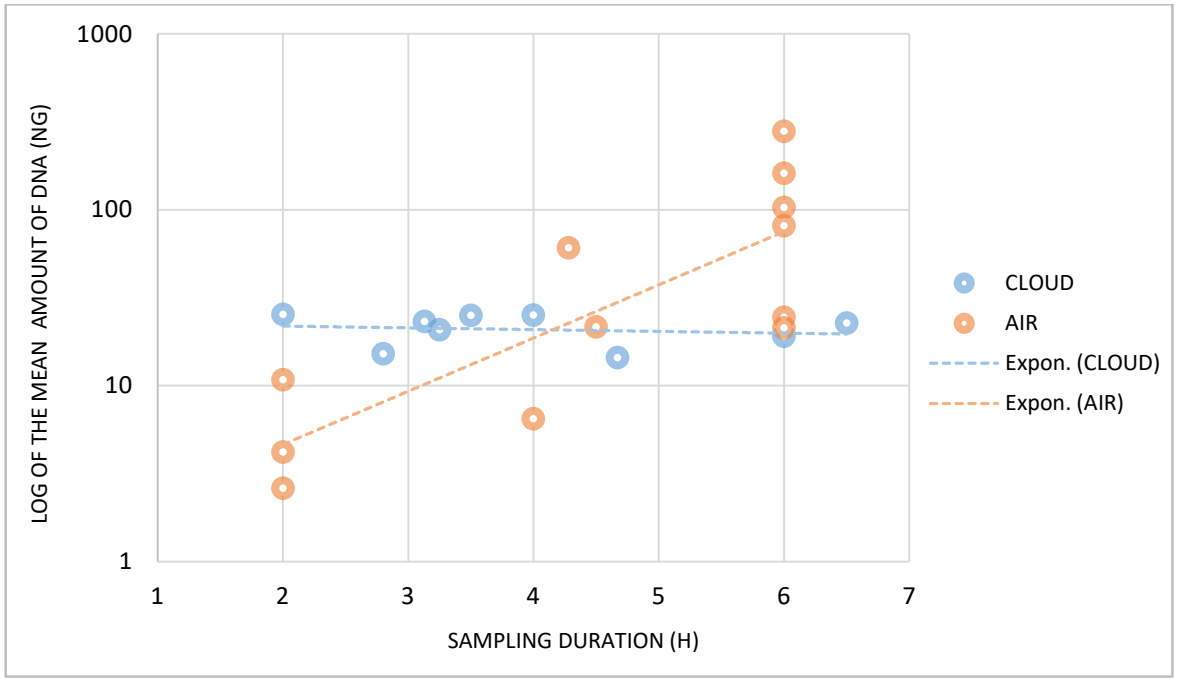


Figure 7: Quantities of DNA in ng extracted from filters as a function of sampling duration in hours. Expon.: exponential trend lines for cloud and aerosol (AIR) samples.

3.3.Quality control for sequencing

Mock communities were elaborated by mixing six strains isolated from cloud samples: *Pseudomonas syringae* 32b-74, *Bacillus sp.* 5b-1, *Sphingomonas sp.* 32b-11, *Rhodococcus enclensis* 23b-28, *Staphylococcus equorum* 5b-16 and *Flavobacterium tructae* 57b-18. Two types of mock communities were constructed: (1) the strains were pooled at given concentrations (**Table 2**) and the DNA from the pool was extracted (termed "Mock cloud") and sequenced; (2) the DNA was extracted separately from each strain and the same volumes of the six DNA extracts were mixed and sequenced (termed "Mock DNA"). These mock communities were used as controls for sequencing. The complete methodology and results are detailed in **Article 1 section 5**.

Table 2: Mock cloud and mock DNA composition

Bacterial strain name	**Cell concentration in Mock cloud**		**DNA extract concentration for Mock DNA**	
	cell mL^{-1}	**%**	**ng μL^{-1}**	**%**
***Pseudomonas syringae* 32b-74**	1.05×10^{8}	24 %	1.50	13 %
***Sphingomonas sp.* 32b-11**	1.05×10^{8}	24 %	2.34	20 %
***Rhodococcus enclensis* 23b-28**	6.61×10^{7}	15 %	0.17	1 %
***Bacillus sp.* 5b-1**	4.97×10^{7}	11 %	0.78	7 %
***Staphylococcus equorum* 5b-16**	3.45×10^{6}	1 %	0.82	7 %
***Flavobacterium tructae* 57b-18**	1.05×10^{8}	24 %	6.07	52 %

3.4.Comparison of nucleic acid extraction kits

Three extraction kits were tested: DNeasy PowerWater kit (QIAGEN), NucleoSpin Soil kit and NucleoMag® DNA/RNA Water kit (Macherey-Nagel). The latter was chosen because it allows the recovery of DNA and RNA simultaneously and was as efficient (i.e., recovered the same amounts of DNA) as the "Soil kit" which was itself better than the QIAGEN Water kit. The results are presented in **section 5**.

4. Collaborations

Several collaborations were undertaken during this thesis work:

- The sampling was carried out with the OPGC and LAMP laboratories on their meteorological stations (Laurent Deguillaume, Pascal Renard, Angelica Bianco). The origin of the air masses, the backward trajectories and the boundary layer height (BLH) were also provided (Jean-Luc Baray).
- The processing of the sequencing data was performed using the resources of the "Mésocentre Clermont Auvergne" and of the AuBI (Auvergne BioInformatique) network.
- The bioinformatics treatment of the sequencing data was carried out with the advice of François Enault (LMGE, BioADAPT team).
- The bioinformatic workflow for the processing of metagenomic and metatranscriptomic data (**Chapter 4**) has been improved with the collaboration of Bérénice Batut and Engy Nasr from the Galaxy team of Freiburg University. It is also planned to make this workflow publicly available on the Galaxy Europe platform.
- It is planned to improve the taxonomic affiliation of metagenomic and metatranscriptomic data in collaboration with Pierre Peyret and Sophie Marre (INRAE, UMR454 MEDIS; Clermont-Ferrand) by 16S rRNA reconstruction.
- The processing of 16S amplicon data for the study of clouds and aerosols bacterial diversity (**Chapter 2 - section 5**) has been improved by the MIGALE bioinformatics platform of INRAE (Jouy-en-Josas, FRANCE) with the work developed by Olivier Rué.

5. Results

5.1.Introduction to the studies

The atmosphere is a challenging environment to studied with nucleic acids-based approach given the low biomass of airborne microorganisms and the variability in diversity and richness of atmospheric microbial communities. For our goal of sequencing without prior amplification step, sufficient nucleic acid concentrations were required (minimum of 5 ng) and the integrity of nucleic acid had to be maintained as best as possible during sampling. In addition, strict controls and decontamination protocol must be used to ensure that these low biomass environmental samples were not contaminated.

This results section is divided into two articles. **Article 1** describes controls and adjustments to the sampling protocol performed to validate our experimental procedure. Through the sequencing controls, issues were discovered regarding the construction of the operational taxonomic unit (OTU) clusters. Two strains were clustered together and could not be distinguished with the initial workflow (FROGS for clustering) used in the amplicon-based study in **Chapter 3**. A collaboration was therefore initiated to correct these issues and improve the clustering with the MIGALE bioinformatics platform (the creators of the FROGS tool). **Article 2** reports the first attempt to describe and compare bacterial diversity in aerosols and cloud events by amplicon sequencing. This study was also a way to check possible variabilities between samplers and the samples collected for a same event. It is indeed important to be aware of possible sampling variations between HFRi as replicates were pooled for downstream sequencing analyzes in the cases of the studies performed in **Chapter 3** and **4**.

This entire section of work supported the main results described in **Chapter 4** on the functional analysis of microbial communities in clouds versus aerosols.

In this work I participated in:

- The collection and processing of cloud and aerosol samples
- The preparation of the NAP buffer
- The realization of the controls experiments (decontamination and extraction tests)
- The extraction and amplification of DNA with tagged primers
- The implementation of the first bioinformatics workflow and data processing
- The writing of publications

5.2. Article 1: "Instrumental procedures for the study of atmospheric eDNA", submitted to the *Environmental DNA* journal

Instrumental procedures and recommendations for the study of atmospheric eDNA

Running Title: Instrumental procedures for atmospheric eDNA study

Raphaëlle Péguilhan[1], Florent Rossi[1], Olivier Rué[2,3], Muriel Joly[1], and Pierre Amato[1]

Author contributions: RP has made major contribution to the acquisition, analysis and interpretation of the data. FR and MJ helped with data acquisition. OR contributed to the data analysis. PA conceived and designed the study, interpreted the data. RP and PA wrote the manuscript.

[1] Université Clermont Auvergne, CNRS, SIGMA Clermont, ICCF, F-63000 CLERMONT-FERRAND, FRANCE

[2] Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France

[3] Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France

Acknowledgment: We thank OPGC and LaMP laboratories staff for the access to the CO-PDD instrumented stations. We are grateful to the Aubi platform and to the Mésocentre Clermont Auvergne for providing computing and storage. We are thankful to flow cytometry department (LMGE laboratory) for the use of their instruments. This research was supported by the French National Research Agency (ANR) (grant no. ANR-17-MPGA-0013), Université Clermont Auvergne, CNRS-INSU and CNES.

Abstract

The atmosphere is a complex environment hosting viable and potentially active microbial communities. They can be transported over long distance and disseminate in surface ecosystems, which can have an impact on the local microbial ecology. They can also impact the chemistry and microphysics of the atmosphere through their metabolic activity, and their ability to be active cloud condensation nuclei and ice nuclei. Therefore, there is great interest in studying the biodiversity and functioning of airborne communities on a global scale through nucleic acid-based analyses, but this comes with many challenges regarding the atmosphere: low biomass, long sampling time to compensate for low biomass, detection of contamination and sequencing artifacts, and preservation of the *in situ* state of the sample during collection. Furthermore, good environmental sampling practices are not always applied in aeromicrobiology: controls, sampling and analysis replicates, etc. Here, we proposed an experimental procedure for sampling and studying nucleic acids from aerosols and clouds. This procedure involves several impingers with nucleic acid preservation buffer as collection liquid, to increase the collection capacity and to fix the cells during collection. The impingers allowed collection of both aerosols and clouds, limiting collection bias, and have a high flow rate that allowed sufficient biomass to be collected in a short time (< 24 h) to avoid overmixing of air masses. The use and combined analysis of blank samples, sequencing quality controls and sequencing replicates permitted monitoring and identification of unavoidable contaminants. Sampling replicates are also necessary for more robust statistics and identification of short term variations. Finally, this experimental procedure leads to the recovery of sufficient amounts of DNA to do amplicon sequencing, or even metagenomics. This method may also be applied for microbial activity studies such as metatranscriptomics thanks to the preservation buffer

Keywords: Clouds, aerosols, eDNA, method, controls, microorganisms

Introduction

The atmosphere is a dynamic environment known to contain and transport microorganisms to high altitudes and over long distances (Després et al., 2012; Smith et al., 2013). Airborne microorganisms can be dispersed to continental scales (Barberán et al., 2015; Smith et al., 2013) and disseminated to local surface ecosystems by dry and wet deposition (Barberán et al., 2014; Morris et al., 2004). A fraction of the airborne microorganisms has been demonstrated to be viable and maintain metabolic activity (Amato et al., 2019; Krumins et al., 2014; Šantl-Temkiv et al., 2017), which can have consequences for atmospheric chemistry (Khaled et al., 2021; Wirgot et al., 2017). Bacteria, as aerosols, also play roles as cloud condensation nuclei (CCN) (Bauer et al., 2003; Zhang et al., 2021) and ice nuclei (IN) (Möhler et al., 2007). Therefore, it is of major interest for microbial ecology and atmospheric chemistry and microphysics to monitor airborne microbial diversity and functioning.

Aeromicrobiology is still an emerging discipline of environmental microbiology (Šantl-Temkiv et al., 2020). As analytical methods improve, deepest investigations of the structural and functional diversity of the atmospheric microbiome, and its environmental drivers, are made possible. Nevertheless, this field is still in its infancy, and most studies do not yet adopt the basics of recommended practices in microbial ecology in other environmental compartments: need for replicate sampling and analysis, controls, etc. (**Table 1**). Constrains related to the difficulty of accessing sufficient biomass in the atmospheric environment (low biomass: 10^2 to 10^6 cells m^{-3} of air) are important. The first prerequisite for meaningful analysis is to be able to distinguish real target(s) from contaminants caused by handling, experimenters, equipment and reagents. In low biomass environments such as the atmosphere, this is not trivial. Additionally, depending on the objectives, sampling must be carried out over short periods of time to avoid smoothing the data, thus missing eventual short-term trends. Long-term sampling using classical means (filtration, impaction) also alters sample integrity and prevents functional

analyses. High throughput sampling solutions circumvent these problems, and impingers allow sampling in liquids including nucleotides, proteins or other fixing agents to preserve the *in situ* state of the sample during collection. Moreover, with the improvement of technologies and the advent of NGS sequencing techniques, new challenges have emerged, such as detection of sequencing bias, artifacts, and contaminating sequences (de Goffau et al., 2018).

Here, we propose an experimental procedure for studying the biodiversity and activity of airborne microorganisms and their drivers using nucleic acid-based analyses applicable to aerosol and clouds. This involves replicated sampling using several high-flow-rate impingers (HFRi) deployed in parallel, a fixative as sampling fluid, controls and characterization of contaminants at several steps of the experimental procedure, and evaluation of the quantitative accuracy of amplicon sequencing.

Table 1: Procedures used in the literature for aerosol sampling, controls, and sequencing.

SAMPLING

Article	Environment type	Sampler type	Sampler flow rate	liquid/filter of collection	Collection liquid volume (mL)	Sampling duration
Amato *et al.* 2017	clouds	Cloud droplet impactor	108 m^3/h	Satured ammonium sulphate solution	200	3 to 24 h
Archer *et al.* 2019	aerosols	High-volume liquid impinger (coriolis µ)	300 L/min	RNA later (Invitrogen)	15	4 hours
Archer *et al.* 2020	aerosols	High-volume liquid impinger (coriolis µ)	300 L/min	Phosphate-buffered saline (PBS)	15	0.5 to 4 hours
Barberán *et al.* 2014	aerosols	Automatic dry/wet passive collector (MTX ARS 1010)	-	Whatman GF/F filter	-	≈15 days
Bowers *et al.* 2011	aerosols	Vacuum filtration	≈30 L/min	0.22 µm cellulose nitrate filter	-	1.5 hours
Bowers *et al.* 2013	aerosols	Dichotomous filter sampler	50 L/min	47 mm quartz fiber filter	-	24 hours
Fröhlich-Nowoisky *et al.* 2012	aerosols	Different high-volume filter samplers	50-1130 L/min	Glass fiber filters/Quartz fiber filter	-	2 to 72 hours / 7 days
Péguilhan *et al.* 2021	clouds	High-flow rate impinger (DS 5600 impinger)	2 m^3/min	Nucleic acid preservation buffer	1,700	2 to 8 hours
Romano *et al.* 2019	aerosols	Low-volume filter sampler (HYDRA-FAI dual sampler)	2.3 m^3/h	47 mm PTFE filter	-	48 to 72 hours
Šantl-Temkiv *et al.* 2018	aerosols	High-flow rate impinger (DS 5600 impinger)	0.8-0.9 m^3/min	High-salt solution	1,500	5 hours
Smith *et al.* 2018	aerosols	Filter sampler	-	Gelatinous filter membrane	-	141 to 250 min
Tignat-Perrier *et al.* 2020	aerosols	High-volume filter sampler (PM10)	30-70 m^3/h	Quartz fiber filter	-	7 days

CONTROL AND SEQUENCING

Article	Blank / Positive control	Sequencing of negative controls	Analysis / Sampling replicate	Technology of sequencing	Amplicon length (bp)	Variable region targeted (16S)
Amato *et al.* 2017	yes/no	no	no	Illumina MiSeq	291	V4
Archer *et al.* 2019	yes/yes	no	no	Illumina MiSeq	464	V3-V4
Archer *et al.* 2020	yes/yes	yes	no	Illumina MiSeq	464	V3-V4
Barberán *et al.* 2014	no/no	no	no	454 pyrosequencing	291	V4
Bowers *et al.* 2011	yes/no	yes	no	454 pyrosequencing	260	V2
Bowers *et al.* 2013	yes/no	no	no	Illumina MiSeq	291	V4
Fröhlich-Nowoisky *et al.* 2012	yes (**not every time**)/yes	yes	no	Cloning and sequencing by PCR	-	-
Péguilhan *et al.* 2021	yes/no	no	no	Illumina MiSeq	394	V5-V6-V7
Romano *et al.* 2019	yes/no	yes	no	Illumina MiSeq	464	V3-V4
Šantl-Temkiv *et al.* 2018	yes/no	yes	no	Ion Torrent PGM	283	V4
Smith *et al.* 2018	yes/no	yes	no	Illumina MiSeq	-	V4
Tignat-Perrier *et al.* 2020	yes/no	yes	no	Illumina MiSeq	464	V3-V4

Materials and methods

Sampling setup with High-Flow-Rate Impingers (HFRi)

The HFRi sampler is a commercial Kärcher DS 5600 or DS6 vacuum cleaner (Kärcher SAS, Bonneuil sur Marne, France) that can contain up to 2 Liters of collection liquid and operates at an airflow rate of 2 m^3 min^{-1} (Šantl-Temkiv et al., 2017). This sampler has been used to collect cloud and aerosol samples.

Sampling of aerosols and clouds was carried out at the summit of puy de Dôme mountain (1,465 m a.s.l, France) using the facilities of the atmospheric station (Baray et al., 2020; Péguilhan et al., 2021). Three to five HFRi were used in parallel (**Figure 1**): one sampler, filled with 1.7 L of autoclaved H_2O (850 mL for clouds), was dedicated to ATP quantification and total cell count; the remaining 2-4 HFRi, filled with 1.7 L of 0.5 X nucleic acid preservation (NAP) buffer solution (850 mL of 1 X NAP buffer for clouds), were dedicated to nucleic acid analysis. The collection liquid was processed (filtrated on 0.2µm) immediately after sampling in a laminar flow hood previously exposed to UV light for 15 min.

NAP buffer was self-made following instructions in Camacho-Sanchez *et al.* 2013. It is composed of 0.019 M of ethylenediaminetetra-acetic acid (EDTA; ref. PB118-500, Fisher BioReagents™), 0.018 M of sodium citrate trisodium salt dehydrate (ref. S/3320/53, Fisher Chemical) and 3.8 M of ammonium sulfate (ref. 446450050, ACROS ORGANICS) in milliQ water, with H_2SO_4 to adjust the pH to 5.2. The NAP buffer was then filtered through Glass Microfiber filters GF/F (47 mm diameter; Whatman, CAT. N. 1825-047; Maidstone, United Kingdom) to remove impurity particles, aliquoted in glass bottles and was finally sterilized by autoclave.

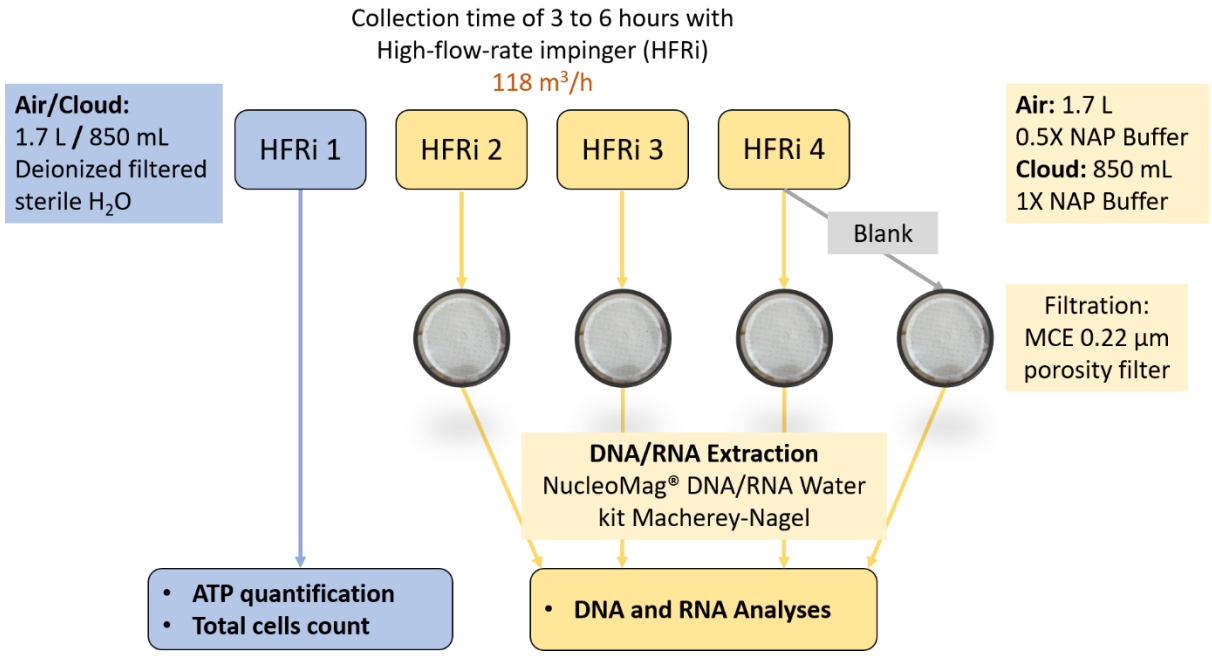


Figure 1: Experimental procedure for the study of nucleic acids in clouds and aerosols. Air: aerosol; HFRi: High-Flow-Rate impinger; NAP buffer: Nucleic Acid Preservation buffer; MCE: Mixed Cellulose Esters.

Decontamination and negative controls

The standard decontamination procedure consisted of: (1) rinsing of the tank with H_2O and dH_2O, (2) exposure to 2 L of 70 % ethanol for 10 min and (3) UV light exposure (254 nm) for 10 min. Ethanol and UV exposures were chosen because they are commonly used to decontaminate materials instead of heating them (Archer et al., 2019; Dommergue et al., 2019; Šantl-Temkiv et al., 2017) and because HFRi tanks cannot be autoclaved because they are made of polypropylene.

To estimate the efficiency of HFRi tanks decontamination, samplers were intentionally contaminated with a *Pseudomonas syringae* 32b-74 culture (10^7 cell mL^{-1}), or adenosine-5'-triphosphate (ATP) disodium salt hydrate (Sigma-Aldrich, Saint-Louis, Missouri, USA) (2.3 mg in 1,7 L of water). These amounts of bacteria and ATP were chosen to be above typical concentrations in atmospheric environmental samples (Vaïtilingom et al., 2012). Three

samplers were used to test three different conditions (**Supplementary Figure 1**): (1) contamination and decontamination with the standard procedure; (2) contamination and rinsing with deionized H_2O; (3) no contamination and decontamination with the standard procedure. To verify whether or not contamination remained after decontamination, 1.7 L of autoclaved water (used to prepare the NAP buffer) was added in the tank and processed in the same conditions as samples for ATP and total cells measurements. A control was also performed on the autoclaved water.

Sampling blanks were performed for each event: 1.7 L of 0.5X NAP buffer (aerosols) or 850 mL of 1X NAP buffer (clouds) was poured into one of the HFRi tanks, left there for 10 min, and then collected in a clean autoclaved bottle before sampling. The blanks were then treated as environmental samples. Four were randomly selected for sequencing (see specific section).

Water blanks were also performed to detect possible contamination during sample processing and nucleic acid extraction. Autoclaved ultrapure water was directly filtered on 0.22 µm MCE membrane filter (cf reference below), processed, and sent to be sequenced like the environmental samples.

Quality control for biodiversity profiling

Mock communities were constructed by mixing six bacterial strains isolated from the atmosphere (Amato et al., 2007; Lallement et al., 2017; Vaïtilingom et al., 2012): *Pseudomonas syringae* 32b-74 (GenBank ID: HQ256872), *Bacillus* sp. 5b-1 (DQ512749), *Sphingomonas* sp. 32b-11 (HQ256831), *Rhodococcus enclensis* 23b-28 (DOVD00000000), *Staphylococcus equorum* 5b-16 (DQ512761) and *Flavobacterium* sp. 57b-18 (KR922118.1).

DNA was extracted either from independent cultures and mixed after extraction (Mock DNA) or from mixed cultures with known cell concentration for each strain (Mock Cloud). The bacterial strains were cultured separately in 10 mL of liquid R2A at 17°C.

153 For Mock DNA samples, DNA extraction was performed following the protocol of the
154 QIAamp DNA Mini kit (QIAGEN, Hilden, Germany) with minor changes: 1 mL of each culture
155 was centrifuged at 14,000 g after 4 days of incubation and the pellets were re-suspended in 180
156 µL of TE (1X), with 25 µL of lysozyme (50 mg/mL) and 5 µL of RNase (1 mg/mL). The
157 mixture was vortexed and incubated 30 min at 37°C. Twenty microliters of Protease K and 200
158 µL of Buffer AL were added. The mixture was vortexed again and incubated first during 30
159 min at 56°C, then for 15 min at 95°C. After extraction, DNA was quantified using the Quant-
160 iT™ PicoGreen® dsDNA kit (Invitrogen; Thermo Fisher Scientific, Waltham, MA USA) and
161 "Mock DNA" aliquots were prepared by adding 2 µL of each strain DNA extract, and stored at
162 -80°C (**Supplementary Figure 2**; **Supplementary Table 1**).
163 For Mock cloud samples, cell concentration of each culture was estimated by flow
164 cytometry at their exponential phase (21-44h incubation). The six strains were mixed at known
165 concentrations (**Supplementary Table 1**) and 1 mL aliquots were stored in 10% glycerol at -
166 80°C. Three aliquots were processed as the environmental samples and DNA extraction was
167 performed using the NucleoMag® DNA/RNA Water (Macherey-Nagel, Hoerdt, France). Three
168 additional aliquots were extracted using the QIAamp DNA Mini kit (QIAGEN; Hilden,
169 Germany) as above, and named "Mock Cloud QIAamp" (**Supplementary Figure 2**). The DNA
170 extracts were stored at -80°C.
171 Theoretical distributions were calculated for both "Mock cloud" and "Mock DNA"
172 samples, and were normalized by the average ribosome copy number from the Ribosomal RNA
173 Database (rrnDB, v 5.7) for each genus (**Supplementary Table 1**), except for *Sphingomonas*
174 sp. for which we had the related genome containing 4 ribosome copies.
175
176

Total cell count and ATP quantification

Total cells were quantified by flow cytometry with BD FacsCalibur instrument (Becton Dickinson, Franklin Lakes, NJ), following the protocol in (Amato et al., 2017). Adenosine-5'-triphosphate (ATP) quantifications were performed by bioluminescence (ATP Biomass Kit HS; BioThema; Handen, Sweden) as in (Vaïtilingom et al., 2013).

Nucleic acid extraction and amplification

Three nucleic acid extraction kits were compared: DNeasy PowerWater kit (QIAGEN; Hilden, Germany), NucleoSpin Soil, and NucleoMag® DNA/RNA Water kits (Macherey-Nagel, Hoerdt, France), referred to as the "Water", "Soil" and "Air" kits, respectively. The Water and Soil kits, and the Soil and Air kits were tested in pairs during two sampling events each and the DNA extractions were performed in triplicate using three samplers. After each sampling event, the collection liquid (1.7 L of NAP buffer by HFRi) was filtered on 0.22 µm mixed cellulose esters (MCE) membranes (47 mm diameter; ClearLine 0421A00023). The filters were then cut equally into two pieces for extraction with 2 kits and stored at -80°C until processing.

Prior to use, all "working" surfaces were treated with RNase away spray solution (Thermo Scientific; Waltham, USA). Extractions were performed following the manufacturers' protocols for the Water and Soil kits. In the case of the Air kit, the following adaptations were made to the protocol: immediately after sampling, the collection liquid from the tank of each HFRi was filtered through 0.22 µm MCE membranes, bead-beating was applied to the filters with 1,200 µL of lysis buffer MWA1, and stored at -80°C in Beads Tubes 5 mL Type A (Macherey-Nagel, ref. 740799.50). For DNA extraction, ~600 µL of beads-beating lysate were processed following the protocol adapted for 47 mm filter membranes. The lysate was RNA-treated by adding 1:50 volumes of RNase A (12 mg/mL, stock solution). Finally, DNA was

eluted in 50 µL of RNase-free H_2O with an incubation time of 5 min at 56°C. DNA was quantified using Quant-iT™ PicoGreen® dsDNA kit (Invitrogen; Thermo Fisher Scientific, Waltham, MA USA).

Amplification of the 16S subunit of bacterial ribosomal gene was performed from genomic DNA extracts by PCR targeting the V4 region, using the primers 515f (5'-GTGYCAGCMGCCGCGGTAA-3') (Parada et al., 2016) and 806r (5'-GGACTACNVGGGTWTCTAAT-3') (Apprill et al., 2015). The PCR mix was modified as follows: each 50 µL reaction volume contained 2 µL of sample, 10 µL of 5X Platinum II PCR Buffer (Invitrogen; Thermo Fisher Scientific, Waltham, MA USA), 5 µL of Platinum GC Enhancer, 1 µL of 10 nM dNTPs (Sigma-Aldrich; Merck, Darmstadt Germany), 1 µl of 10 µM forward and reverse primers and 0.2 µL of Platinum II Taq HS DNA pol (Invitrogen). PCR amplification conditions are described on the "Earth Microbiome Project" website (https://earthmicrobiome.org/). Amplicons were purified using the QIAquick PCR Purification kit (QIAGEN; Hilden, Germany) and sequenced by Illumina Miseq 2*250 bp (GenoScreen; Lille, France).

Bioinformatics data processing and statistics

Amplicon sequence variants (ASVs) were obtained from raw 16S reads with the dada2 package (v 1.20.0) (Callahan et al., 2016); functions filterAndTrim, learnErrors, dada, mergePairs, makeSequenceTable and removeBimeraDenovo following authors guidelines. Then, FROGS (Bernard et al., 2021) was used to affiliate ASVs versus SILVA v138.1 (Quast et al., 2013). When the BLAST assignation was questionable (i.e., multi-affiliations, percent identity <95%, or percent query coverage <98%), they were verified using the RDP assignation and the EzBioCloud 16S rRNA gene-based ID database (Yoon et al., 2017; https://www.ezbiocloud.net/, update 2021.07.07). ASVs affiliated to *Chloroplast* (110), to

Mitochondria (285), to Archaea (5) and ASVs without Blast affiliation (439) were removed. All samples were processed together to the rarefaction step; the common abundance table was divided into three parts containing the mock samples (**Supplementary Table 2**), blanks (**Supplementary Table 3**), and environmental samples (only three aerosols presented here; **Supplementary Table 4**), respectively. The mock, blanks and environmental samples were rarefied to 28,100, 1,770 and 2,770 sequences respectively (corresponding to the sample with the lowest number of reads), using "FROGS Abundance normalization". This left 22, 152 and 862 ASVs respectively.

The ASV abundance data were centered log-ratio (CLR)-transformed, as recommended by Gloor et al. (2017) to account for their compositional nature. Data analysis was performed and represented using the *R* environment (v 4.0.3) (R Core Team (2019)). The *zCompositions* package (v 1.3.4) (Palarea-Albaladejo and Martín-Fernández, 2015) was used to replace null counts in our compositional data based on a Bayesian-multiplicative method (function *cmultRepl* using CZM method and an input format in pseudo-counts) and to CLR-transform the abundance table (*clr* function). Heatmaps were obtained using the packages *pheatmap* (v 1.0.12) (Raivo Kolde, 2019) and *ggdendro* (v 0.1.22) (Andrie de Vries and Brian D. Ripley, 2016). Statistical tests were performed using PAST (v 4.02) (Hammer et al., 2001).

Results and discussion

Controls and decontamination

Negative controls were performed throughout sampling and sample processing to monitor potential contaminants. First, total cell counts and ATP quantification were similar to blank concentrations after intentional contamination and decontamination with the standard protocol or with a H_2O rinse (Mann-Whitney tests between blanks and concentrations after decontamination had a *p-value* > 0.1) (**Figure 2**; **Supplementary Figure 3**). Thus, a simple H_2O rinse does as well as the decontamination procedure. However, DNA concentrations were not quantified, and to be sure of the effectiveness of the decontamination, we choose to continue with the standard procedure.

Second, we did not detect any common contamination between autoclaved water and NAP buffer left for 10 min in HFRi tanks (**Figure 3**). Some genera such as *Blastococcus*, *Pelomonas*, *Stenotrophomonas*, *Rhodococcus* or *Sphingomonas* were detected, but were only present in few samples (two to three), with an abundance not exceeding 630 reads. Additionally, the sampling and water blanks were not clearly distinct, indicating the absence of significant contamination from the clean HFRi tanks compared to what is found in the filtration step and nucleic acid extraction kits.

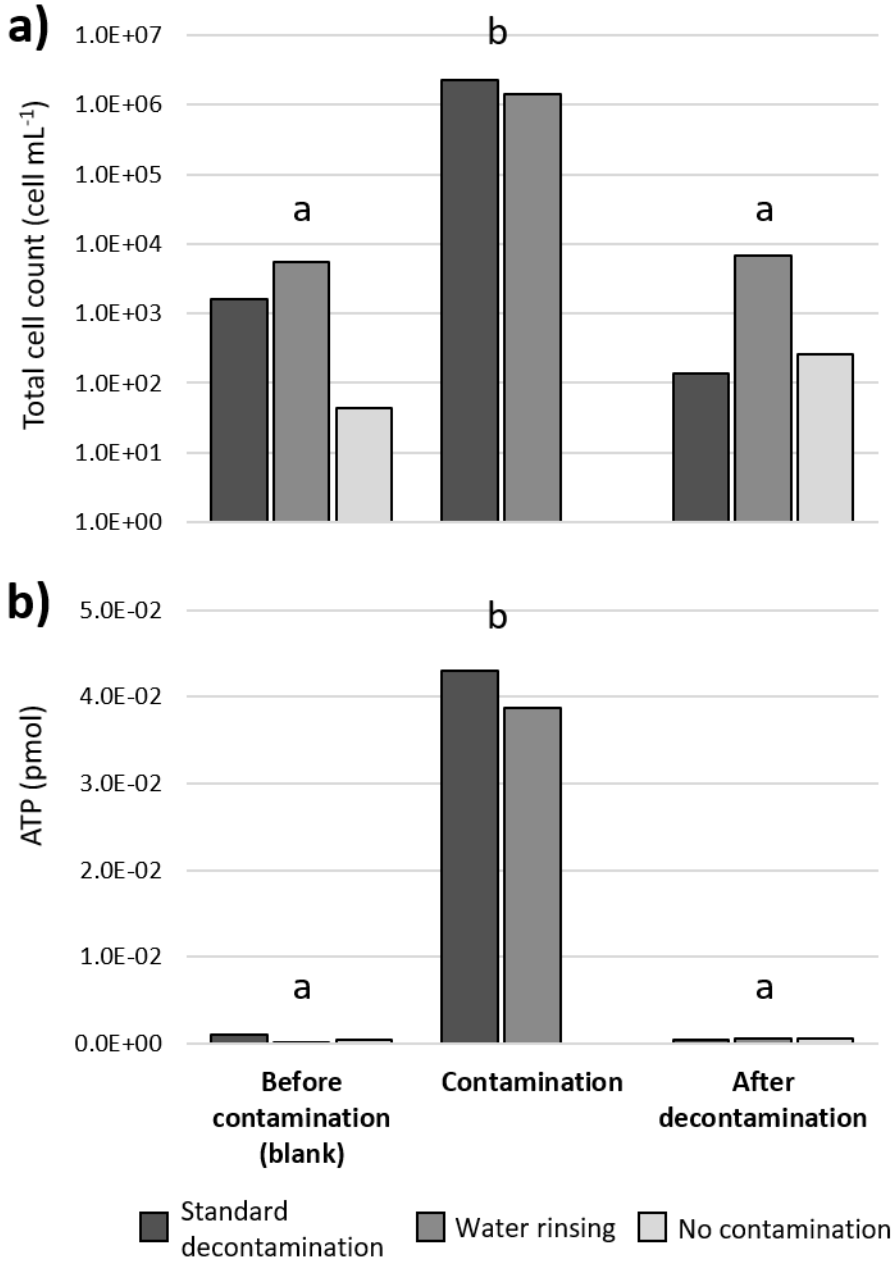


Figure 2: Total cell (a) and ATP concentrations (b) before and after intentional contamination and decontamination. a, b: Mann-Whitney tests between blanks and concentrations after decontamination had a *p-value* > 0.1.

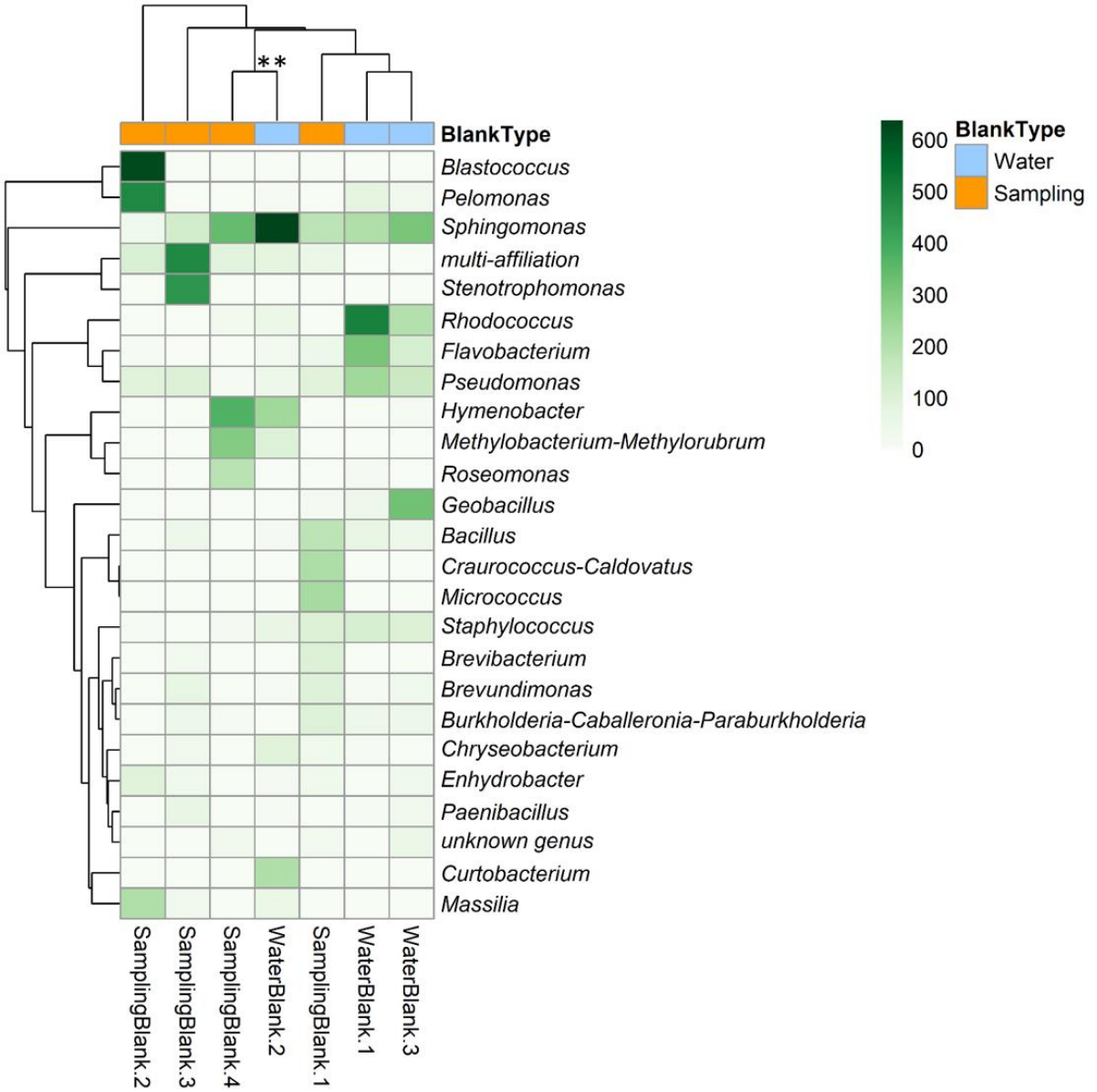


Figure 3: The 25 most abundant bacterial genera present in blank samples. Water blanks were autoclaved ultra-pure water and sampling blanks were NAP buffer solution in contact with High-Flow-Rate impinger tank for 10 min before sampling. Green scale represents the number of sequencing reads. Hierarchical clustering was performed with the Euclidean distance and the Ward's method (ward.D2). **: the approximately unbiased (AU) p-value is > 99 ('pvclust' R package).

Positive controls (i.e., sequencing controls) were also performed to check for potential amplification and sequencing biases. The six genera composing the mock community were successfully recovered from bioinformatics processing and ASV affiliation with 8 highly abundant ASVs (two affiliated to *Flavobacterium* and two to *Staphylococcus*; **Supplementary Table 2**). A few other low abundant ASVs were most likely sequencing artifacts (in total, less than 0.05% of the reads corresponding to < 40 reads per variant sequence). The use of mock communities as a positive control may therefore provide clues to identify potential contaminants and artifacts such as in our dataset: *Kocuria* sp., *Methylobacterium* sp., or *Curtobacterium* sp. For the "Mock DNA" controls, *Flavobacterium* and *Rhodococcus* were less represented in the sequenced samples than would be expected based on the amounts of DNA in the original samples (theoretical "Mock DNA") (**Figure 4**). For the "Mock cloud" controls, the proportions were well conserved, except for *Staphylococcus*, which was overrepresented, and *Bacillus*, which was underrepresented in the sequenced samples compared to the expected distribution. Underestimated strains may have been poorly targeted by the primers (Parada et al., 2016). In the case of *Staphylococcus*, this genus is known to make cellular aggregates (Zeng et al., 2008) and thus, may have been under-quantified by flow cytometry. As a conclusion on the positive controls, they demonstrated that we should be careful with the quantitative aspect of amplicon sequencing datasets and with rare species, which are most likely contaminants or artifacts. Sequencing quality controls, like our mock communities, are rarely performed in biodiversity profiling studies (**Table 1**) but are nevertheless important for being aware of sequencing biases and artefacts (de Goffau et al., 2018) and for estimating the efficiency of primers to amplify and differentiate certain species of interest (Parada et al., 2016; Rajendhran and Gunasekaran, 2011; Reysenbach et al., 1992).

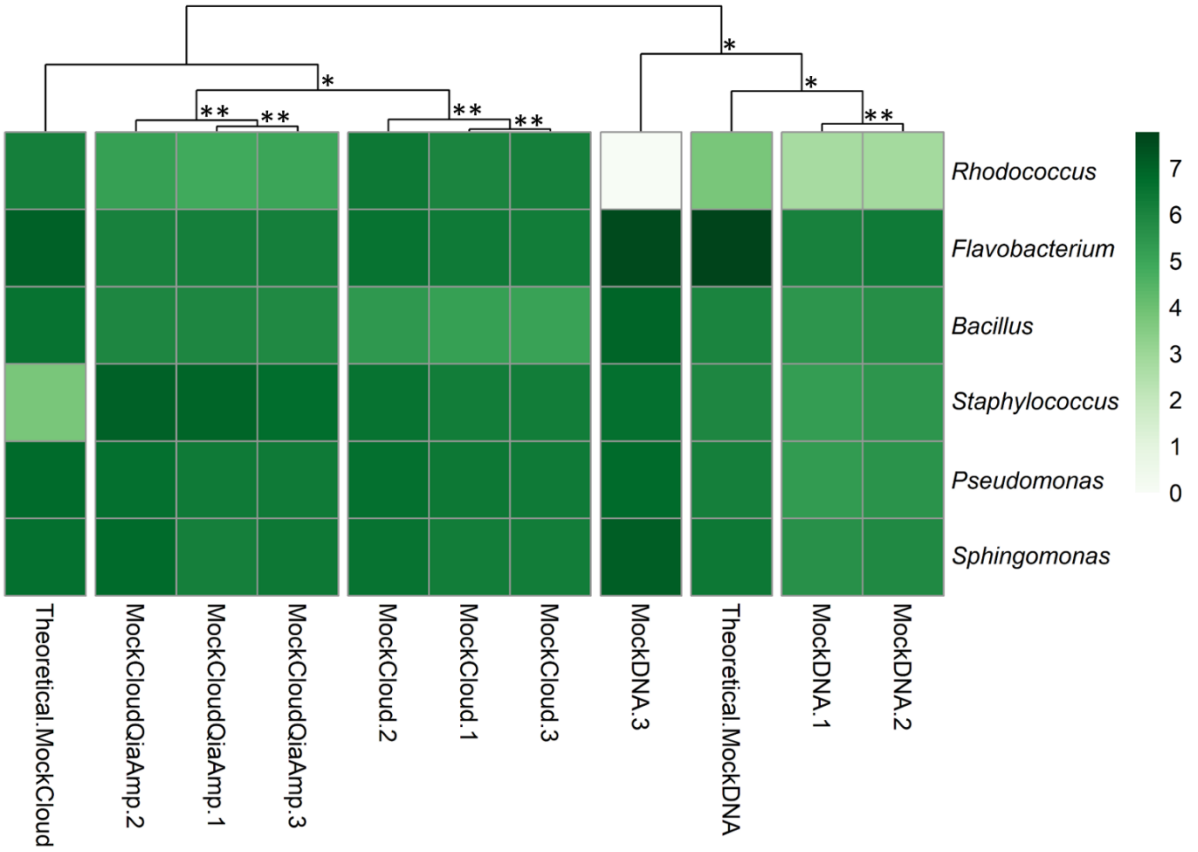


Figure 4: Composition of the "mock cloud" and "mock DNA" samples and their theoretical compositions. Green scale represents centered-log ratios (clr)-transformed number of affiliated sequences. Hierarchical clustering were performed with the Euclidean distance and the Ward's method (ward.D2). *: the approximately unbiased (AU) p-value is ≥ 95; **: AU p-value > 99.

The presence of negative controls is essential, especially in aerobiology where the collected biomass is very low at high altitudes (~10^4 cell m^{-3}) and thus can be easily confounded with contaminants, reagent microbiome or sequencing artifacts (de Goffau et al., 2018; Šantl-Temkiv et al., 2020). Blank samples are applied most of the time (**Table 1**), but to be most efficient, negative controls should be performed at several stages of sample collection and processing (de Goffau et al., 2018), and should be fully analyzed like other samples (in our case sequenced), which is not always achieved in environmental studies. Similarly, decontamination of all

equipment, must be done and monitored, especially when the material is not easily decontaminated such as HFRi tanks that cannot be autoclaved.

Validation of the experimental procedure

One of the main objectives of this study was to optimize the experimental procedure to collect enough biomass in a short time (< 24 h) to obtain the minimum amount of nucleic acid required for typical Illumina protocols (> 5 ng; Dommergue et al., 2019; Quick et al., 2017); and this without any prior amplification step.

Experimental procedures have been proposed to study nucleic acids from the outdoor atmosphere as in Dommergue *et al.* 2019, based on filtration sampling on quartz fiber filters for 7 days. However, this methodology still has limitations such as the small amounts of DNA recovered, the mixing of air masses over a week (which does not allow for short-term variations to be observed), and the use of filtration that can be very destructive to cell integrity and activity, depending on the analyses we want to perform.

The HFRi was chosen for our sampling procedure because it has one of the highest airflow rate (2 m^3 min^{-1}) and one of the largest volume capacities to hold the collection liquid (up to 2 L) (**Table 1**). Sampling by impingement is less destructive to cell integrity and viability than filtration (Griffin et al., 2011) and can also allow the use of a highly saline solution as NAP buffer to fix cells during sampling and thus preserve the *in situ* state of the sample. Three of the collected aerosol events are presented with their replicates (**Figure 5**). The replicates were all clustered by sampling date (*p-value* > 95), even though the events were collected on consecutive days, demonstrating reproducibility between samplers for the same event. Therefore, sample replicates can be pooled for the same event for subsequent sequencing analyses requiring larger amounts of nucleic acids. Sample replicates are rarely performed in environmental nucleic acid-based studies (**Table 1**), however, triplicate analyses to ASVs affiliation provide more robust

statistics and the ability to discriminate even closely related samples. In addition, replicates for the same sample can detect potential contaminants by revealing taxonomic groups absent in some of the replicates (de Goffau et al., 2018).

Several acid nucleic extraction kits were tested to select the most suitable for low biomass and high diversity samples. Each nucleic acid extraction kit must be compared in pairs for the same event, as each sampling event had specific meteorological conditions and collection time (**Figure 6**). The Soil kit performed better than the Water kit (i.e. more DNA was extracted). Therefore, the Soil kit was retained and compared to the Air kit. This time, the Air kit performed the best for low concentration samples. The Air kit also allows the extraction of DNA and RNA at the same time and was therefore selected.

This kit was then tested on 11 aerosols and 8 cloud samples and recovered 0.05 – 9.01 ng μL^{-1} of DNA for aerosols and 0.25 – 0.50 ng μL^{-1} for clouds (for a total of ~360-720 m^3 of air collected) (**Table 2**), which is on average an order of magnitude higher than what was recovered in (Dommergue et al., 2019) on filters with one week of sampling (total of ~5,040-11,760 m^3 of air collected) and the DNeasy PowerWater kit (i.e. "Water" kit): 0.08 ± 0.03 ng μL^{-1}.

The ATP concentrations and total cell numbers for the three aerosol events presented averaged 6.39 ± 3.18 pmol m^{-3} and $1.98\times10^3 \pm 5.53\times10^2$ cell m^{-3} of air, which correspond to typical concentrations found in the atmosphere (Bauer et al., 2002; Vaïtilingom et al., 2012). This again confirms the ability of HFRi to efficiently collect airborne microorganisms.

To conclude here, the amounts of nucleic acid recovered from the clouds and aerosols were sufficient to performed Illumina sequencing for biodiversity profiling (16S rRNA amplicon) and for additional sequencing analyses. The experimental procedure is therefore validated for nucleic acid-based atmospheric analyses.

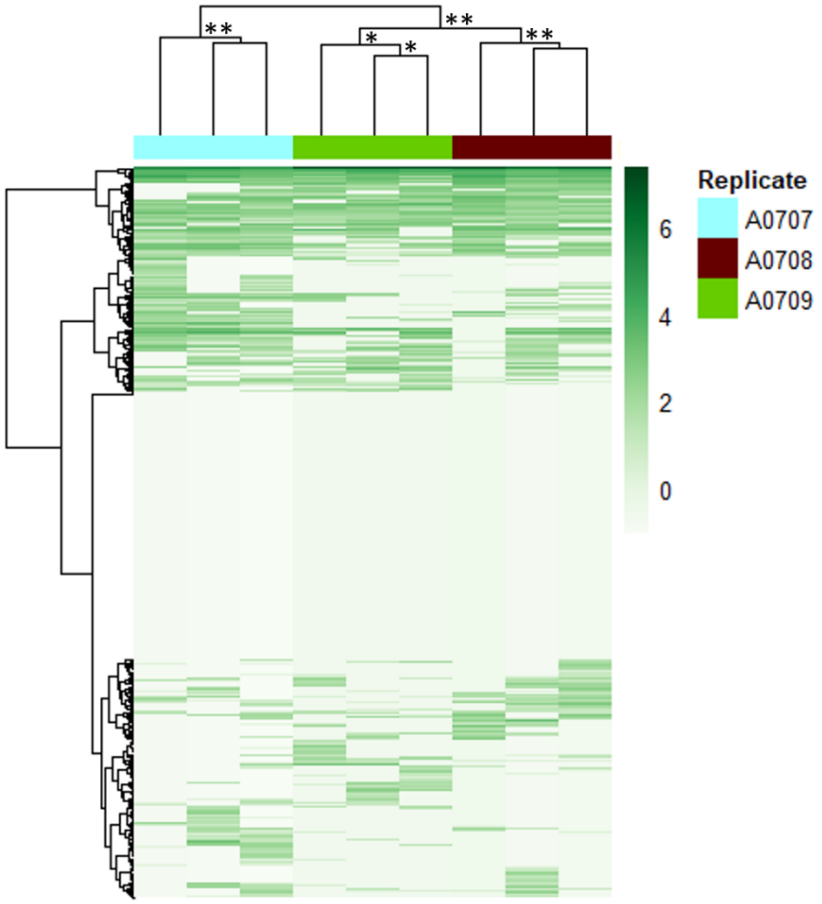
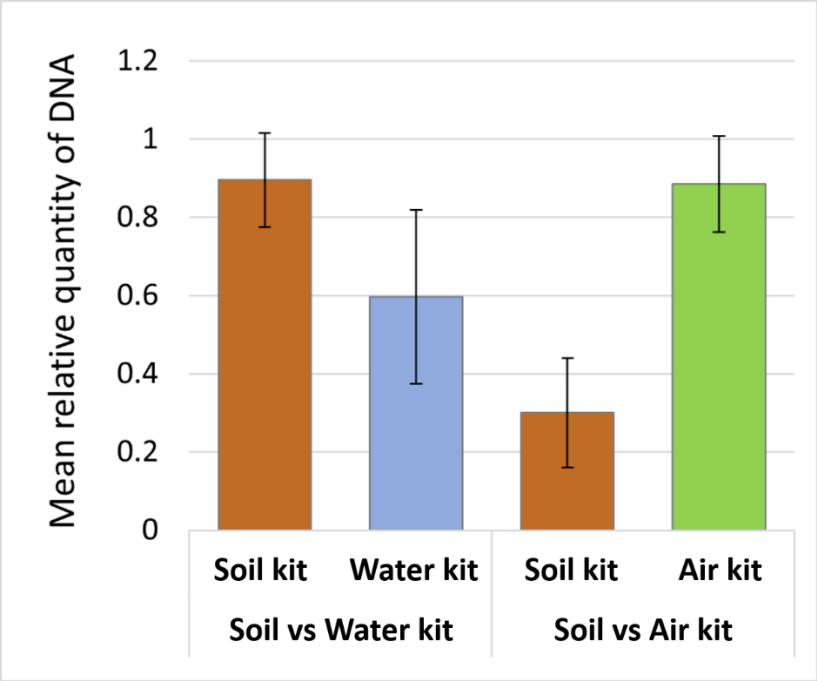


Figure 5: **Total bacterial genera in three aerosol samples collected in triplicate on consecutive days.** Green scale represent the number of affiliated sequences centered-log ratios (clr)-transformed. Hierarchical clustering was performed with the Euclidean distance and the Ward's method (ward.D2). *: the approximately unbiased (AU) p-value is ≥ 95; **: AU p-value ≥ 99.



368

369 **Figure 6: Paired-comparison of three nucleic acid extraction kits. Water**: DNeasy
370 PowerWater kit (QIAGEN; Hilden, Germany); **Soil**: NucleoSpin Soil kit (Macherey-Nagel,
371 Hoerdt, France); **Air**: NucleoMag® DNA/RNA Water kit (Macherey-Nagel, Hoerdt, France).

372

373 **Table 2: DNA concentrations in environmental samples extracted with the Air kit.**
374 Extracted with the Air kit (NucleoMag DNA/RNA Water kit, Macherey-Nagel).

	Mean DNA concentration (ng µl^{-1}) in aerosols (n = 11)*	**Mean DNA concentration (ng µl^{-1}) in clouds (n = 8)***
Minimum	0.05	0.25
Maximum	9.01	0.50
Median	1.22	0.40
Mean	2.20	0.40
Standard error	2.80	0.09

375 *In total volume of 50 µl

Conclusion

In conclusion, the use of multiple negative and positive controls validated our experimental procedure and is essential for good practice in nucleic acid-based studies. The presence of contaminants or sequencing artifacts is inevitable, as evidence by blank samples. It is therefore essential to be able to monitor and identify them. Multiple approaches are needed to achieve this, such as combined analysis of blank samples (negative controls), sequencing quality controls (positive controls), and sampling replicates. Finally, our experimental procedure based on multiple HFRi with NAP buffer as collection liquid allowed the recovery of sufficient amounts of DNA from atmospheric samples in a short period of time (≤ 6 h) to perform nucleic acid-based analyses (several amplicon sequencing, metagenomics) without overmixing the air masses. Another advantage of HFRi is the possibility to collect clouds and aerosols with the same sampler, thus limiting collection bias. The NAP buffer also allows to study the potential activity of airborne communities as for example by metatranscriptomics. In addition, the sampling replicates will allow for more robust statistics and differentiation of short-term variations.

Data Archiving Statement: Demultiplexed sequencing files were deposited at the European Nucleotide Archive and have accession numbers ERR9924984 to ERR9924999, and ERR9924950 to ERR9924958.

References

Amato, P., Parazols, M., Sancelme, M., Laj, P., Mailhot, G., and Delort, A.M. (2007). Microorganisms isolated from the water phase of tropospheric clouds at the Puy de Dôme: Major groups and growth abilities at low temperatures. In FEMS Microbiology Ecology, pp. 242–254.

Amato, P., Joly, M., Besaury, L., Oudart, A., Taib, N., Moné, A.I., Deguillaume, L., Delort, A.M., and Debroas, D. (2017). Active microorganisms thrive among extremely diverse communities in cloud water. PLoS One *12*, 1–22.

Amato, P., Besaury, L., Joly, M., Penaud, B., Deguillaume, L., and Delort, A.M. (2019). Metatranscriptomic exploration of microbial functioning in clouds. Sci. Rep. *9*.

Andrie de Vries and Brian D. Ripley (2016). ggdendro: Create Dendrograms and Tree Diagrams Using "ggplot2". R package version 0.1-20.

Apprill, A., Mcnally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat. Microb. Ecol. *75*, 129–137.

Archer, S.D.J., Lee, K.C., Caruso, T., Maki, T., Lee, C.K., Cary, S.C., Cowan, D.A., Maestre, F.T., and Pointing, S.B. (2019). Airborne microbial transport limitation to isolated Antarctic soil habitats. Nat. Microbiol. *4*, 925–932.

Archer, S.D.J., Lee, K.C., Caruso, T., King-Miaow, K., Harvey, M., Huang, D., Wainwright, B.J., and Pointing, S.B. (2020). Air mass source determines airborne microbial diversity at the ocean–atmosphere interface of the Great Barrier Reef marine ecosystem. ISME J. *14*, 871–876.

Baray, J.L., Deguillaume, L., Colomb, A., Sellegri, K., Freney, E., Rose, C., Baelen, J. Van, Pichon, J.M., Picard, D., Fréville, P., et al. (2020). Cézeaux-Aulnat-Opme-Puy de Dôme: A multi-site for the long-term survey of the tropospheric composition and climate change. Atmos. Meas. Tech. *13*, 3413–3445.

Barberán, A., Henley, J., Fierer, N., and Casamayor, E.O. (2014). Structure, inter-annual recurrence, and global-scale connectivity of airborne microbial communities. Sci. Total Environ. *487*, 187–195.

Barberán, A., Ladau, J., Leff, J.W., Pollard, K.S., Menninger, H.L., Dunn, R.R., and Fierer, N. (2015). Continental-scale distributions of dust-associated bacteria and fungi. PNAS *112*, 5756–5761.

Bauer, H., Kasper-Giebl, A., Zibuschka, F., Hitzenberger, R., Kraus, G.F., and Puxbaum, H. (2002). Determination of the carbon content of airborne fungal spores. Anal. Chem. *74*, 91–95.

Bauer, H., Giebl, H., Hitzenberger, R., Kasper-Giebl, A., Reischl, G., Zibuschka, F., and Puxbaum, H. (2003). Airborne bacteria as cloud condensation nuclei. J. Geophys. Res.

Atmos. *108*, 4658.

Bernard, M., Rué, O., Mariadassou, M., and Pascal, G. (2021). FROGS: a powerful tool to analyse the diversity of fungi with special management of internal transcribed spacers. Brief. Bioinform. *22*.

Bowers, R.M., McLetchie, S., Knight, R., and 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.

Bowers, R.M., Clements, N., Emerson, J.B., Wiedinmyer, C., Hannigan, M.P., and Fierer, N. (2013). Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. Environ. Sci. Technol. *47*, 12097–12106.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 2016 137 *13*, 581–583.

Camacho-Sanchez, M., Burraco, P., Gomez-Mestre, I., and Leonard, J.A. (2013). Preservation of RNA and DNA from mammal samples under field conditions. Mol. Ecol. Resour. *13*, 663–673.

Després, V.R., Alex Huffman, J., Burrows, S.M., Hoose, C., Safatov, A.S., Buryak, G., Fröhlich-Nowoisky, J., Elbert, W., Andreae, M.O., Pöschl, U., et al. (2012). Primary biological aerosol particles in the atmosphere: A review. Tellus, Ser. B Chem. Phys. Meteorol. *64*.

Dommergue, A., Amato, P., Tignat-Perrier, R., Magand, O., Thollot, A., Joly, M., Bouvier, L., Sellegri, K., Vogel, T., Sonke, J.E., et al. (2019). Methods to investigate the global atmospheric microbiome. Front. Microbiol. *10*.

Fröhlich-Nowoisky, J., Burrows, S.M., Xie, Z., Engling, G., Solomon, P.A., Fraser, M.P., Mayol-Bracero, O.L., Artaxo, P., Begerow, D., Conrad, R., et al. (2012). Biogeography in the air: Fungal diversity over land and oceans. Biogeosciences *9*, 1125–1136.

de Goffau, M.C., Lager, S., Salter, S.J., Wagner, J., Kronbichler, A., Charnock-Jones, D.S., Peacock, S.J., Smith, G.C.S., and Parkhill, J. (2018). Recognizing the reagent microbiome. Nat. Microbiol. *3*, 851–853.

Griffin, D.W., Gonzalez, C., Teigell, N., Petrosky, T., Northup, D.E., and Lyles, M. (2011). Observations on the use of membrane filtration and liquid impingement to collect airborne microorganisms in various atmospheric environments. Aerobiologia (Bologna). *27*, 25–35.

Hammer, Ø., Harper, D.A.T., and Ryan, P.D. (2001). Past: Paleontological statistics software package for education and data analysis. Palaeontol. Electron. *4*, 1–9.

Khaled, A., Zhang, M., Amato, P., Delort, A.M., and Ervens, B. (2021). Biodegradation by bacteria in clouds: An underestimated sink for some organics in the atmospheric multiphase system. Atmos. Chem. Phys. *21*, 3123–3141.

Krumins, V., Mainelis, G., Kerkhof, L.J., and Fennell, D.E. (2014). Substrate-Dependent rRNA Production in an Airborne Bacterium. Environ. Sci. Technol. Lett. *1*, 376–381.

Lallement, A., Besaury, L., Eyheraguibel, B., Amato, P., Sancelme, M., Mailhot, G., and Delort, A.M. (2017). Draft Genome Sequence of Rhodococcus enclensis 23b-28, a Model Strain Isolated from Cloud Water. Genome Announc. *5*.

Möhler, O., DeMott, P.J., Vali, G., and Levin, Z. (2007). Microbiology and atmospheric processes: The role of biological particles in cloud physics. Biogeosciences *4*, 1059–1071.

Morris, C.E., Georgakopoulos, D.G., and Sands, D.C. (2004). Ice nucleation active bacteria and their potential role in precipitation. J. Phys. IV *121*, 87–103.

Palarea-Albaladejo, J., and Martín-Fernández, J.A. (2015). ZCompositions - R package for multivariate imputation of left-censored data under a compositional approach. Chemom. Intell. Lab. Syst. *143*, 85–96.

Parada, A.E., Needham, D.M., and Fuhrman, J.A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ. Microbiol. *18*, 1403–1414.

Péguilhan, R., Besaury, L., Rossi, F., Enault, F., Baray, J., Deguillaume, L., and Amato, P. (2021). Rainfalls sprinkle cloud bacterial diversity while scavenging biomass. FEMS Microbiol. Ecol. 1–15.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. *41*, 590–596.

Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T.F., Beutler, N.A., et al. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat. Protoc. *12*, 1261–1266.

R Core Team (2019) R Core Team (2019). R: A language and environment for statistical computing.

Raivo Kolde (2019). pheatmap: Pretty Heatmaps. R package version 1.0.12.

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

Reysenbach, A.L., Giver, L.J., Wickham, G.S., and Pace, N.R. (1992). Differential amplification of rRNA genes by polymerase chain reaction. Appl. Environ. Microbiol. *58*, 3417–3418.

Romano, S., Di Salvo, M., Rispoli, G., Alifano, P., Perrone, M.R., and Talà, A. (2019). Airborne bacteria in the Central Mediterranean: Structure and role of meteorology and air mass transport. Sci. Total Environ. *697*.

Šantl-Temkiv, T., Amato, P., Gosewinkel, U., Thyrhaug, R., Charton, A., Chicot, B., Finster,

K., Bratbak, G., and Löndahl, J. (2017). High-Flow-Rate Impinger for the Study of Concentration, Viability, Metabolic Activity, and Ice-Nucleation Activity of Airborne Bacteria. Environ. Sci. Technol. *51*, 11224–11234.

Šantl-Temkiv, T., Gosewinkel, U., Starnawski, P., Lever, M., and Finster, K. (2018). Aeolian dispersal of bacteria in southwest Greenland: Their sources, abundance, diversity and physiological states. FEMS Microbiol. Ecol. *94*, 1–10.

Šantl-Temkiv, T., Sikoparija, B., Maki, T., Carotenuto, F., Amato, P., Yao, M., Morris, C.E., Schnell, R., Jaenicke, R., Pöhlker, C., et al. (2020). Bioaerosol field measurements: Challenges and perspectives in outdoor studies. Aerosol Sci. Technol. *54*, 520–546.

Smith, D.J., Timonen, H.J., Jaffe, D.A., Griffin, D.W., Birmele, M.N., Perry, K.D., Ward, P.D., and Roberts, M.S. (2013). Intercontinental dispersal of bacteria and archaea by transpacific winds. Appl. Environ. Microbiol. *79*, 1134–1139.

Smith, D.J., Ravichandar, J.D., Jain, S., Griffin, D.W., Yu, H., Tan, Q., Thissen, J., Lusby, T., Nicoll, P., Shedler, S., et al. (2018). Airborne bacteria in earth's lower stratosphere resemble taxa detected in the troposphere: Results from a new NASA Aircraft Bioaerosol Collector (ABC). Front. Microbiol. *9*, 1–20.

Tignat-Perrier, R., Dommergue, A., Thollot, A., Magand, O., Amato, P., Joly, M., Sellegri, K., Vogel, T.M., and Larose, C. (2020). Seasonal shift in airborne microbial communities. Sci. Total Environ. 137129.

Vaïtilingom, M., Attard, E., Gaiani, N., Sancelme, M., Deguillaume, L., Flossmann, A.I., Amato, P., and Delort, A.M. (2012). Long-term features of cloud microbiology at the puy de Dôme (France). Atmos. Environ. *56*, 88–100.

Vaïtilingom, M., Deguillaume, L., Vinatier, V., Sancelme, M., Amato, P., Chaumerliac, N., and Delort, A.-M. (2013). Potential impact of microbial activity on the oxidant capacity and organic carbon budget in clouds. Proc. Natl. Acad. Sci. *110*, 559–564.

Wirgot, N., Vinatier, V., Deguillaume, L., Sancelme, M., and Delort, A.M. (2017). H2O2 modulates the energetic metabolism of the cloud microbiome. Atmos. Chem. Phys. *17*, 14841–14851.

Yoon, S.H., Ha, S.M., Kwon, S., Lim, J., Kim, Y., Seo, H., and Chun, J. (2017). Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int. J. Syst. Evol. Microbiol. *67*, 1613–1617.

Zeng, P., Moy, B.Y.P., Song, Y.H., and Tay, J.H. (2008). Biodegradation of dimethyl phthalate by Sphingomonas sp. isolated from phthalic-acid-degrading aerobic granules. Appl. Microbiol. Biotechnol. *80*, 899–905.

Zhang, M., Khaled, A., Amato, P., Delort, A.M., and Ervens, B. (2021). Sensitivities to biological aerosol particle properties and ageing processes: Potential implications for aerosol-cloud interactions and optical properties. Atmos. Chem. Phys. *21*, 3699–3724.

Supplementary information to:

Instrumental procedures and recommendations for the study of atmospheric eDNA

Raphaëlle Péguilhan[1], Florent Rossi[1], Olivier Rué[2,3], Muriel Joly[1], and Pierre Amato[1]

[1] Université Clermont Auvergne, CNRS, SIGMA Clermont, ICCF, F-63000 CLERMONT-FERRAND, France
[2] Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France
[3] Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France

List of supplements:

- **Supplementary Tables and Figures**

Supplementary Table 1: Cell and DNA concentrations of the mock communities.

Supplementary Table 2: Taxonomy and count table for the mock samples. Number of sequences affiliated to the total 22 ASVs detected in mock samples.

Supplementary Table 3: Taxonomy table for the blank samples. Number of sequences affiliated to the 20 most abundant ASVs detected in blank samples (over 152 ASVs).

Supplementary Table 4: Taxonomy table for the three aerosol samples. Number of sequences affiliated to the 20 most abundant ASVs (over 862).

Supplementary Figure 1: Schematic protocol of the High-Flow rate impinger (HFRi) contamination experiment. Black arrow: sampling point.

Supplementary Figure 2: Protocol for building mock communities.

Supplementary Figure 3: ATP amount in pmol during contamination experiment with ATP. a, b: Mann-Whitney tests between blanks and concentrations after decontamination had a *p-value* > 0.1

Supplementary Table 1: Cell and DNA concentrations of the mock communities.

Bacterial strain name	Cell concentration in Mock cloud sample (cell mL^{-1})	Theoretical gene copies (mL^{-1})	DNA concentration in Mock DNA sample (ng μL^{-1})	Normalized DNA concentration (ng μL^{-1})*
***Pseudomonas syringae* 32b-74**	1.05×10^{8}	5.26×10^{8}	1.50	7.48
***Sphingomonas* sp. 32b-11**	1.05×10^{8}	4.21×10^{8}	2.34	9.35
***Rhodococcus enclensis* 23b-28**	6.61×10^{7}	2.64×10^{8}	0.17	0.68
***Bacillus* sp. 5b-1**	4.97×10^{7}	3.98×10^{8}	0.78	6.26
***Staphylococcus equorum* 5b-16**	3.45×10^{6}	2.41×10^{7}	0.82	5.77
***Flavobacterium tructae* 57b-18**	1.05×10^{8}	6.31×10^{8}	6.07	36.42

*Normalized by ribosome copy number in genome

Supplementary Table 2: Taxonomy table for the mock samples. Number of sequences affiliated to the total 22 ASVs detected in mock samples.

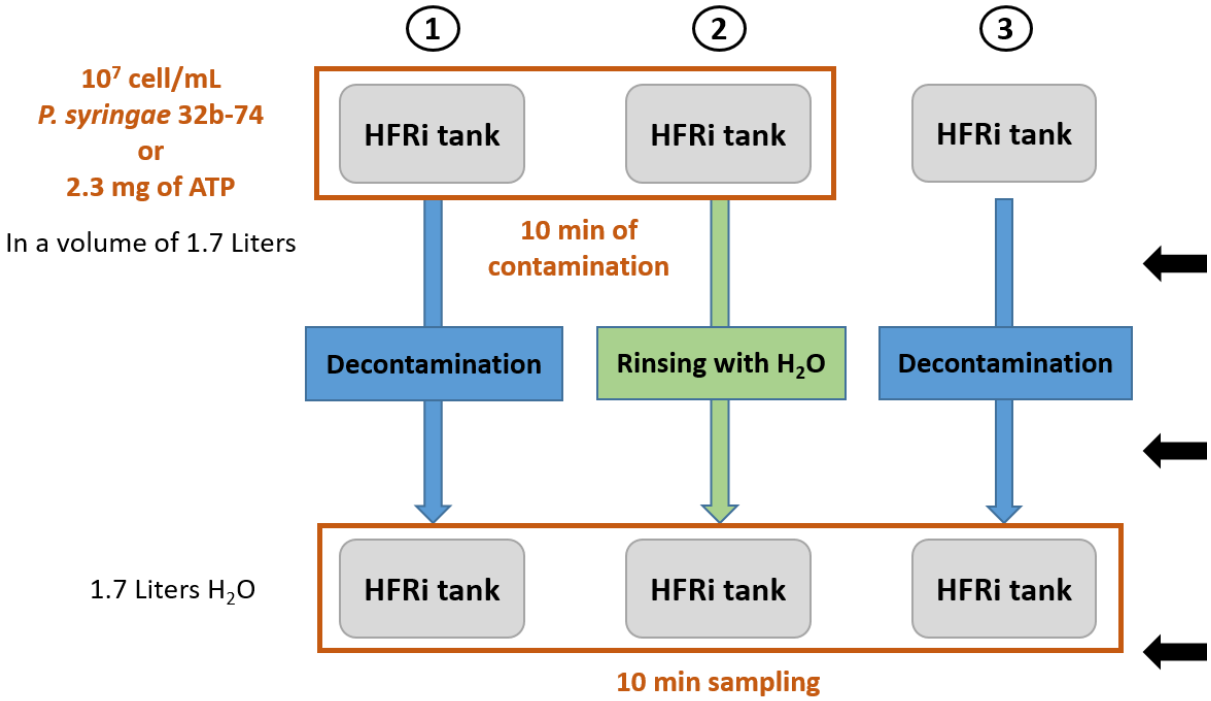
ASV	Genus	Total	MockCloud.1	MockCloud.2	MockCloud.3	MockCloud QiaAmp.1	MockCloud QiaAmp.2	MockCloud QiaAmp.3	MockDNA.1	MockDNA.2	MockDNA.3
Cluster_4	*Sphingomonas*	50267	5353	5383	5421	4416	6519	5697	5738	5837	5903
Cluster_6	*Flavobacterium*	46734	4683	4380	4479	3689	2775	3756	7591	7499	7882
Cluster_2	*Pseudomonas*	46202	5862	5746	5805	5062	5807	5388	4213	4150	4169
Cluster_5	*Staphylococcus*	44295	4390	4415	4452	7919	6781	6637	3178	3263	3260
Cluster_8	*Bacillus*	29350	1698	1768	1634	3554	2935	3176	4774	4834	4977
Cluster_13	*Rhodococcus*	18177	4206	4632	4819	1223	1299	1395	322	281	0
Cluster_30	*Flavobacterium*	9202	1019	918	777	715	581	717	1554	1565	1356
Cluster_33	*Staphylococcus*	8548	882	858	702	1509	1403	1322	680	639	553
Cluster_3	Multi-affiliation	35	5	0	0	0	0	12	7	11	0
Cluster_28	*Pelomonas*	29	0	0	0	13	0	0	8	8	0
Cluster_38	*Kocuria*	15	0	0	0	0	0	0	15	0	0
Cluster_68	*Methylobacterium-Methylorubrum*	14	0	0	0	0	0	0	14	0	0
Cluster_79	*Curtobacterium*	9	0	0	9	0	0	0	0	0	0
Cluster_94	*Acidiphilium*	5	0	0	0	0	0	0	0	5	0
Cluster_66	*Romboutsia*	5	0	0	0	0	0	0	0	5	0
Cluster_99	unknown genus	3	0	0	0	0	0	0	3	0	0
Cluster_863	*Escherichia-Shigella*	2	2	0	0	0	0	0	0	0	0
Cluster_210	*Luteimonas*	2	0	0	0	0	0	0	2	0	0
Cluster_248	*Methylobacterium-Methylorubrum*	2	0	0	2	0	0	0	0	0	0
Cluster_370	unknown genus	2	0	0	0	0	0	0	0	2	0
Cluster_137	*Acinetobacter*	1	0	0	0	0	0	0	0	1	0
Cluster_251	Multi-affiliation	1	0	0	0	0	0	0	1	0	0

Supplementary Table 3: Taxonomy table for the blank samples. Number of sequences affiliated to the 20 most abundant ASVs detected in blank samples (over 152 ASVs).

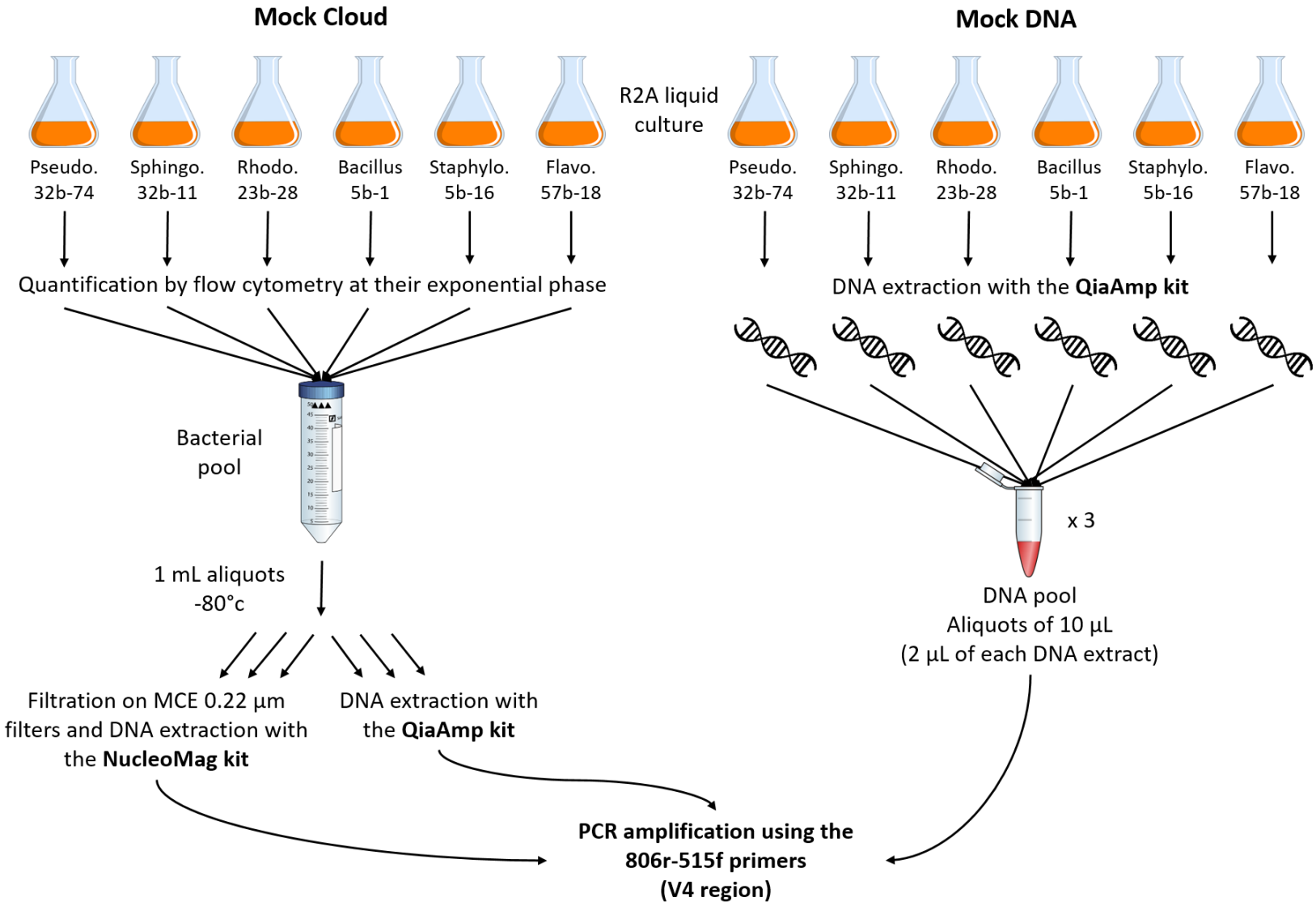
ASV	Genus	Total	SamplingBlank.1	SamplingBlank.2	SamplingBlank.3	SamplingBlank.4	WaterBlank.1	WaterBlank.2	WaterBlank.3
Cluster_4	*Sphingomonas*	1287	68	11	64	79	209	544	312
Cluster_13	*Rhodococcus*	739	0	3	0	0	489	52	195
Cluster_2	*Pseudomonas*	637	94	27	103	8	238	20	147
Cluster_44	*Blastococcus*	620	0	620	0	0	0	0	0
Cluster_28	*Pelomonas*	595	0	479	0	0	77	9	30
Cluster_6	*Flavobacterium*	508	49	8	0	4	311	19	117
Cluster_418	*Stenotrophomonas*	451	0	0	451	0	0	0	0
Cluster_64	multi-affiliation	418	0	0	418	0	0	0	0
Cluster_5	*Staphylococcus*	357	57	12	0	15	115	52	106
Cluster_322	*Geobacillus*	323	0	0	0	0	0	0	323
Cluster_174	*Burkholderia-Caballeronia-Paraburkholderia*	258	100	5	45	11	43	6	48
Cluster_68	*Methylobacterium-Methylorubrum*	254	0	0	0	156	0	98	0
Cluster_52	*Massilia*	230	0	151	28	0	0	51	0
Cluster_107	*Micrococcus*	226	223	0	0	0	0	3	0
Cluster_92	*Hymenobacter*	223	0	0	0	198	0	13	12
Cluster_89	*Enhydrobacter*	212	38	90	33	0	0	13	38
Cluster_1355	*Craurococcus-Caldovatus*	211	211	0	0	0	0	0	0
Cluster_581	*Brevundimonas*	195	98	0	69	1	12	0	15
Cluster_158	*Roseomonas*	183	0	0	0	183	0	0	0
Cluster_141	*Hymenobacter*	182	0	0	0	0	0	182	0

Supplementary Table 4: Taxonomy table for the three aerosol samples. Number of sequences affiliated to the 20 most abundant ASVs (over 862).

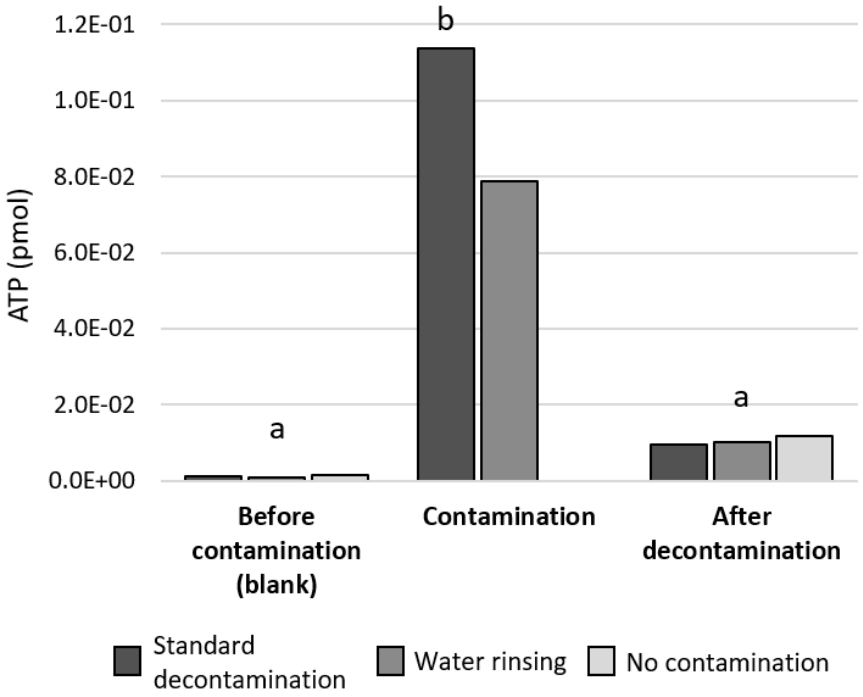
ASV	Genus	Total	20200707 AIRPDDK1	20200707 AIRPDDK2	20200707 AIRPDDK3	20200708 AIRPDDK1	20200708 AIRPDDK2	20200708 AIRPDDK3	20200709 AIRPDDK1	20200709 AIRPDDK2	20200709 AIRPDDK3
Cluster_3	multi-affiliation	6368	216	204	167	677	961	869	1069	1090	1115
Cluster_4	Sphingomonas	903	133	126	141	82	104	124	56	67	70
Cluster_31	multi-affiliation	593	83	77	59	54	75	71	55	72	47
Cluster_36	Bacillus	435	44	43	34	53	42	33	57	63	66
Cluster_57	Bacillus	303	34	32	32	27	31	32	28	48	39
Cluster_64	multi-affiliation	291	50	72	26	38	49	28	7	17	4
Cluster_32	Sphingomonas	256	33	43	26	27	34	26	24	19	24
Cluster_82	Bradyrhizobium	256	24	39	45	32	41	18	22	20	15
Cluster_99	unknown genus	248	39	44	34	26	23	26	18	13	25
Cluster_80	multi-affiliation	234	47	32	45	27	0	28	28	0	27
Cluster_122	multi-affiliation	232	30	34	32	24	24	24	12	32	20
Cluster_106	Nakamurella	191	22	17	20	19	30	23	16	20	24
Cluster_96	Methylobacterium-Methylorubrum	188	0	0	7	0	0	0	181	0	0
Cluster_147	Massilia	185	39	30	19	16	25	29	10	11	6
Cluster_102	unknown genus	184	9	9	16	38	40	20	10	21	21
Cluster_133	multi-affiliation	177	27	20	14	19	0	30	17	31	19
Cluster_69	Knoellia	169	23	18	13	23	18	23	35	16	0
Cluster_52	Massilia	166	21	27	25	17	23	12	10	17	14
Cluster_218	multi-affiliation	162	24	24	15	21	15	15	14	18	16
Cluster_179	Sphingomonas	157	17	27	15	10	22	12	13	23	18



Supplementary Figure 1: Schematic protocol of the High-Flow rate impinger (HFRi) contamination experiment. Black arrow: sampling point.



Supplementary Figure 2: Protocol for building mock communities.



Supplementary Figure 3: ATP amount in pmol during contamination experiment with ATP. a, b: Mann-Whitney tests between blanks and concentrations after decontamination had a *p-value* > 0.1

5.3. Article 2: "Comparative study of bacterial communities in clouds and aerosols", prepared to be submitted to a scientific journal

Comparative study of bacterial communities in clouds and aerosols

Raphaëlle Péguilhan[1], Florent Rossi[1], Olivier Rué[2,3], and Pierre Amato[1]

[1] Université Clermont Auvergne, CNRS, SIGMA Clermont, ICCF, F-63000 CLERMONT-FERRAND, FRANCE

[2] Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France

[3] Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France

Abstract

Aerosols and clouds have never been studied together in terms of biological content, although these both situations are key steps in the atmospheric cycling of microorganisms. Airborne bacteria can serve as cloud condensation nuclei (CCN) and thus impact cloud formation and chemistry. When embedded in clouds, bacteria can fall back to surface ecosystems with precipitation. Therefore, it is of major interest to microbial ecology and atmospheric physics to better understand what bacterial diversity exists in aerosols and clouds. We propose here a comparative study of the bacterial communities present in aerosols and clouds. For this purpose, aerosol and cloud events were collected in replicates at high altitude at the top of puy de Dôme mountain (1,465 m a.s.l) with several High-Flow-Rate impingers. Aerosols showed a higher bacterial richness than cloudy situations. They were also much more variable in terms of biodiversity and biological content (ATP, cell, and DNA concentrations) than clouds. This could be due to a seasonal effect, as aerosols were collected mainly in summer, and to the sampling that was partially performed in the planetary boundary layer, with more local and point sources of microorganism emission. The cloud events were much more homogeneous, perhaps highlighting specific phenomena or selective pressures controlling the structure of bacterial communities. Finally, the impact of condensed water situations on bacterial diversity was not conclusive here and further investigations should be conducted regarding bacterial activity in these two atmospheric situations.

Keywords

Clouds, aerosols, bacterial diversity, seasonal effect

Introduction

Bacteria are present in the atmosphere up to high altitude and in the clouds (Amato et al., 2007; Després et al., 2012; Sattler et al., 2001). They originate from multiple sources including vegetation, soil etc., whose airborne microbial emission plumes are mixed and transported regionally and continentally (Smith et al., 2013). The airborne microbiota is therefore composed of diverse and variable assemblages depending on the heterogeneity and variations of sources, such as those caused by surface occupation and use of surfaces (oceanic, rural, urban, agricultural, etc.) (Bowers et al., 2011; Burrows et al., 2009), seasonality (Tignat-Perrier et al., 2020), or meteorology (turbulent fluxes, precipitation etc.) (Butterworth and McCartney, 1992; Carotenuto et al., 2017; Joung et al., 2017).

Relative humidity (rH) in the air is defined by water vapour content and temperature. Aerosol particles attract water vapour and eventually reach a deliquescence point, becoming partially liquid even at rH <100% (Ge et al., 1998; Tang and Munkelwitz, 1993). When rH increases to values exceeding 100% (supersaturation), for example with altitude, cloud droplets are formed due to condensation of water vapour on the surface of particles. The propensity of aerosol particles to condense water depends on their size (Kelvin effect) and hygroscopicity/solubility (Raoult effect). Largest and more soluble aerosols are activated as cloud condensation nuclei (CCN) at lower supersaturation, so earlier in cloud's lifetime, than small hydrophobic insoluble particles (Koehler principle; Koehler *et al.*, 2006; Wex *et al.*, 2008). Because of their large size (~1 µm) and surface properties, airborne bacteria are considered excellent CCNs (Möhler et al., 2007), so they are likely to be among the first particles to be incorporated into the liquid phase of clouds when they form (Bauer et al., 2003; Lazaridis, 2019; Möhler et al., 2007).

Clouds are thought to provide favorable conditions for airborne living microbes as compared with dry situations: condensed water protects cells from desiccation and direct

sunlight, and it also potentially facilitates access to the nutrients solubilized from aerosols particles and gas (Šantl-Temkiv et al., 2022). In clouds, bacterial cells impact microphysics and chemistry through properties of ice nucleation (IN) (Morris et al., 2004) and metabolic activity (Ariya and Amyot, 2004; Khaled et al., 2021; Krumins et al., 2014; Wirgot et al., 2017), and they eventually could multiply (Ervens and Amato, 2020; Fuzzi et al., 1996; Sattler et al., 2001). Additionally clouds largely contribute to the deposition of high-altitude aerosols including bacteria though precipitation (Aho et al., 2019; Morris et al., 2004; Péguilhan et al., 2021; Triadó-Margarit et al., 2019; Woo and Yamamoto, 2020). These introduce distant microbial diversity into surface ecosystems, with measurable impacts (Jalasvuori, 2020; Noirmain et al., 2022). The distribution of bacterial diversity in aerosols and clouds is therefore a key element to study in the atmospheric cycle of bacteria. However, it is not clear yet whether microbiological characteristics are affected by clouds.

To date, comparative analyses between the wet and dry phases of the atmosphere have largely focused on chemistry (Mekic and Gligorovski, 2020; Möhler et al., 2007; Sellegri et al., 2003). Few studies examined bacterial diversity in atmospheric deposits, and demonstrated slight differences between wet and dry processes (e.g., Els et al., 2019; Triadó-Margarit et al., 2019). Here we aim to comparatively investigate airborne biological bacteria in the wet (i.e., clouds) and dry phases of the atmosphere, at a single site and using similar methods, which to our knowledge have never been assessed. Differences may relate with taxa-dependent propensity of cells to act as CCNs and/or IN, resist stresses and maintain integrity facing osmotic and freeze-thaw shocks (Joly et al., 2015), or eventually multiply (Ervens and Amato, 2020; Fuzzi et al., 1996). More than 50 samples of clouds and aerosols were collected at 19 dates throughout the year from the high-altitude atmospheric station of puy de Dôme mountain (1,465 m a.s.l.) using several high-flow rate impingers deployed in parallel as replicates. Microbial biomass, viability and biodiversity were examined by flow cytometry, ATP quantification and amplicon sequencing, respectively. The results indicate that microbial characteristics are much less variable in clouds than in the dry atmosphere, suggesting a

convergence linked with the presence of condensed water. Some taxa were found significantly more abundant in clouds (e.g., *Enhydrobacter*, *Kocuria*, *Staphylococcus*), and vice versa (e.g., *Bacillus*). But the overall diversity of bacteria could not be clearly discriminated between the two circumstances and completely disentangled from, notably, seasonal variations.

Materials and methods

Sampling procedure

From autumn 2019 to autumn 2020, 19 atmospheric samples were collected, including 11 aerosol events and 8 cloudy events, representing a total of 53 environmental samples. Throughout the manuscript, we refer to sampling dates as "events", and to sampling replicates as "samples".

Samples were collected at the top of puy de Dôme Mountain (PUY; 1,465 m a.s.l., 45.772° N, 2.9655° E; France) and, at one occasion at the Opme station located in the plain underneath PUY (OPM; 680 m a.s.l., 45.7125° N, 3.090278° E; France) (Baray et al., 2020; Péguilhan et al., 2021). The geographical origin of the air masses sampled and the fraction of time spent over sea or land at high (> 1 km) or low (< 1 km) altitude over the three days preceding sampling were recovered from 72-hours backward trajectory plots, computed using the CAT trajectory model (Baray et al., 2020). In addition, each event was qualified as "In" or "Out" the boundary layer, as in Péguilhan *et al.* (2021), based on ECMWF ERA5 data (Hoffmann et al., 2019). More details are provided as **supplementary material and Péguilhan *et al.* (Article 1).** Variables pertaining to sample collection are presented in **Table 1 and supplementary Figure 1.** Cloudy events occurred mostly during fall, in the free troposphere, and from air masses originating from West (Atlantic Ocean). Aerosols were collected during summer and fall, partially in the boundary layer and from air masses originating from West and North.

For sampling, 3 to 5 High-Flow-Rate impingers (HFRi) (DS6, Kärcher SAS, Bonneuil sur Marne, France) (Šantl-Temkiv et al., 2017) were ran in parallel for ~2 to ~6h, corresponding to ~360-720 cubic meters of air based on instrument specifications. HFRi were filled with either 850 mL or 1,700 mL of collection liquid for cloudy and dry situations, respectively (clouds are defined as situations where liquid water content was > 0 g/m3, based on real-time measurements at the PUY station). One of the samplers was filled with autoclaved ultrapure water for total cell counts and ATP quantification, from triplicate subsamples of 450 and 50 µL fixed with 0.5% glutaraldehyde and B/S extractant, respectively. The others were filled with nucleic acid preservation (NAP) buffer for DNA analyses and treated as replicates; here the entire volumes of liquids were filtered independently through 0.22 µm porosity (mixed ester cellulose filters, 47 mm) to recover microorganisms. DNA was extracted from filters using NucleoMag® DNA/RNA Water extraction kit (Macherey-Nagel, Hoerdt, France), subjected to metabarcoded PCR targeting the V3-V4 region of the 16S rRNA gene of bacteria, and sequencing (Illumina Miseq 2*250bp). More details on the methods are given as **supplementary material**.

The presence of eventual contaminants in the collection liquids were checked before and after exposure to the samplers, with no aspiration. We detected negligible numbers of sequences (<630 total reads per ASV) and richness (152 ASVs against 862 in environmental samples), and no core contaminant. Detailed information is provided in **Péguilhan et al (Article 1)**.

Table 1: Samples main information

Sample name	Sampling date (dd/mm/yyyy)	Season	Sampling site	Main geographical origin[±]	Sampling duration (h)	In/out PBL	Temperature (°C)[#]	Relative humidity (%)[#]	Wind speed (m s^{-1})[#]
AEROSOLS									
20190712AIRPDD	12/07/2019	Summer	PUY	W	2.0	NA*	14.8	62	5.8
20190918AIRPDD	18/09/2019	Summer	PUY	NW	2.1	Out	11.3	67	4.4
20200206AIRPDD	06/02/2020	Winter	PUY	N	4.5	Out	4.7	20	6.0
20200518AIROPME	18/05/2020	Spring	OPME	NA*	5.1	In	15.2	24	NA*
20200610AIRPDD	10/06/2020	Spring	PUY	N	4.3	In	6.3	90	1.7
20200707AIRPDD	07/07/2020	Summer	PUY	NW	6.5	In	11.1	61	3.6
20200708AIRPDD	08/07/2020	Summer	PUY	NW	6.1	In	14.2	53	3.1
20200709AIRPDD	09/07/2020	Summer	PUY	N	6.0	In	20.3	48	3.4
20200922AIRPDD	22/09/2020	Fall	PUY	W	5.9	Out	12.4	78	1.0
20201118AIRPDD	18/11/2020	Fall	PUY	W	5.8	Out	14.1	41	6.4
20201124AIRPDD	24/11/2020	Fall	PUY	W	6.0	Out	8.6	50	3.4
Min	-	-	-	-	**2.0**	-	**4.7**	**20**	**1.0**
Max	-	-	-	-	**6.5**	-	**20.3**	**90**	**6.4**
Median	-	-	-	-	**5.8**	-	**12.4**	**53**	**3.5**
Mean	-	-	-	-	**4.9**	-	**12.1**	**54**	**3.9**
Standard error	-	-	-	-	**1.6**	-	**4.4**	**21**	**1.8**
CLOUDS									
20191022CLOUD	22/10/2019	Fall	PUY	S	6.4	Out	5.7	100	8.7
20200311CLOUD	11/03/2020	Winter	PUY	W	4.1	Out	5.0	100	7.4
20200717CLOUD	17/07/2020	Summer	PUY	NW	3.3	Out	10.1	100	1.6
20201016CLOUD	16/10/2020	Fall	PUY	NE	4.7	Out	1.1	100	1.8
20201028CLOUD	28/10/2020	Fall	PUY	W	6.0	Out	5.2	100	11.0
20201103CLOUD	03/11/2020	Fall	PUY	W	3.5	In	2.2	100	8.7
20201110CLOUD	10/11/2020	Fall	PUY	SW	3.1	Out	5.9	100	2.5
20201119CLOUD	19/11/2020	Fall	PUY	W	2.8	Out	0.3	100	7.7
Min	-	-	-	-	**2.8**	-	**0.3**	**100**	**1.6**
Max	-	-	-	-	**6.4**	-	**10.1**	**100**	**11.0**
Median	-	-	-	-	**3.8**	-	**5.1**	**100**	**7.5**
Mean	-	-	-	-	**4.2**	-	**4.4**	**100**	**6.2**
Standard error	-	-	-	-	**1.4**	-	**3.2**	**0**	**3.7**

[±] From air masses backward trajectory plots over 72 h preceding sampling **(**https://www.opgc.fr/data-center/public/data/copdd/trajectory);

[#] **Average over the sampling period;**

NA*: No data available;

PBL: Planetary Boundary Layer

Bioinformatics and data analysis

Sequences were processed as in **Article 1** for taxonomic affiliation from amplicon sequence variants (ASVs) using dada2 package (v 1.20.0) (Callahan et al., 2016) and FROGS (Bernard et al., 2021; Escudié et al., 2018) against SILVA v138.1 (Quast et al., 2013). Details on the methods are given as **supplementary material.** Briefly, a total of 2,770 sequences from each sample were used for analyses, corresponding to 862 amplicon sequence variants (ASV). One of the replicate sample for aerosols (20190712AIRPDDK1) exhibited low number of sequences (424) so this event was removed from analysis of diversity.

Results and Discussion

Meteorological context

Clouds were sampled mostly during fall and winter, and dry atmosphere during spring and summer. Consistently, ambient temperatures were colder during cloud sampling than during aerosol collection (Mann-Whitney test, *p-value* = 0.003) (**Table 1**). Dry aerosols were partially sampled in the planetary boundary layer while cloudy situations were mainly collected in the free troposphere. Based on 72-hours backward trajectories, the air masses for both situations spent most of their time at high altitude (> 1 km a.g.l.) over continental or marine environments before reaching the sampling site (**Supplementary Figure 1**).

Biomass, ATP and DNA contents

Cell and DNA concentrations tended to be lower, although not significantly different, during cloudy than during dry situations, with means of 1.62×103 ± 1.89×103 and 3.37×103 ± 2.92×103 cell.m-3, and 0.28 ± 0.08 and 1.01 ± 1.35 ng DNA.m-3, respectively (**Figure 1; Table 2**). ATP concentration was also significantly lower in the clouds by a factor of ~28 on average (0.15 ± 0.11 pmol m-3 in clouds vs 4.22 ± 3.27 pmol m-3 in aerosols). For ATP and DNA

concentration, dry situations were much more variable than clouds, with coefficients of variations of 77.5 and 134.5 versus 71.2 and 30.7 respectively. This can be explained by the wider range of meteorological situations investigated, notably regarding boundary layer height and humidity (**Table 1**). Higher ATP per cell ratios in aerosols (2.14×10-5-1.03×10-2 pmol.cell-1) than in clouds (1.20×10-5-1.52×10-3 pmol.cell-1) suggests a higher viability and potential biological activity. In a previous study (Péguilhan et al., 2021), ATP-to-cell ratio was found significantly higher in cloud water compared to precipitation, suggesting lower values in the aerosols scavenged. Here differences of temperature between clouds and dry conditions (colder in clouds) could explain the differences.

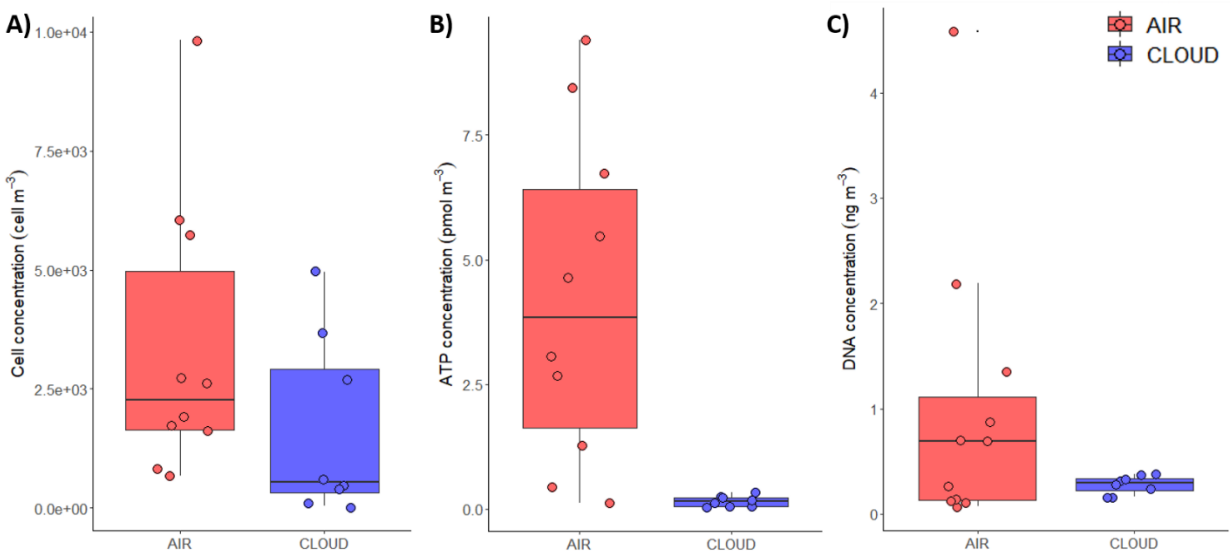


Figure 1: Cell (A), ATP (B), and DNA (C) concentrations in the presence and in the absence of cloud.

Table 2: Events biological information.

Sample name	Microbial cell number concentration (cells m^{-3} of air)[#]	ATP concentration (pmol m^{-3} of air)[#]	DNA Average (ng m^{-3} of air)[±]	ASV number average
AEROSOLS				
20190712AIRPDD	9.82×10^{3}	4.63	0.07	NA[$]
20190918AIRPDD	6.05×10^{3}	0.44	0.11	298
20200206AIRPDD	NA[$]	NA[$]	0.26	133
20200518AIROPME	8.21×10^{2}	8.45	4.58	56
20200610AIRPDD	6.82×10^{2}	2.66	0.69	198
20200707AIRPDD	1.62×10^{3}	6.72	2.19	285
20200708AIRPDD	1.72×10^{3}	3.06	1.35	239
20200709AIRPDD	2.62×10^{3}	9.40	0.87	224
20200922AIRPDD	1.91×10^{3}	5.47	0.70	80
20201118AIRPDD	2.72×10^{3}	1.27	0.14	173
20201124AIRPDD	5.72×10^{3}	0.12	0.12	370
Minimum	**6.82×10^{2}**	**0.12**	**0.07**	**56**
Maximum	**9.82×10^{3}**	**9.40**	**4.58**	**370**
Median	**2.26×10^{3}**	**3.84**	**0.69**	**211**
Mean	**3.37×10^{3}**	**4.22**	**1.01**	**205**
Standard error	**2.92×10^{3}**	**3.27**	**1.35**	**99**
CLOUDS				
20191022CLOUD	4.95×10^{3}	0.06	0.24	136
20200311CLOUD	4.75×10^{2}	0.04	0.31	73
20200717CLOUD	1.20×10^{2}	0.18	0.33	125
20201016CLOUD	2.67×10^{3}	0.24	0.16	95
20201028CLOUD	3.67×10^{3}	0.33	0.16	130
20201103CLOUD	6.17×10^{2}	0.22	0.37	120
20201110CLOUD	3.90×10^{2}	0.12	0.38	171
20201119CLOUD	3.66×10^{1}	0.02	0.28	100
Minimum	**3.66×10^{1}**	**0.02**	**0.16**	**73**
Maximum	**4.95×10^{3}**	**0.33**	**0.38**	**171**
Median	**5.46×10^{2}**	**0.15**	**0.29**	**122**
Mean	**1.62×10^{3}**	**0.15**	**0.28**	**119**
Standard error	**1.89×10^{3}**	**0.11**	**0.08**	**30**
P-value (Mann-Whitney test)	**0.08**	**0.002****	**0.59**	**0.045***

NA[$]: no data available;
[#] concentrations per m^3 of air were calculated based on the duration of sampling and the sampler air-flow rate for aerosols, and based on liquid water content (LWC) for clouds;
[±] concentrations per m^3 of air were calculated based on the duration of sampling and the HFR impinger air-flow rate for both aerosols and clouds;
* p-value significant ($0.05 > \text{p-value} \geq 0.01$);
** p-value highly significant ($0.01 > \text{p-value}$).

Biodiversity in the wet and dry phases of the atmosphere

Also with respect to biodiversity, clouds were more similar than dry situations (**Supplementary Figures 2 and 3**), and bacteria richness was higher during dry than during cloudy situations (**Table 2**; **Supplementary Figure 4**). Although 85 % of the bacterial genera were shared between both atmospheric situations (**Figure 2**). The low variability among cloud events along with the overrepresentation of certain taxa in one or the other situation suggest possible underlying phenomena influencing bacteria diversity in clouds and leading them to converge.

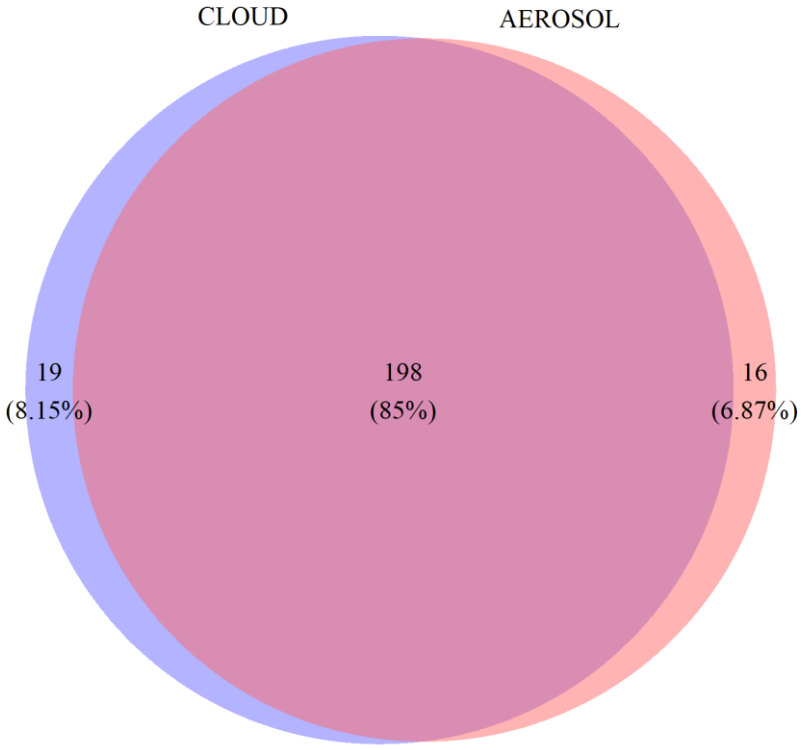


Figure 2: Distribution of 233 bacterial genera between clouds and aerosols.

Samples were collected under multiple conditions, resulting, as expected, in inter-event variability of bacteria diversity due to the influences of numerous environmental factors. True sampling replicates (by opposition to technical replicates, where a single sample is subsampled for replicated analyses) are necessary for accounting for intra-event variation and allow for more powerful statistical analyses. The replicates of most events (12 out of 18) were more similar to each other than to other events in terms of biodiversity, highlighting the high reproducibility of sampling with several samplers displayed in parallel at a single sampling site.

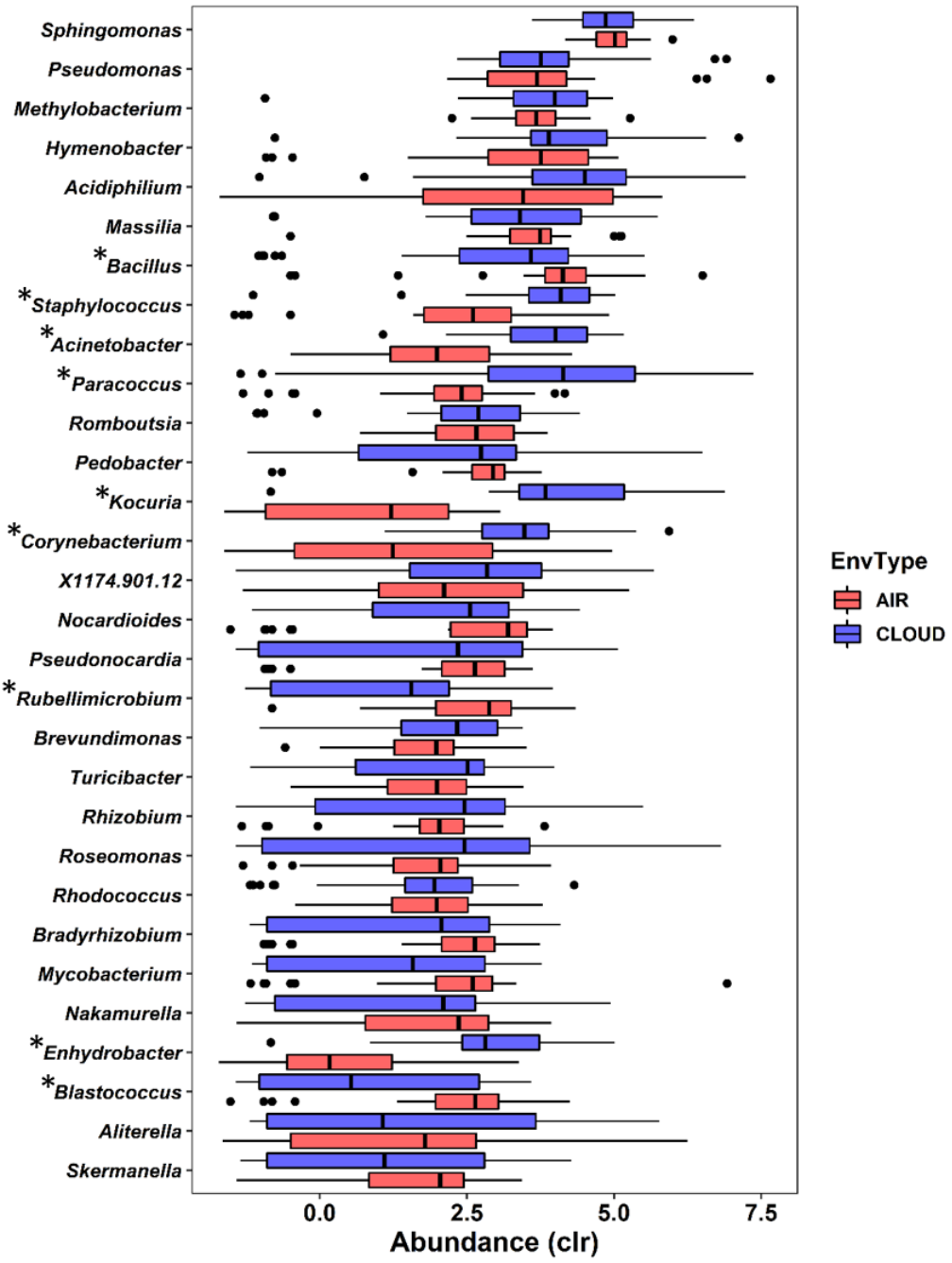
The most abundant taxa (> 100 reads) were consistently retrieved between replicates, with low variability (CV< 1), in particular during cloudy situations (**Supplementary Figure 5**). In turn, rarest taxa's abundances were by essence more variable between replicates, due to necessary increased heterogeneity and larger impacts of eventual analytic bias. Therefore, for consistency, we focused our analysis on abundant groups common to both situations.

Both cloud and aerosol samples harbored bacterial phyla often reported as dominant in the atmosphere over continental areas, such as Proteobacteria, Actinobacteria, Firmicutes and Bacteroidetes (over 17 phyla in total) (Amato et al., 2017; Bowers et al., 2013; Tignat-Perrier et al., 2020). The genera *Sphingomonas*, *Methylobacterium* (Alphaproteobacteria), *Pseudomonas* (Gammaproteobacteria), *Massilia* (Betaproteobacteria), *Hymenobacter* (Bacteroidetes) and *Bacillus* (Firmicutes) were among the most abundant (**Supplementary Table 2**). These are generally reported in association with soil and plants (Barberán et al., 2015; Bowers et al., 2011; Lindemann et al., 1982).

The relative abundances of epiphytes such as *Sphingomonas*, *Pseudomonas*, and *Massilia* were not different between wet and dry situations (**Figure 3**). These are widespread taxa in the atmosphere all over the globe (Després et al., 2012).

Several genera were significantly more represented during cloudy situations, such as *Staphylococcus*, *Acinetobacter*, *Paracoccus*, *Kocuria*, *Corynebacterium*, *Enhydrobacter*,

220 *Streptococcus* and *Aerococcus* (Kruskal-Wallis test, *p-value* <0.05; **Figure 3**; **Supplementary**
221 **Table 3**). Conversely, *Bacillus*, *Rubellimicrobium*, and *Blastococcus* were more abundant
222 under dry conditions.



223

224 **Figure 3: The 30 most abundant bacterial genera in clouds and aerosols ordered by total**
225 **abundance from top to down.** Scale is represented in centered-log ratio (clr). EnvType:
226 Environment type; *: the genus is significantly differently abundant between cloud and aerosol,
227 kruskal-Wallis test between clouds and aerosols, *p-value* <0.05.

Clouds provide an aqueous environment for cells and maybe perhaps better conditions for survival and development, if osmotic shocks are not detrimental. Given the short lifetime of clouds and the short residence time in the atmosphere for bacteria cells (~3 days; Burrows et al., 2009), it was estimated that bacterial multiplication in cloud droplets should not significantly affect community' structure (Ervens and Amato, 2020). Rather, the differences could be due to the selective exclusion of certain taxa in relation to physical and biological factors.

Osmotic shock and freeze-thaw cycles at high rH and in the presence of condensed water could have disrupted a significant portion of the cells (Christian and Ingram, 1959; Gutierrez et al., 1995). In our dataset, *Staphylococcus* and *Corynebacterium* were more represented in clouds than in the dry atmosphere. Interestingly these were previously reported to be also significantly more abundant in clouds at the same site compared to the associated precipitation (Péguilhan *et al*., 2021). *Staphylococcus* is known to have high tolerance to osmotic shocks (Graham and Wilkinson, 1992; Gutierrez et al., 1995), which could have contributed to maintain cell integrity during cloud formation. In turn, *Bacillus* was more present in aerosols than in clouds here, and was previously reported more abundant in clouds than in the precipitation below (Péguilhan et al., 2021). This relates with its widespread presence at high altitudes (Jaing et al., 2020; Smith et al., 2013). *Bacillus* is a bacterium known to form endospores that can indeed endure extreme conditions. Previous reports indicated its high abundance throughout the year at this sampling site (Tignat-Perrier et al., 2020), but the reason for its relative depletion in clouds compared with dry situations here is unclear. This may relate with higher emissions from dryer soils, a likely strong source of *Bacillus,* and/or propensity to be subjected early to wet deposition due to large cell size (Knaysi, 1929). Aerosols, of which airborne bacteria, serve as CCN, whose efficiency depends on their size and hygroscopicity (Asmi et al., 2012; Petters and Kreidenweis, 2007). Bacteria have a size of ~0.1 to 2 µm and

253 are considered efficient CCN (Andreae and Crutzen, 1997; Bauer et al., 2003), with no
254 distinction between taxa. In theory a particle in that size range cannot avoid water condensation
255 when the air is supersaturated with water given Koehler's theory, so similar diversity should be
256 found based solely on such physical aspects. Nevertheless, it is likely that larger cells are more
257 prone to deposit during wet situations, as they already sediment more rapidly than smaller cells
258 under dry conditions. Their abundance decreases when the sources are turned down due to
259 elevated humidity, which limits dust emissions.

260 Hydrophobicity of cells (Dahlbäck et al., 1985) or its associated organic/inorganic
261 compounds could also be a factor of partitioning for the integration of cloud droplets. However,
262 a high hydrophobicity of the particle will only delay the condensation of water on it and might
263 is not sufficient to avoid the formation of a cloud droplet when the air becomes more strongly
264 supersaturated with water (Deleon-Rodriguez, 2015; Van Der Mei et al., 1998; Ogawa et al.,
265 2016).

266

267 **Seasonal variations**

268 We focused our analysis on the impact of the presence/absence of condensed water, which
269 defines clouds. Several factors are often intimately related to each other, such as in this case the
270 season and the presence of clouds, which also goes hand in hand with temperature, and thus
271 their respective influences could not be clearly distinguished in our data set. Concomitant
272 situations were indeed investigated between meteorological situations and season, *i.e.*, clouds
273 were more frequently sampled during the fall and winter, and the dry atmosphere during spring
274 and summer (**Table 1**). Bacteria diversity was much higher during summer than during fall and
275 winter (**Figure 4**), as already observed at this site (Tignat-Perrier et al., 2020), probably in
276 relation with sources (vegetation) and boundary layer altitude. Nevertheless, clouds collected

during summer resembled more other clouds than dry aerosols collected during the same season (**Figure 5**), still suggesting some extent of determinism of the presence of liquid water.

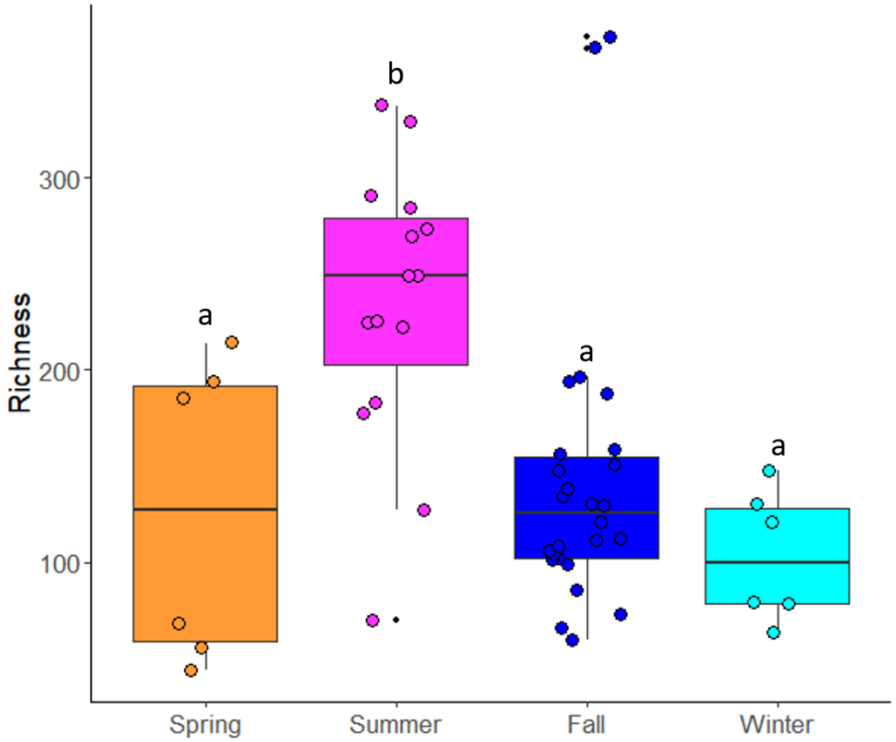


Figure 4: Seasonal effect on ASV distribution between samples. Letters represent pairwise Mann-Whitney results. ASV: amplicon sequence variant.

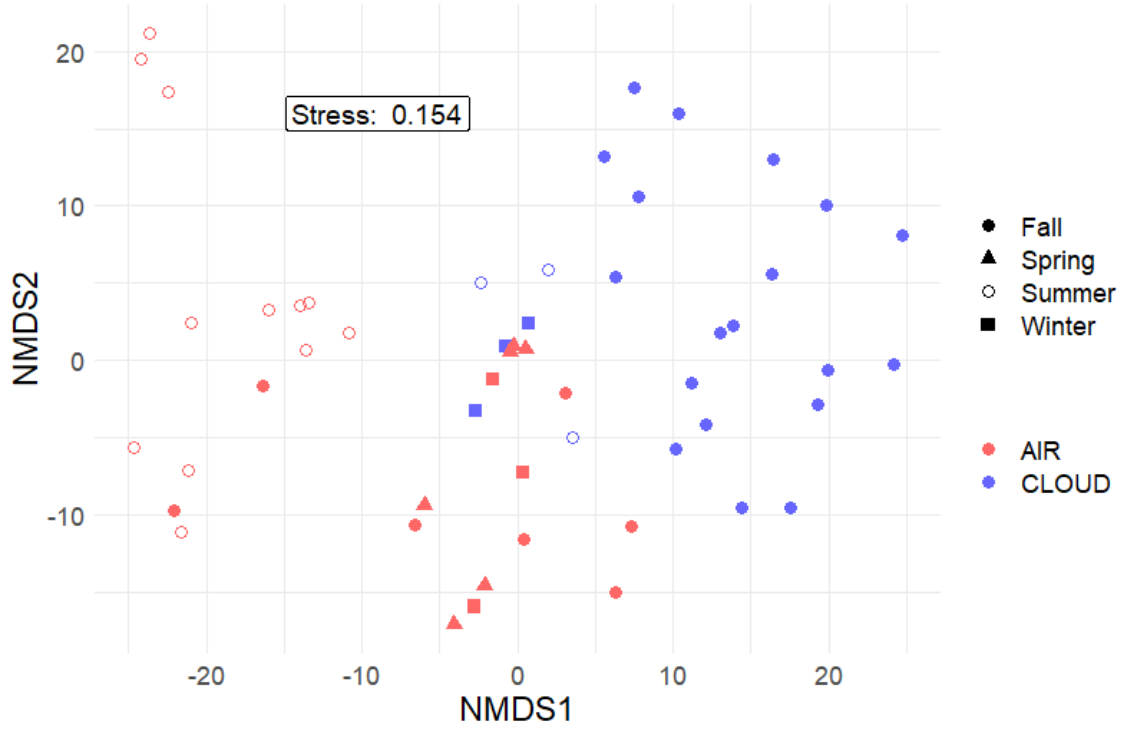


Figure 5: Non-metric multidimensional scaling (NMDS) analysis of the biodiversity in clouds and dry atmosphere, based on the abundance of the total 862 ASVs (Euclidean distance). All sample replicates are represented.

Concluding remarks

Several significant variations of biological content in the atmosphere were detected between the wet and dry situations. However, these differences can be explained by seasonal variations, as studied by Tignat-Perrier *et al.* (2020) at the same sampling site. The low variability between clouds points to underlying phenomena converging bacterial diversity. Selective processes such as osmotic shocks and freeze-thaw cycles could play roles. Further investigations will be needed to clarify this aspect.

Bacterial cells were demonstrated to be better adapted to the drying-rewetting cycle when they undergo several successive cycles in soil (Leizeaga et al., 2022), which also occurs in the atmosphere. It is estimated that water undergoes 10 to 11 condensation-evaporation cycles

300 before precipitating (Pruppacher and Jaenicke, 1995), so it is conceivable that bacteria also are
301 exposed to several cloud cycles before precipitating. Cloud processing could select and "train"
302 cells to respond rapidly to changing conditions and the presence of liquid water, and prepare
303 them to colonize new surfaces as active cells prior being deposited with precipitation, which
304 can be seen as beneficial.
305

306 ## Acknowledgment

307 We thank L. Deguillaume, P. Renard, M. Brissy and C. Ghaffar for help in the field, JL.
308 Baray for the meteorological data and backward trajectory analysis, F. Enault for bioinformatic
309 support and B. Ervens for advice. We are thankful to the Aubi platform and to the Mésocentre
310 Clermont Auvergne for providing support, computing and storage resources and to the
311 ECMWF's computing and archive facilities. The sampling has been performed using the CO-
312 PDD instrumented sites of the OPGC observatory and LaMP laboratory. This research has been
313 supported by the French National Research Agency (ANR) (grant no. ANR-17-MPGA-0013),
314 Université Clermont Auvergne, CNRS-INSU and CNES.

References

Aho, K., Weber, C.F., Christner, B.C., Vinatzer, B.A., Morris, C.E., Joyce, R., Failor, K., Werth, J.T., Bayless-Edwards, A.L.H., and Schmale III, D.G. (2019). Spatiotemporal patterns of microbial composition and diversity in precipitation. Ecol. Monogr. *0*, 1–26.

Amato, P., Parazols, M., Sancelme, M., Laj, P., Mailhot, G., and Delort, A.M. (2007). Microorganisms isolated from the water phase of tropospheric clouds at the Puy de Dôme: Major groups and growth abilities at low temperatures. In FEMS Microbiology Ecology, pp. 242–254.

Amato, P., Joly, M., Besaury, L., Oudart, A., Taib, N., Moné, A.I., Deguillaume, L., Delort, A.M., and Debroas, D. (2017). Active microorganisms thrive among extremely diverse communities in cloud water. PLoS One *12*, 1–22.

Andreae, M.O., and Crutzen, P.J. (1997). Atmospheric Aerosols: Biogeochemical Sources and Role in Atmospheric Chemistry. Science (80-.). *276*, 1052–1058.

Ariya, P.A., and Amyot, M. (2004). New Directions: The role of bioaerosols in atmospheric chemistry and physics. Atmos. Environ. *38*, 1231–1232.

Asmi, E., Freney, E., Hervo, M., Picard, D., Rose, C., Colomb, A., and Sellegri, K. (2012). Aerosol cloud activation in summer and winter at puy-de-Dôme high altitude site in France. Atmos. Chem. Phys. *12*, 11589–11607.

Baray, J.L., Deguillaume, L., Colomb, A., Sellegri, K., Freney, E., Rose, C., Baelen, J. Van, Pichon, J.M., Picard, D., Fréville, P., et al. (2020). Cézeaux-Aulnat-Opme-Puy de Dôme: A multi-site for the long-term survey of the tropospheric composition and climate change. Atmos. Meas. Tech. *13*, 3413–3445.

Barberán, A., Ladau, J., Leff, J.W., Pollard, K.S., Menninger, H.L., Dunn, R.R., and Fierer, N. (2015). Continental-scale distributions of dust-associated bacteria and fungi. PNAS *112*, 5756–5761.

Bauer, H., Giebl, H., Hitzenberger, R., Kasper-Giebl, A., Reischl, G., Zibuschka, F., and Puxbaum, H. (2003). Airborne bacteria as cloud condensation nuclei. J. Geophys. Res. Atmos. *108*, 4658.

Bernard, M., Rué, O., Mariadassou, M., and Pascal, G. (2021). FROGS: a powerful tool to analyse the diversity of fungi with special management of internal transcribed spacers. Brief. Bioinform. *22*.

Bowers, R.M., McLetchie, S., Knight, R., and 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.

Bowers, R.M., Clements, N., Emerson, J.B., Wiedinmyer, C., Hannigan, M.P., and Fierer, N. (2013). Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. Environ. Sci. Technol. *47*, 12097–12106.

Burrows, S.M., Elbert, W., Lawrence, M.G., and 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.

Butterworth, J., and McCartney, A. (1992). The Removal and Dispersal of Foliar Bacteria by Rain Splash. Release Genet. Modif. Microorg. 2 187–189.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 2016 137 *13*, 581–583.

Carotenuto, F., Georgiadis, T., Gioli, B., Leyronas, C., Morris, C.E., Nardino, M., Wohlfahrt, G., and Miglietta, F. (2017). Measurements and modeling of surface-atmosphere exchange of microorganisms in Mediterranean grassland. Atmos. Chem. Phys. *17*, 14919–14936.

Christian, J.H., and Ingram, M. (1959). Lysis of Vibrio costicolus by osmotic shock. J. Gen. Microbiol. *20*, 32–42.

Dahlbäck, B., Hermansson, M., Kjelleberg, S., and Norkrans, B. (1985). The Hydrophobieity of Bacteria-An Important Factor in Their Initial Adhesion at the Air-Water Interface (Springer-Verlag).

Deleon-Rodriguez, N. (2015). Microbes in the Atmosphere : Prevalence , Species Composition , and Relevance To Cloud Formation Copyright © 2015 Natasha Deleon-Rodriguez Microbes in the Atmosphere : Prevalence , Species Composition , and Relevance To.

Després, V.R., Alex Huffman, J., Burrows, S.M., Hoose, C., Safatov, A.S., Buryak, G., Fröhlich-Nowoisky, J., Elbert, W., Andreae, M.O., Pöschl, U., et al. (2012). Primary biological aerosol particles in the atmosphere: A review. Tellus, Ser. B Chem. Phys. Meteorol. *64*.

Els, N., Larose, C., Baumann-Stanzer, K., Tignat-Perrier, R., Keuschnig, C., Vogel, T.M., and Sattler, B. (2019). Microbial composition in seasonal time series of free tropospheric air and precipitation reveals community separation. Aerobiol. 2019 354 *35*, 671–701.

Ervens, B., and Amato, P. (2020). The global impact of bacterial processes on carbon mass. Atmos. Chem. Phys. Discuss. *20*, 1–25.

Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., Maman, S., Hernandez-Raquet, G., Combes, S., and Pascal, G. (2018). FROGS: Find, Rapidly, OTUs with Galaxy Solution. Bioinformatics *34*, 1287–1294.

Fuzzi, S., Mandrioli, P., and Perfetto, A. (1996). Fog droplets - An atmospheric source of secondary biological aerosol particles. Atmos. Environ. *31*, 287–290.

Ge, Z., Wexler, A.S., and Johnston, M. V (1998). Deliquescence behavior of multicomponent aerosols. J. Phys. Chem. A *102*, 173–180.

Graham, J.E., and Wilkinson, B.J. (1992). Staphylococcus aureus osmoregulation: Roles for choline, glycine betaine, proline, and taurine. J. Bacteriol. *174*, 2711–2716.

Gutierrez, C., Abee, T., and Booth, I.R. (1995). Physiology of the osmotic stress response in microorganisms. Int. J. Food Microbiol. *28*, 233–244.

Hoffmann, L., Günther, G., Li, D., Stein, O., Wu, X., Griessbach, S., Heng, Y., Konopka, P., Müller, R., Vogel, B., et al. (2019). From ERA-Interim to ERA5: The considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. Atmos. Chem. Phys. *19*, 3097–3214.

Jaing, C., Thissen, J., Morrison, M., Dillon, M.B., Waters, S.M., Graham, G.T., Be, N.A., Nicoll, P., Verma, S., Caro, T., et al. (2020). Sierra Nevada sweep: metagenomic measurements of bioaerosols vertically distributed across the troposphere. Sci. Rep. *10*, 12399.

Jalasvuori, M. (2020). Silent rain: does the atmosphere-mediated connectivity between microbiomes influence bacterial evolutionary rates? FEMS Microbiol. Ecol. *96*, 96.

Joly, M., Amato, P., Sancelme, M., Vinatier, V., Abrantes, M., Deguillaume, L., and Delort, A.M. (2015). Survival of microbial isolates from clouds toward simulated atmospheric stress factors. Atmos. Environ. *117*, 92–98.

Joung, Y.S., Ge, Z., and Buie, C.R. (2017). Bioaerosol generation by raindrops on soil. Nat. Commun. *8*, 1–10.

Khaled, A., Zhang, M., Amato, P., Delort, A.M., and Ervens, B. (2021). Biodegradation by bacteria in clouds: An underestimated sink for some organics in the atmospheric multiphase system. Atmos. Chem. Phys. *21*, 3123–3141.

Knaysi, G. (1929). The Cell Structure and Cell Division of Bacillus subtilis. J. Bacteriol. *19*, 113–115.

Koehler, K.A., Kreidenweis, S.M., DeMott, P.J., Prenni, A.J., Carrico, C.M., Ervens, B., and Feingold, G. (2006). Water activity and activation diameters from hygroscopicity data - Part II: Application to organic species. Atmos. Chem. Phys. *6*, 795–809.

Krumins, V., Mainelis, G., Kerkhof, L.J., and Fennell, D.E. (2014). Substrate-Dependent rRNA Production in an Airborne Bacterium. Environ. Sci. Technol. Lett. *1*, 376–381.

Lazaridis, M. (2019). Bacteria as Cloud Condensation Nuclei (CCN) in the Atmosphere. Atmosphere (Basel). *10*, 786.

Leizeaga, A., Meisner, A., Rousk, J., and Bååth, E. (2022). Repeated drying and rewetting cycles accelerate bacterial growth recovery after rewetting. Biol. Fertil. Soils *58*, 365–374.

Lindemann, J., Constantinidou, H.A., Barchet, W.R., and Upper, C.D. (1982). Plants as sources of airborne bacteria, including ice nucleation-active bacteria. Appl. Environ. Microbiol. *44*, 1059–1063.

Van Der Mei, H.C., Bos, R., and Busscher, H.J. (1998). A reference guide to microbial cell surface hydrophobicity based on contact angles. Colloids Surfaces B Biointerfaces *11*, 213–221.

Mekic, M., and Gligorovski, S. (2020). Ionic strength effects on heterogeneous and multiphase chemistry: Clouds versus aerosol particles. Atmos. Environ. *244*, 117911.

Möhler, O., DeMott, P.J., Vali, G., and Levin, Z. (2007). Microbiology and atmospheric processes: The role of biological particles in cloud physics. Biogeosciences *4*, 1059–1071.

Morris, C.E., Georgakopoulos, D.G., and Sands, D.C. (2004). Ice nucleation active bacteria and their potential role in precipitation. J. Phys. IV *121*, 87–103.

Noirmain, F., Baray, J., Tridon, F., Cacault, P., Billard, H., Voyard, G., Baelen, J. Van, and Latour, D. (2022). Interdisciplinary strategy to survey phytoplankton dynamics of a eutrophic lake under rain forcing: description of the instrumental set-up and first results. Biogeosciences.

Ogawa, S., Setoguchi, Y., Kawana, K., Nakayama, T., Ikeda, Y., Sawada, Y., Matsumi, Y., and Mochida, M. (2016). Hygroscopicity of aerosol particles and CCN activity of nearly hydrophobic particles in the urban atmosphere over Japan during summer. J. Geophys. Res. *121*, 7215–7234.

Péguilhan, R., Besaury, L., Rossi, F., Enault, F., Baray, J., Deguillaume, L., and Amato, P. (2021). Rainfalls sprinkle cloud bacterial diversity while scavenging biomass. FEMS Microbiol. Ecol. 1–15.

Petters, M.D., and Kreidenweis, S.M. (2007). A single parameter representation of hygroscopic growth and cloud condensation nucleus activity. Atmos. Chem. Phys. *7*, 1961–1971.

Pruppacher, H.R., and Jaenicke, R. (1995). The processing of water vapor and aerosols by atmospheric clouds, a global estimate. Atmos. Res. *38*, 283–295.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. *41*, 590–596.

Šantl-Temkiv, T., Amato, P., Gosewinkel, U., Thyrhaug, R., Charton, A., Chicot, B., Finster, K., Bratbak, G., and Löndahl, J. (2017). High-Flow-Rate Impinger for the Study of Concentration, Viability, Metabolic Activity, and Ice-Nucleation Activity of Airborne Bacteria. Environ. Sci. Technol. *51*, 11224–11234.

Šantl-Temkiv, T., Amato, P., Casamayor, E.O., Lee, P.K.H., and Pointing, S.B. (2022). Microbial ecology of the atmosphere. FEMS Microbiol. Rev.

453 Sattler, B., Puxbaum, H., and Psenner, R. (2001). Bacterial growth in supercooled cloud
454 droplets. Geophys. Res. Lett. *28*, 239–242.

455 Sellegri, K., Laj, P., Marinoni, A., Dupuy, R., Legrand, M., and Preunkert, S. (2003).
456 Contribution of gaseous and particulate species to droplet solute composition at the Puy de
457 Dôme, France. Atmos. Chem. Phys. *3*, 1509–1522.

458 Smith, D.J., Timonen, H.J., Jaffe, D.A., Griffin, D.W., Birmele, M.N., Perry, K.D., Ward, P.D.,
459 and Roberts, M.S. (2013). Intercontinental dispersal of bacteria and archaea by transpacific
460 winds. Appl. Environ. Microbiol. *79*, 1134–1139.

461 Tang, I.N., and Munkelwitz, H.R. (1993). Composition and temperature dependence of the
462 deliquescence droperties of hygroscopic aerosols. Atmos. Environ. *27*, 467–473.

463 Tignat-Perrier, R., Dommergue, A., Thollot, A., Magand, O., Amato, P., Joly, M., Sellegri, K.,
464 Vogel, T.M., and Larose, C. (2020). Seasonal shift in airborne microbial communities. Sci.
465 Total Environ. 137129.

466 Triadó-Margarit, X., Caliz, J., Reche, I., and Casamayor, E.O. (2019). High similarity in
467 bacterial bioaerosol compositions between the free troposphere and atmospheric depositions
468 collected at high-elevation mountains. Atmos. Environ. 79–86.

469 Wex, H., Stratmann, F., Topping, D., and McFiggans, G. (2008). The Kelvin versus the raoult
470 term in the köhler equation. J. Atmos. Sci. *65*, 4004–4016.

471 Wirgot, N., Vinatier, V., Deguillaume, L., Sancelme, M., and Delort, A.M. (2017). H2O2
472 modulates the energetic metabolism of the cloud microbiome. Atmos. Chem. Phys. *17*, 14841–
473 14851.

474 Woo, C., and Yamamoto, N. (2020). Falling bacterial communities from the atmosphere.
475 Environ. Microbiomes *15*, 22.

476

1

Supplementary information to:

Comparative study of bacterial communities in clouds and aerosols

Raphaëlle Péguilhan[1], Florent Rossi[1], Olivier Rué[2,3] and Pierre Amato[1]

[1] Université Clermont Auvergne, CNRS, SIGMA Clermont, ICCF, F-63000 CLERMONT-FERRAND, France
[2] Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France
[3] Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France

List of supplements:

- **Supplementary materials**

Meteorological data and air mass origin

Total cell counts and ATP quantification

Nucleic acid extraction and amplification

Bioinformatics and data analysis

- **Supplementary Tables and Figures**

Supplementary Table 1: Multiplexing information for sequencing.

Supplementary Table 2 (electronic .xlsx file): Taxonomy table.

Supplementary Table 3: Kruskal-Wallis test results for significantly differently represented bacterial genera between clouds and aerosols.

Supplementary Figure 1: Percentage of time spent over continent (land) or ocean (sea) at high (>1 km above ground level) or low (<1 km a.g.l.) altitude within the 72 h preceding sampling, for each sample collected at the puy de Dôme mountain. Data were recovered from OPGC data center: https://www.opgc.fr/data-center/public/data/copdd/trajectory, from the backward trajectories over the time of sampling. Sample name is composed of the date (yyyymmdd) and the environment type (cloud or air, i.e. aerosol).

Supplementary Figure 2: ASV distribution among the 53 environmental samples.

Supplementary Figure 3: Hierarchical clustering with p-values for the 53 environmental samples. Clustering associated to the Supplementary Figure 2.

Supplementary Figure 4: Rarefaction curves (53 environmental samples collected from 18 events). ASV: amplicon sequence variant.

Supplementary Figure 5: Relationship between read abundance and variability between replicates, at the genus level, in clouds and in the dry atmosphere. Expon.: exponential trends.

Supplementary materials

Meteorological data and air mass origin

Ambient temperature, relative humidity, and wind speed were measured at a frequency of 5 min at PUY and OPM meteorological stations; these are publicly available at https://www.opgc.fr/data-center/public/data/copdd/pdd. The boundary layer height (BLH) was extracted from ECMWF ERA5 data reanalysis (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5) (Hoffmann et al., 2019), and each event was qualified as "In" or "Out" the boundary layer, as in Péguilhan et al. (2021). The geographical origin of the air masses and the fraction of time spent over sea or land at high (> 1 km) or low (< 1 km) altitude over the three days preceding sampling were recovered from 72-hours backward trajectory plots, computed using the CAT trajectory model (Baray et al., 2020). These are publicly available for PUY site on the OPGC's database at https://www.opgc.fr/data-center/public/data/copdd/trajectory.

Total cell counts and ATP quantification

Total microbial cell concentrations were estimated by flow cytometry using BD FacsCalibur instrument (Becton Dickinson, Franklin Lakes, NJ), as in (Amato et al., 2017). Adenosine-5'-triphosphate (ATP) was quantified by bioluminescence (ATP Biomass Kit HS; BioThema; Handen, Sweden) as in (Vaïtilingom et al., 2013).

Nucleic acid extraction and amplification

DNA was extracted from MCE filters using NucleoMag® DNA/RNA Water kit (Macherey-Nagel, Hoerdt, France) and quantified by fluorescence using Quant-iT™ PicoGreen® dsDNA kit (Invitrogen; Thermo Fisher Scientific, Waltham, MA USA).

The V4 region of the bacterial16S sub-unit of ribosomal gene was amplified by PCR using the 515f-806r tagged primers (Apprill et al., 2015; Parada et al., 2016) (Supplementary Table 1). PCR product were purified using the QIAquick PCR Purification kit (Qiagen, Hilden, Germany). Sequencing was performed using Illumina Miseq 2*250 bp (GenoScreen; Lille, France). Demultiplexed sequencing files were deposited to the European Nucleotide Archive and have the accession numbers ERR9924931 to ERR9924983.

Bioinformatics and data analysis

Sequence numbers were rarefied to 2,770 sequences (corresponding to the sample with the lowest number of reads), using the FROGS "Abundance normalization" function, corresponding to 862 ASVs (Supplementary Table 2).

ASV abundance data were centered log-ratio (CLR)-transformed, as recommended by Gloor and colleagues (Gloor et al., 2017). Representations such as principal component analyses (PCA), heatmaps and Venn diagrams were obtained using the packages factoextra (v 1.0.7) (Alboukadel Kassambara and Fabian Mundt, 2019), pheatmap (v 1.0.12) (Raivo Kolde, 2019), ggdendro (v 0.1.22) (Andrie de Vries and Brian D. Ripley, 2016) and VennDiagram (v 1.6.20) (Chen and Boutros, 2011) using the R environment (v 4.0.3) (R Core Team (2019)). Statistics were performed using PAST software (v 4.02) (Hammer et al., 2001).

Tables and Figures

Supplementary Table 1: Multiplexing information for sequencing

Sample information			**Primer 16S - 515F (forward) GTG-YCA-GCM-GCC-GCG-GTA-A**		**Primer 16S - 806R (reverse) GGA-CTA-CNV-GGG-TWT-CTA-AT**	
Sample ID	**Type**	**Date (yyyymmdd)**	**N° primer**	**Tag sequence**	**N° primer**	**Tag sequence**
20191002CLOUD_K1K2	CLOUD	20191002	**F2**	acagcaca	**F1**	acacacac
20191022CLOUD_K0	CLOUD	20191022	**F2**	acagcaca	**F2**	acagcaca
20191022CLOUD_K1	CLOUD	20191022	**F2**	acagcaca	**F3**	gtgtacat
20191022CLOUD_K2	CLOUD	20191022	**F2**	acagcaca	**F4**	tatgtcag
20191022CLOUD_K3	CLOUD	20191022	**F2**	acagcaca	**F5**	tagtcgca
20200311CLOUD_K1	CLOUD	20200311	**F2**	acagcaca	**F6**	tactatac
20200311CLOUD_K2	CLOUD	20200311	**F2**	acagcaca	**F10**	gtcgtaga
20200311CLOUD_K3	CLOUD	20200311	**F2**	acagcaca	**F12**	gactgatg
20200717CLOUD_K1	CLOUD	20200717	**F2**	acagcaca	**F13**	agactatg
20200717CLOUD_K2	CLOUD	20200717	**F3**	gtgtacat	**F1**	acacacac
20200717CLOUD_K3	CLOUD	20200717	**F3**	gtgtacat	**F2**	acagcaca
20201016CLOUD_K1	CLOUD	20201016	**F3**	gtgtacat	**F3**	gtgtacat
20201016CLOUD_K2	CLOUD	20201016	**F3**	gtgtacat	**F4**	tatgtcag
20201016CLOUD_K3	CLOUD	20201016	**F3**	gtgtacat	**F5**	tagtcgca
20201028CLOUD_K1	CLOUD	20201028	**F3**	gtgtacat	**F6**	tactatac
20201028CLOUD_K2	CLOUD	20201028	**F3**	gtgtacat	**F10**	gtcgtaga
20201028CLOUD_K3	CLOUD	20201028	**F3**	gtgtacat	**F12**	gactgatg
20201103CLOUD_K1	CLOUD	20201103	**F3**	gtgtacat	**F13**	agactatg
20201103CLOUD_K2	CLOUD	20201103	**F4**	tatgtcag	**F1**	acacacac
20201103CLOUD_K3	CLOUD	20201103	**F4**	tatgtcag	**F2**	acagcaca
20201110CLOUD_K1	CLOUD	20201110	**F4**	tatgtcag	**F3**	gtgtacat
20201110CLOUD_K2	CLOUD	20201110	**F4**	tatgtcag	**F4**	tatgtcag
20201110CLOUD_K3	CLOUD	20201110	**F4**	tatgtcag	**F5**	tagtcgca
20201119CLOUD_K1	CLOUD	20201119	**F4**	tatgtcag	**F6**	tactatac
20201119CLOUD_K2	CLOUD	20201119	**F4**	tatgtcag	**F10**	gtcgtaga
20201119CLOUD_K3	CLOUD	20201119	**F4**	tatgtcag	**F12**	gactgatg
20190712AIR_PDD_K1	AIR	20190712	**F4**	tatgtcag	**F13**	agactatg
20190712AIR_PDD_K2	AIR	20190712	**F5**	tagtcgca	**F1**	acacacac
20190712AIR_PDD_K3	AIR	20190712	**F5**	tagtcgca	**F2**	acagcaca
20190918AIR_PDD_K1	AIR	20190918	**F5**	tagtcgca	**F3**	gtgtacat
20190918AIR_PDD_K2	AIR	20190918	**F5**	tagtcgca	**F4**	tatgtcag
20190918AIR_PDD_K3	AIR	20190918	**F5**	tagtcgca	**F5**	tagtcgca
20200206AIR_PDD_K1	AIR	20200206	**F5**	tagtcgca	**F6**	tactatac
20200206AIR_PDD_K2	AIR	20200206	**F5**	tagtcgca	**F10**	gtcgtaga
20200206AIR_PDD_K3	AIR	20200206	**F5**	tagtcgca	**F12**	gactgatg
20200518AIR_OPME_K1	AIR	20200518	**F5**	tagtcgca	**F13**	agactatg
20200518AIR_OPME_K2	AIR	20200518	**F6**	tactatac	**F1**	acacacac
20200518AIR_OPME_K3	AIR	20200518	**F6**	tactatac	**F2**	acagcaca

20200610AIR_PDD_K1	AIR	20200610	**F6**	tactatac	**F3**	gtgtacat
20200610AIR_PDD_K2	AIR	20200610	**F6**	tactatac	**F4**	tatgtcag
20200610AIR_PDD_K3	AIR	20200610	**F6**	tactatac	**F5**	tagtcgca
20200707AIR_PDD_K1	AIR	20200707	**F6**	tactatac	**F6**	tactatac
20200707AIR_PDD_K2	AIR	20200707	**F6**	tactatac	**F10**	gtcgtaga
20200707AIR_PDD_K3	AIR	20200707	**F6**	tactatac	**F12**	gactgatg
20200708AIR_PDD_K1	AIR	20200708	**F6**	tactatac	**F13**	agactatg
20200708AIR_PDD_K2	AIR	20200708	**F10**	gtcgtaga	**F1**	acacacac
20200708AIR_PDD_K3	AIR	20200708	**F10**	gtcgtaga	**F2**	acagcaca
20200709AIR_PDD_K1	AIR	20200709	**F10**	gtcgtaga	**F3**	gtgtacat
20200709AIR_PDD_K2	AIR	20200709	**F10**	gtcgtaga	**F4**	tatgtcag
20200709AIR_PDD_K3	AIR	20200709	**F10**	gtcgtaga	**F5**	tagtcgca
20200922AIR_PDD_K2	AIR	20200922	**F10**	gtcgtaga	**F6**	tactatac
20200922AIR_PDD_K3	AIR	20200922	**F10**	gtcgtaga	**F10**	gtcgtaga
20200922AIR_PDD_K4	AIR	20200922	**F10**	gtcgtaga	**F12**	gactgatg
20201118AIR_PDD_K1K2	AIR	20201118	**F10**	gtcgtaga	**F13**	agactatg
20201118AIR_PDD_K3K4	AIR	20201118	**F12**	gactgatg	**F1**	acacacac
20201124AIR_PDD_K1K2	AIR	20201124	**F12**	gactgatg	**F2**	acagcaca
20201124AIR_PDD_K3K4	AIR	20201124	**F12**	gactgatg	**F3**	gtgtacat

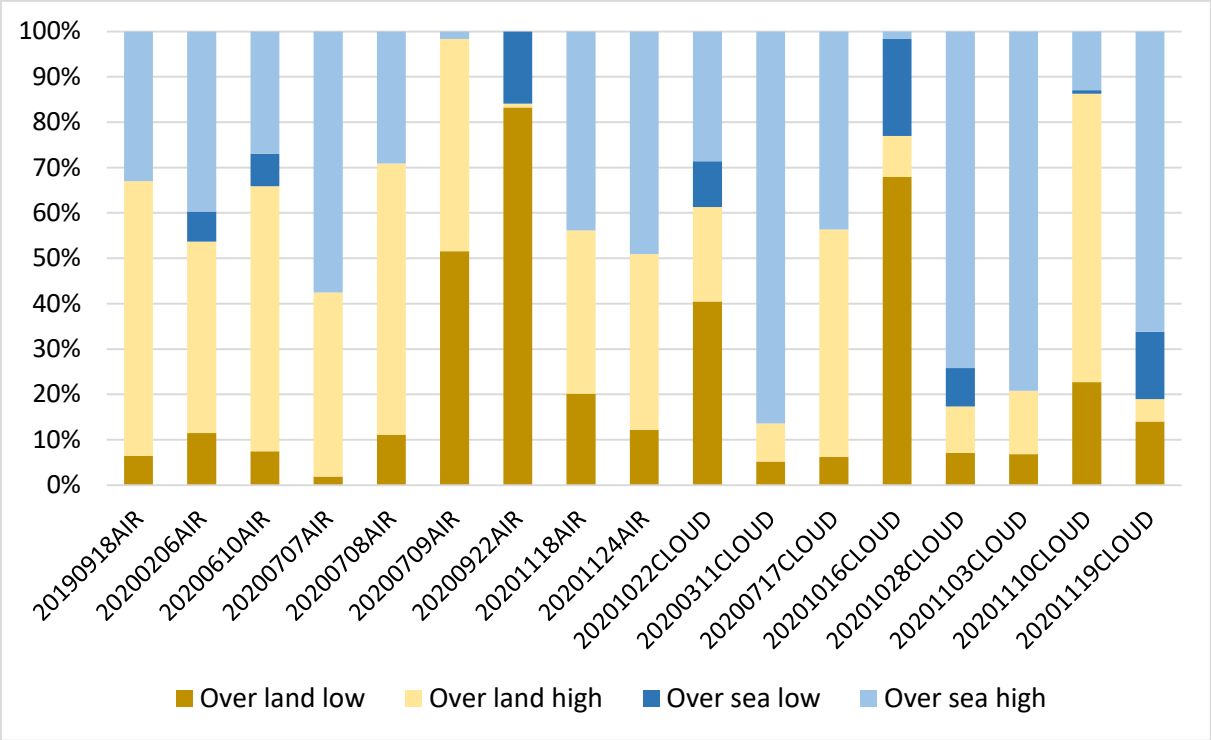
Supplementary Table 2 (electronic .xlsx file): Taxonomy table.

Supplementary Table 3: Kruskal-Wallis test results for significantly differently represented bacterial genera between clouds and aerosols.

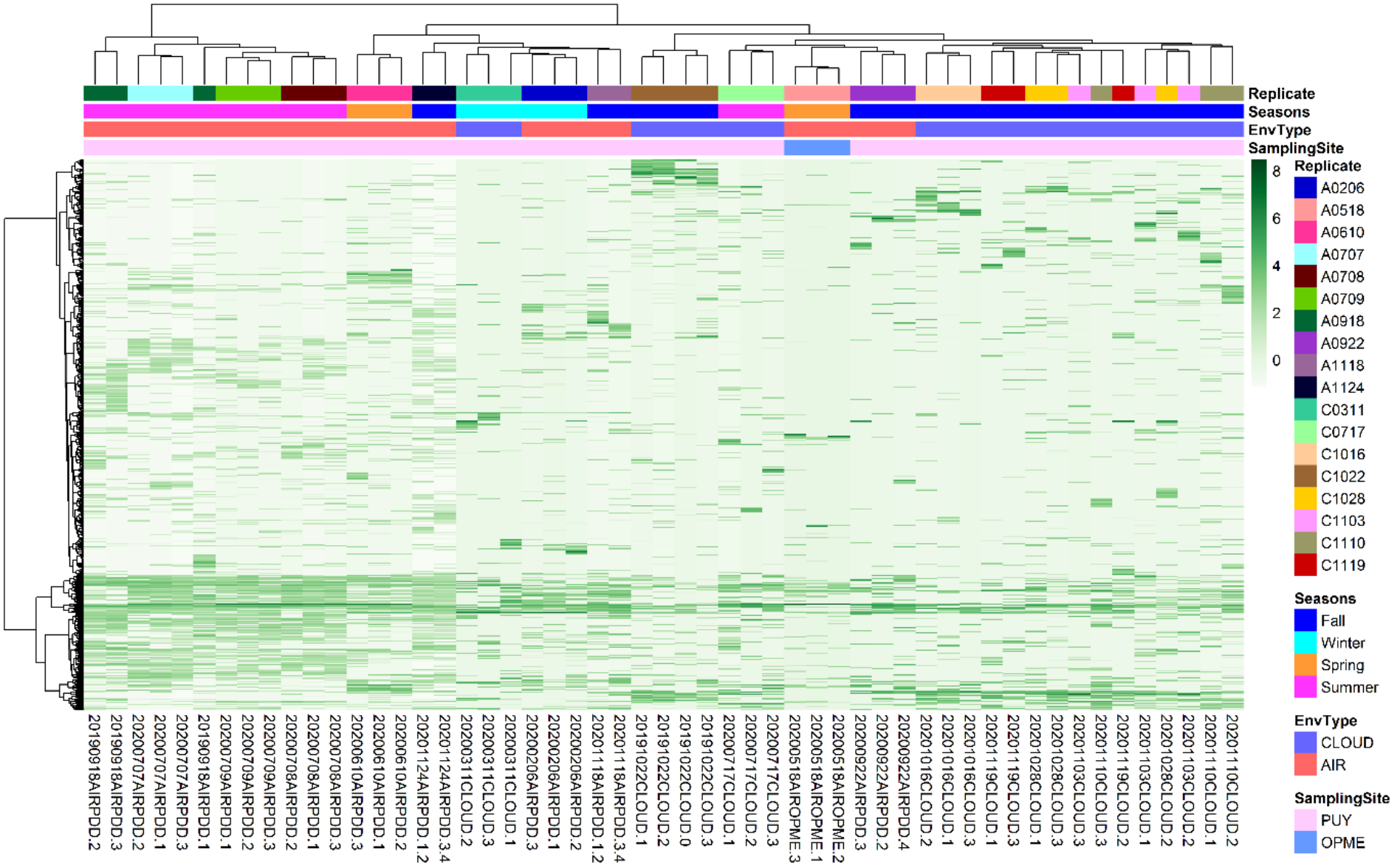
Genus	chi-squares	df	p-value	Mean number of sequence in Clouds (clr)*	Mean number of sequence in Aerosols (clr)*
Acidothermus	5.79	1	0.0162	-0.647	0.567
Acidovorax	5.96	1	0.0146	-0.659	-1.262
Acinetobacter	20.48	1	6.02E-06	3.835	1.986
Actinomycetospora	10.06	1	0.0015	-0.113	1.332
Aerococcus	6.13	1	0.0133	0.987	-0.328
Alloprevotella	5.37	1	0.0205	-0.663	-1.222
Altererythrobacter	3.98	1	0.0460	-0.862	-1.080
Amycolatopsis	4.57	1	0.0325	-0.899	-1.153
Aquabacterium	9.17	1	0.0025	-0.752	-1.262
Arenimonas	6.31	1	0.0120	-0.547	0.011
Bacillus	4.96	1	0.0259	2.802	3.895
Blastococcus	6.58	1	0.0103	0.754	2.189
Brachybacterium	5.70	1	0.0170	1.165	0.071
Brevibacillus	8.33	1	0.0039	-0.657	-1.262
Brevibacterium	16.65	1	4.50E-05	2.022	0.288
Bryobacter	12.45	1	0.0004	-0.940	0.332
Burkholderia	24.53	1	7.30E-07	1.891	-0.663
Candidatus Carsonella	6.58	1	0.0103	-0.866	-1.262
Candidatus Protochlamydia	6.13	1	0.0133	-0.489	-1.133
Candidatus Spiroplasma	22.13	1	2.55E-06	0.907	-1.197
Candidatus Sulcia	7.33	1	0.0068	-0.756	-1.262
Candidatus Udaeobacter	13.87	1	0.0002	0.389	1.921
Candidatus Xiphinematobacter	5.28	1	0.0215	-0.800	-0.017
Chryseobacterium	4.05	1	0.0441	0.569	1.487
Cloacibacterium	6.67	1	0.0098	-0.468	-1.100
Clostridium sensu stricto 1	10.52	1	0.0012	1.810	0.431
Clostridium sensu stricto 13	4.96	1	0.0259	-0.590	-0.055
Cnuella	4.88	1	0.0271	-0.973	-0.247
Conexibacter	6.58	1	0.0103	-0.586	0.463
Corynebacterium	17.53	1	2.83E-05	3.440	1.308
Craurococcus-Caldovatus	5.96	1	0.0146	-1.028	-1.262
Dactylosporangium	4.80	1	0.0284	-0.717	-1.133
Defluviicoccus	4.27	1	0.0388	-0.823	-1.133
Enhydrobacter	23.66	1	1.15E-06	2.885	0.391
Facklamia	4.13	1	0.0422	-0.533	-1.005
Ferruginibacter	6.77	1	0.0093	0.566	1.680
Frankia	6.05	1	0.0139	-0.861	-1.262
Fusibacter	9.50	1	0.0021	-0.417	-1.262
Gaiella	9.72	1	0.0018	0.264	1.423
Gemella	11.83	1	0.0006	0.065	-1.262
Geobacillus	11.10	1	0.0009	-0.452	-1.262

Geodermatophilus	5.79	1	0.0162	-0.590	-1.197
Gilliamella	5.62	1	0.0178	-0.936	-0.842
Granulicatella	7.05	1	0.0079	-0.776	-1.262
Haemophilus	8.13	1	0.0044	-0.303	-1.262
Halomonas	9.39	1	0.0022	-0.684	-1.262
Herbiconiux	3.98	1	0.0460	-0.873	-1.165
Hydrotalea	6.58	1	0.0103	-1.000	-1.262
Ilumatobacter	5.45	1	0.0196	-0.759	-0.240
Janthinobacterium	5.20	1	0.0226	-0.900	-1.165
JGI 0001001-H03	9.95	1	0.0016	-0.701	0.140
Kineosphaera	6.05	1	0.0139	-0.843	-1.262
Knoellia	8.03	1	0.0046	0.042	1.847
Kocuria	32.92	1	9.63E-09	4.143	0.770
Kurthia	4.88	1	0.0271	-0.744	-1.153
Leuconostoc	6.58	1	0.0103	-0.066	-0.844
Ligilactobacillus	6.31	1	0.0120	-0.549	-1.118
Lysinibacillus	8.96	1	0.0028	-0.133	1.158
Macrococcus	7.05	1	0.0079	-0.540	-1.140
Marmoricola	6.86	1	0.0088	0.073	1.053
Mesorhizobium	12.20	1	0.0005	-0.945	0.111
Micrococcus	19.68	1	9.14E-06	2.768	0.088
Neisseria	10.06	1	0.0015	0.015	-1.133
Nesterenkonia	6.58	1	0.0103	-0.792	-1.262
Noviherbaspirillum	9.83	1	0.0017	-0.298	1.210
Novosphingobium	8.23	1	0.0041	-0.227	0.848
Paracoccus	11.10	1	0.0009	3.697	2.053
Pelomonas	4.13	1	0.0422	-0.179	-0.711
Psychrobacter	8.13	1	0.0044	-0.573	-1.262
Reyranella	4.13	1	0.0422	-0.984	-0.329
Rhodopseudomonas	4.57	1	0.0325	-0.285	-0.988
Rickettsia	6.05	1	0.0139	0.002	-0.671
Rothia	8.13	1	0.0044	-0.382	-1.262
Rubellimicrobium	6.95	1	0.0084	1.099	2.488
Saccharopolyspora	5.28	1	0.0215	-0.402	-0.985
Serratia	5.12	1	0.0236	-1.072	-1.047
Spiroplasma	3.98	1	0.0460	-1.072	-1.053
Spirosoma	4.96	1	0.0259	0.076	1.154
Staphylococcus	15.37	1	8.86E-05	3.784	2.330
Streptococcus	8.64	1	0.0033	1.038	-0.527

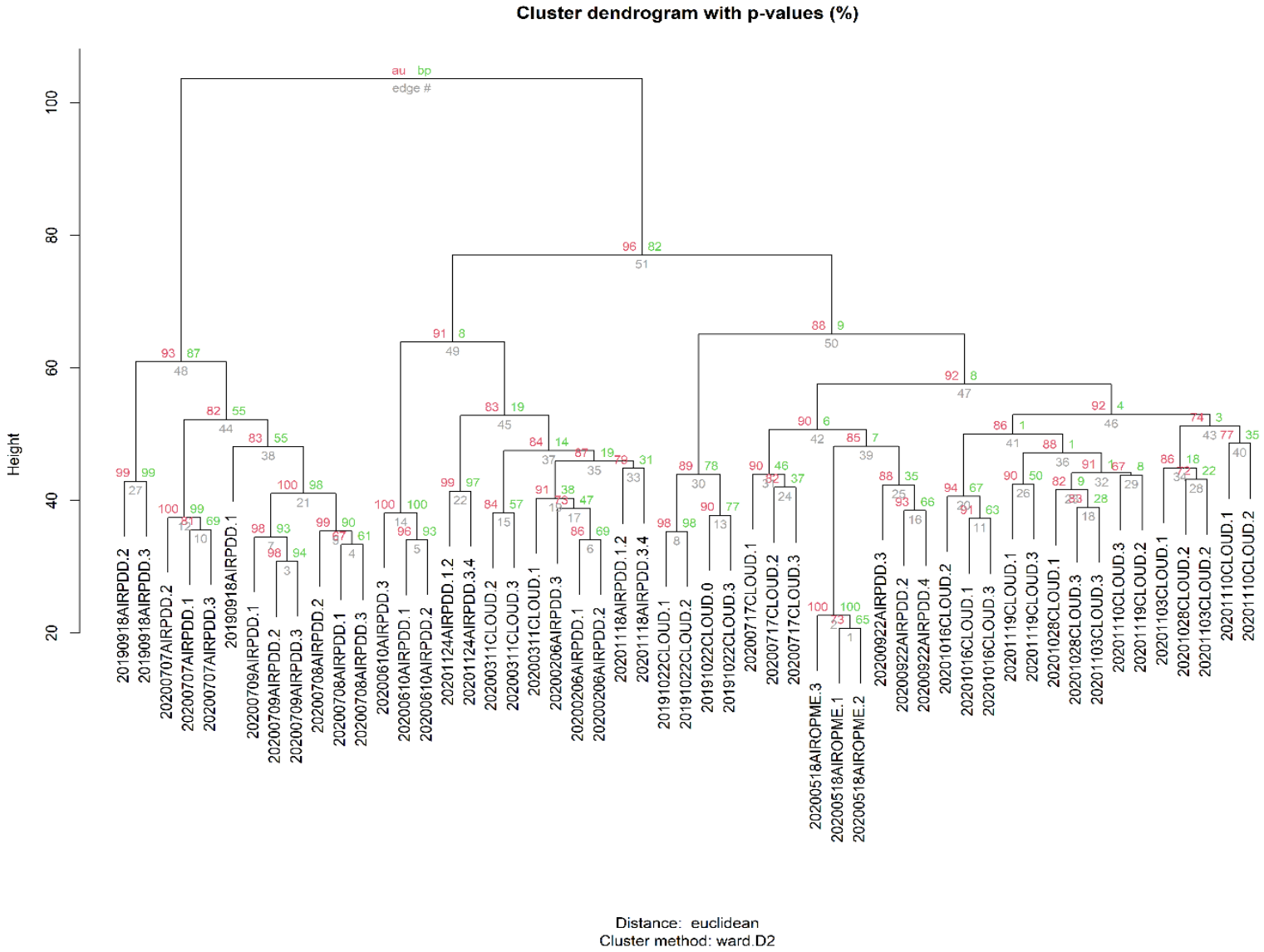
*clr: data transformed in centered-log ratio



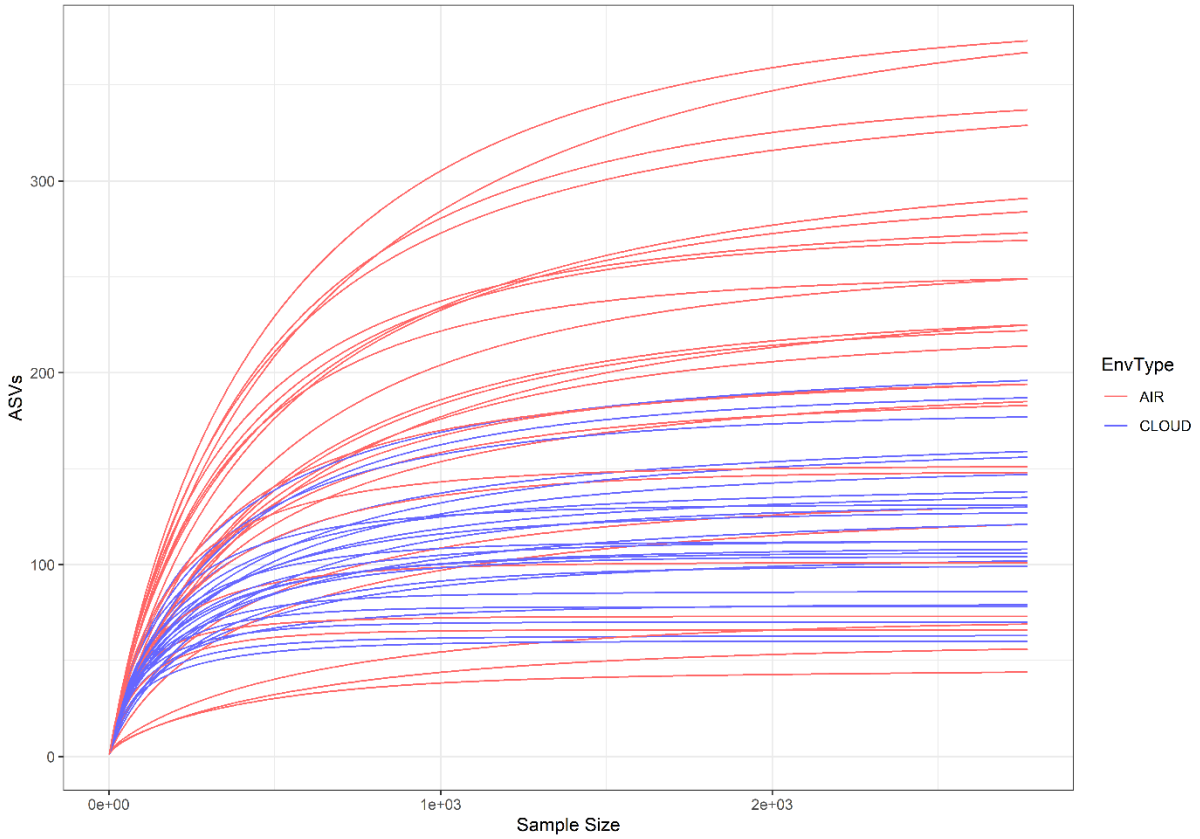
Supplementary Figure 1: Percentage of time spent over continent (land) or ocean (sea) at high (>1 km above ground level) or low (<1 km a.g.l.) altitude within the 72 h preceding sampling, for each sample collected at the puy de Dôme mountain. Data were recovered from OPGC data center: https://www.opgc.fr/data-center/public/data/copdd/trajectory, from the backward trajectories over the time of sampling. Sample name is composed of the date (yyyymmdd) and the environment type (cloud or air, i.e. aerosol).



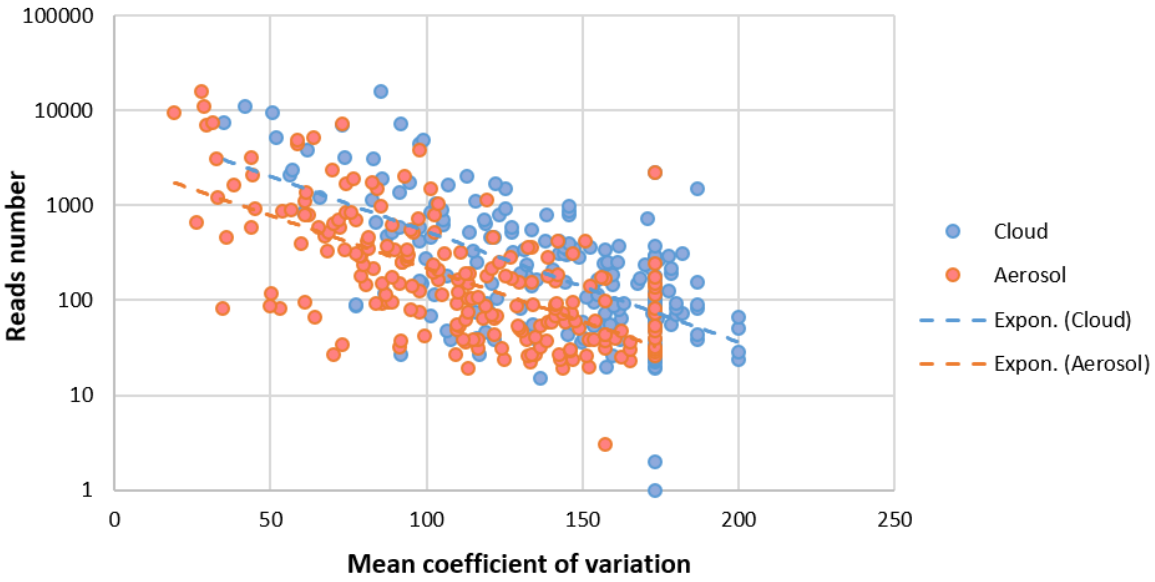
Supplementary Figure 2: ASV distribution among the 53 environmental samples.



Supplementary Figure 3: Hierarchical clustering with p-values for the 53 environmental samples. Clustering associated to the **Supplementary Figure 2.**



Supplementary Figure 4: Rarefaction curves (53 environmental samples collected from 18 events). ASV: amplicon sequence variant.



Supplementary Figure 5: Relationship between read abundance and variability between replicates, at the genus level, in clouds and in the dry atmosphere. Expon.: exponential trends.

6. Conclusion

To conclude on this chapter, multiple verifications, negative and positive controls were established to validate our experimental procedure for nucleic acid-based analyses of the atmosphere. The final nucleic acid concentrations obtained were sufficient to perform multiple amplicon sequencing analyses (cloud vs air, and cloud vs rain) and to directly sequence MG and MT for our functional analysis. Good sampling and sample processing practices were recalled and are strongly recommended, particularly in a low biomass environment. Sampling replicates are also important to verify sampler reproducibility and be aware of possible intra-event variations. This step was necessary before pooling samples for further analysis.

Bacterial diversity was studied for the first time, to our knowledge, under aerosol and cloud conditions. Aerosol events appeared much more variable than cloud events, which may imply an underlying phenomenon controlling the variability of bacterial diversity in clouds. Sampling clouds in the free troposphere and aerosols partially in the ABL could also explain these observations, as microbial emission sources are more variable in the ABL. Despite this, aerosols and clouds appeared to harbor similar communities influenced by a seasonal effect.

To go further, it is now interesting to study the next stage of the microbial atmospheric cycle, which is the transition from clouds to surface ecosystems *via* precipitation. This part is studied in the next **Chapter 3**. It is also of primary interest to explore the potential activity of the bacterial communities described in clouds and aerosols. Biodiversity may be similar, but there may be a specific partitioning regarding metabolic functioning. Indeed, clouds may provide an aqueous microenvironment for cells and perhaps more favorable conditions for their development. This issue is explored in **Chapter 4**.

Chapter 3: Partitioning of bacterial communities between clouds and precipitation

1. Introduction

This study started in 2016 in the context of the "Chlorofilter" ANR (Agence nationale de la recherche) project which has for main goal to study methylotroph bacteria in different environment. Cloud, rain, and snow samples were collected and analyzed but, because of the lack of samples and conclusive data, the project was set aside.

This work has been continued in 2019 within the framework of this thesis with a redefined objective. The problematic here was to investigate the partitioning of bacterial communities between clouds and precipitation to better understand the impact of below cloud scavenging on the bacterial composition of precipitation reaching the Earth's surface. Bacteria in precipitation can impact surface ecosystems through the input of new species and genetic material and are therefore of interest in terms of microbial ecology. This phenomenon is also interesting in the field of health with the monitoring of potential pathogenic agents for humans or for plants (crops).

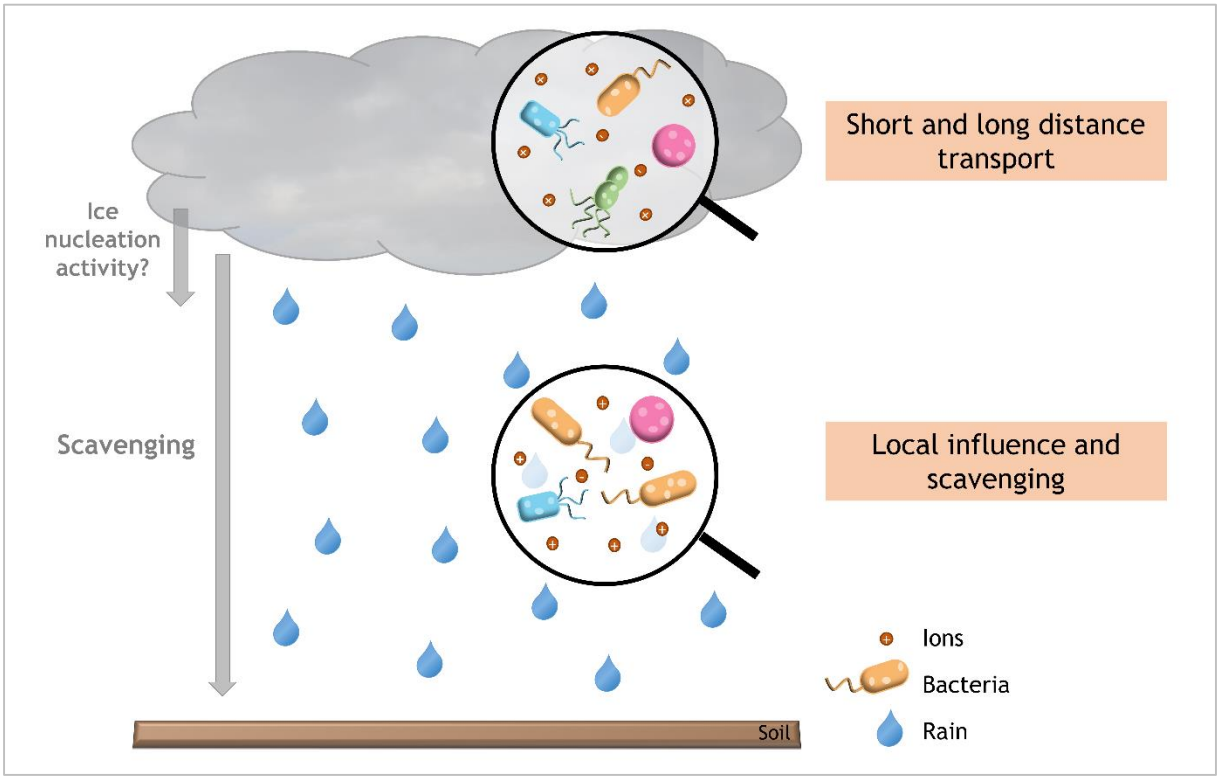
In this purpose new cloud and rain samples were collected at the same sites than in 2016-2017 to complete the previous dataset. Only five "old" samples were kept out of the thirteen due to their too low number of sequenced reads and ten new samples were added from 2019. The sequencing of the amplicons was done with same primers as in the project "Chlorofilter" (primers designed to avoid chloroplast) for the 2019 samples and the bioinformatics processing was completely redone for all the samples in order to standardize the analysis.

I presented this work at the **EGU General Assembly 2020 online congress** and at the "17ème rencontre des microbiologistes du pôle Clermontois" (17th meeting of the microbiologists from Clermont-Ferrand). It was also **published in the journal FEMS Microbiology Ecology in 2021**. Section 2 of this chapter 3 consists of the publication.

In this work I participated in:

- The collection of rain and cloud samples in 2019.
- The biological analysis of samples (cell count and ATP quantification).
- The DNA extraction and amplification with tagged primers.
- The setup of the bioinformatics workflow and the data processing.
- The writing of the publication (methodology, results, discussion) and the follow-up of the publication process.

2. Article 3: "Rainfalls sprinkle cloud bacterial diversity while scavenging biomass", published in FEMS Microbiology Ecology



Schematic representation (featured image) of the main objective of the paper "Rainfalls sprinkle cloud bacterial diversity while scavenging biomass", Péguilhan *et al.* 2021, FEMS Microbiology Ecology.

FEMS Microbiology Ecology, 97, 2021, fiab144

https://doi.org/10.1093/femsec/fiab144
Advance Access Publication Date: 3 November 2021
Research Article

RESEARCH ARTICLE

Rainfalls sprinkle cloud bacterial diversity while scavenging biomass

Raphaëlle Péguilhan[1,*,†], Ludovic Besaury[1,‡], Florent Rossi[1], François Enault[2], Jean-Luc Baray[3,4], Laurent Deguillaume[3,4] and Pierre Amato[1,#]

[1]Université Clermont Auvergne, CNRS, SIGMA Clermont, ICCF, F-63000 CLERMONT-FERRAND, France, [2]Université Clermont Auvergne, CNRS, Laboratoire Microorganismes: Genome et Environnement, F-63000 CLERMONT-FERRAND, France, [3]Université Clermont Auvergne, CNRS, Observatoire de Physique du Globe de Clermont-Ferrand, UMS 833, F-63000 CLERMONT-FERRAND, France and [4]Université Clermont Auvergne, CNRS, Laboratoire de Météorologie Physique, UMR 6016, F-63000 CLERMONT-FERRAND, France

*Corresponding author: Institut de Chimie de Clermont-ferrand (ICCF), UMR6296 CNRS-UCA-Sigma, 63178 AUBIERE Cedex, France. Tel: +33 (0)4 73 40 52 84; E-mail: raphaelle.peguilhan@uca.fr

One sentence summary: Comparative study of bacterial diversity in clouds and precipitation reveals distinct structures indicative of bacteria distribution and dispersal in the outdoor atmosphere.

Editor: Lee Kerkhof

[†]Raphaëlle Péguilhan, https://orcid.org/0000-0002-7206-0120
[‡]Ludovic Besaury, https://orcid.org/0000-0002-1463-1239
[#]Pierre Amato, https://orcid.org/0000-0003-3168-0398

ABSTRACT

Bacteria circulate in the atmosphere, through clouds and precipitation to surface ecosystems. Here, we conducted a coordinated study of bacteria assemblages in clouds and precipitation at two sites distant of ~800 m in elevation in a rural vegetated area around puy de Dôme Mountain, France, and analysed them in regard to meteorological, chemical and air masses' history data. In both clouds and precipitation, bacteria generally associated with vegetation or soil dominated. Elevated ATP-to-cell ratio in clouds compared with precipitation suggested a higher proportion of viable cells and/or specific biological processes. The increase of bacterial cell concentration from clouds to precipitation indicated strong below-cloud scavenging. Using ions as tracers, we derive that 0.2 to 25.5% of the 1.1×10^7 to 6.6×10^8 bacteria cell/m^2/h^1 deposited with precipitation originated from the source clouds. Yet, the relative species richness decreased with the proportion of inputs from clouds, pointing them as sources of distant microbial diversity. Biodiversity profiles, thus, differed between clouds and precipitation in relation with distant/local influencing sources, and potentially with bacterial phenotypic traits. Notably *Undibacterium*, *Bacillus* and *Staphylococcus* were more represented in clouds, while epiphytic bacteria such as *Massilia*, *Sphingomonas*, *Rhodococcus* and *Pseudomonas* were enriched in precipitation.

Keywords: bacterial diversity; atmosphere; cloud; precipitation; scavenging; bioaerosol dispersal

INTRODUCTION

Earth' surface continuously exchanges biological material with the atmosphere. It was estimated that, globally, ~10^{24} bacteria cells are aerosolized from surface ecosystems each year (Burrows *et al.* 2009). With a modeled atmospheric residence time of up to several days and a half-life of several hours (Amato *et al.* 2015), airborne bacteria are then prone to be dispersed over

Received: 29 April 2021; **Accepted:** 27 October 2021

regional and continental scales (Hervàs *et al.* 2009; Smith *et al.* 2013; Barberán *et al.* 2015; Griffin *et al.* 2017; Weil *et al.* 2017), carried up to high altitudes (Smith *et al.* 2018) and integrate clouds and the atmospheric water cycle (e.g. Bauer *et al.* 2002; Amato *et al.* 2007).

Typical bacteria number concentration in the outdoor atmosphere ranges between ~10^2 and ~10^6 cells/m^3 of air (Després *et al.* 2012) and ~10^2- ~10^5 mL in condensed water (Vaïtilingom *et al.* 2012; Pouzet *et al.* 2017). The bacterial assemblage is highly diverse and often dominated by Alpha-, Beta- and Gamma-Proteobacteria, along with Bacteroidetes, Actinobacteria and Firmicutes species (e.g. Amato *et al.* 2017a; Šantl-Temkiv *et al.* 2018). Wet processes are major in the deposition of aerosol particles from the atmosphere, of which bacteria, whose atmospheric cycle is thus intimately linked with that of water (Morris *et al.* 2008; Evans *et al.* 2019). Aerosols themselves are essential actors of the formation of clouds and precipitation: they serve as nuclei for (i) the condensation of water vapor into cloud droplets (Cloud Condensation Nuclei CCN), and (ii) freezing of supercooled water into ice crystals (Ice Nuclei IN; e.g. Möhler *et al.* 2007); the latter often triggers precipitation in mixed-phase clouds and is of primary interest in atmospheric sciences. Certain bacteria affiliated or close to *Pseudomonas* species produce proteins that were identified as the most active IN existing in nature (Lindow, Arny and Upper 1978), suggesting a role of biological ice nucleation in precipitation (Morris *et al.* 2008).

The composition of water reaching the surface with precipitation does not directly reflect that in the cloud. On their path to the ground, falling raindrops collect aerosol particles by below-cloud scavenging (Jaffrezo and Colin 1988; Bourcier *et al.* 2012). This is a complex phenomenon involving physical processes such as Brownian diffusion, inertial impaction and interception, whose efficiency largely depends on aerosol particle size and composition, rain drop size and rainfall intensity (Willis and Tattelman 1989; Mircea, Stefan and Fuzzi 2000; Hou *et al.* 2018; Sonwani and Kulshrestha 2019). Heavy rains (> ~10 mm/h) are in general much more efficient in scavenging aerosols particles <2.5 μm in diameter (PM2.5) than light rains (< ~1.0 mm/h), with efficiencies of ~50% vs. ~5%, respectively (Luan *et al.* 2019). This discrimination leads to the enrichment or depletion of certain solutes and particles in rain water with respect to cloud water or to aerosol (Jaffrezo and Colin 1988). However, very little is known currently concerning the possibility of differential scavenging among bacteria aerosols. Interestingly for microbial ecology, a gap (Greenfield gap) in the scavenging efficiency of aerosol particles by rain was identified for particles between ~0.2 and 1.0 μm, i.e. the usual size of bacteria cells (Radke, Hobbs and Eltgroth 1980; Ladino *et al.* 2011; Blanco-Alegre *et al.* 2018).

Clouds and precipitation have been studied in the past for their microbiological contents (Amato *et al.* 2017b; Aho *et al.* 2020). However, to our knowledge, these have never been investigated coordinately in a natural context. Here, with the aim to decipher the atmospheric life history of wet-deposited bacteria, we examined the biological (biodiversity, biomass and activity) and chemical (major inorganic ions) contents along the 'first' steps of the water cycle, in clouds and their precipitation. Samples were collected at two meteorological stations installed at rural sites geographically close, but that differ in elevation by about 800 m (puy de Dôme Mountain summit and the surrounding plateau). Air masse's origins and histories were characterized, and a comparative study was performed between clouds and precipitation for discriminating between bacteria deposited from clouds from those that entered the water cycle after being washed-out from the air column underneath. The data help understanding bacteria dispersal and fate in the atmosphere and atmosphere–surface exchanges.

MATERIAL AND METHODS

The methods are summarized below; more details are provided as supplementary information.

Sample collection

Samples were collected in the rural area of puy de Dôme Mountain, France, located ~400 km east from the Atlantic Ocean, and ~300 km north from the Mediterranean Sea. The surrounding landscape comprises mainly deciduous forests and pastures. A total of two sites separated by 12 km and 785 m in elevation were prospected for cloud water and rain sampling: puy de Dôme Mountain' summit (PUY; 1465 m a.s.l., 45.772°N, 2.9655°E) and Opme station (OPME; 680 m a.s.l., 45.7125°N, 3.090278°E), respectively (Figure S1, Supporting Information). These meteorological stations are part of the Cézeaux-Aulnat-Opme-Puy de Dôme (CO-PDD) instrumented platforms for atmospheric research (Baray *et al.* 2020). In total, four cloud water and one fresh snow samples were collected from PUY, and 10 rain samples from OPME.

Cloud water was collected during spring 2017 and autumn 2019 over periods of ~2 to 7.5 consecutive hours (Table 1). In 2017, only cloud droplet impactors sterilized by autoclave were used for all biological and chemical analyses as in (Amato *et al.* 2019). In 2019, three additional high-flow-rate (HFR) impingers (DS6, Kärcher SAS and Bonneuil sur Marne, France; air-flow rate of 2 m^3/min; Šantl-Temkiv *et al.* 2017) were deployed specifically for nucleic acid analyses (Figure S1C, Supporting Information); other analyses were carried out from the cloud droplet impactors sampling in parallel. HFR impingers were filled with 850 mL of GF/F-filtered autoclaved nucleic acid preservation (NAP) buffer solution (Camacho-Sanchez *et al.* 2013; Menke *et al.* 2017) as the collection liquid. This contains 0.019 M of ethylenediaminetetra-acetic acid (EDTA) disodium salt dihydrate, 0.018 M of sodium citrate trisodium salt dihydrate, 3.8 M of ammonium sulfate and H_2SO_4 to adjust the pH at 5.2. Total volumes of 309–1300 mL of cloud water were collected and processed immediately after sampling using the station's microbiology facility, within a laminar flow hood previously exposed to UV for 15 min.

Rain water was collected over 24-h periods from the plateau underneath the mountain summit in November 2016, March 2017 and October–November 2019 using an automated refrigerated (4°C) rain collector NSA 181/KHS (47.4 cm diameter; surface area = 1764 cm^2; Eigenbrodt; Königsmoor, Germany) as in (Pouzet *et al.* 2017; Figure S1D, Supporting Information). Total volumes of water of 20 to 1124mL were collected and processed in a laminar flow hood within 24 h following collection.

Additionally, one fresh snow sample was collected at the puy de Dôme summit during an intense snowfall event, from the top of a snowpack of several tens of centimetres deep using sterilized stainless steel spoons. The top 10 cm of the snow cover were collected (i.e. snow that accumulated within approximately the last 2 h), after removing the top layer (~2 cm). A total of two sterile 1 L-glass bottles were filled with snow and placed at 4°C for melting, then processed for analyses.

Table 1. Meteorological contexts during sampling.

SampleID*	Sampling location—altitude (m a.s.l.)	Sampling date (mm/dd/yyyy)	Cumulated precipitation or cloud sampling duration (h)	Average temperature (°C)#	Wind speed (average [max]) (m s^{-1})#	Precipitation rate (average [max]) (mm/h)#	Boundary layer altitude (min-max [average]) (m)#,‡	Main geographical origin
SNOW								
20170116SNOW	PUY–1465 m	1/16/2017	0.5	−5.7	3.2 [4.6]	-	844–1012 [928]	NW
RAIN								
20 161 105–06RAIN	Opme–680 m	11/5 and 11/6/2016	23	7.2	NA**	2.5 [12.0]	510–1318 [925]	W
20170307RAIN[a]	Opme–680 m	3/7/2017	3	3.4	5.2 [9.1]	1.3 [3.6]	540–1808 [1399]	W
20170323RAIN	Opme–680 m	3/23/2017	8	6.5	1.1 [3.5]	2.9 [24.0]	557–1346 [856]	SW
20191001RAIN[b]	Opme–680 m	10/1/2019	6	20.1	2.3 [5.0]	4.4 [24.0]	1035–1406 [1165]	W
20191015RAIN	Opme–680 m	10/15/2019	15	NA**	NA**	4.3 [14.4]	727–1532 [1196]	W
20191022RAIN[c]	Opme–680 m	10/22/2019	14	10.0	0.3 [2.4]	1.5 [7.2]	444–777 [569]	S
20191023RAIN[c]	Opme–680 m	10/23/2019	13.5	11.4	1.5 [4.7]	2.7 [24.0]	481–1285 [835]	SE
20191028RAIN	Opme–680 m	10/28/2019	12	14.2	0.3 [2.9]	1.5 [3.6]	441–864 [491]	W
20191031RAIN	Opme–680 m	10/31/2019	9	12.4	1.6 [5.1]	1.3 [6.0]	449–1235 [760]	W
20191104RAIN	Opme–680 m	11/4/2019	19	9.9	4.3 [7.2]	1.9 [8.4]	889–1815 [1476]	W
CLOUDS								
20170308CLOUD[a]	PUY–1465 m	3/8/2017	7.5	2.9	14.9 [19.5]	-	798–1384 [1174]	W
20190925CLOUD	PUY–1465 m	9/25/2019	3.3	6.6	10.8 [13.0]	-	1105–1728 [1321]	W
20191002CLOUD[b]	PUY–1465 m	10/2/2019	2.4	6.5	3.0 [4.6]	-	1263–1318 [1296]	NW
20191022CLOUD[c]	PUY–1465 m	10/22/2019	6.4	NA**	NA**	-	511–777 [623]	SE

*Superscripted letters indicate chronological associations between cloud and rain samples;
#Over the precipitation or cloud sampling period;
‡Data extracted from ECMWF ERA5 model;
Based on CAT 72-h back trajectory plots starting at Opme for rain (ground level—cloud level) and at PUY summit level for clouds and snow; see Table S1 (Supporting Information) for details;
NA**: No data available.

Meteorological data and backward trajectory plots

Meteorological variables were monitored by the meteorological stations installed at PUY and OPME. In addition, vertical profiles of cloud liquid and ice water contents (LWC and IWC) and boundary layer height (BLH) were extracted from the ECMWF ERA5 global reanalysis (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5; Hoffmann *et al.* 2019). The geographical origin of the air masses were obtained from 72-h backward trajectory plots computed using the CAT trajectory model (Baray *et al.* 2020). The model uses dynamical fields extracted from the ERA-5 meteorological data archive with, for the present work, a spatial resolution of 0.5° in latitude and longitude. This tool allowed to compute: (i) air mass backward trajectories starting from Opme (ground and cloud level) and PUY summit; (ii) air masses history, as the density of trajectory points below the BLH and the percentage of trajectory points above and below the BLH, over land and seas; (iii) the percentage of trajectory points near the CO-PDD observatory (distance < 50 km) and in each of eight direction sectors (Renard *et al.* 2020).

Chemical analyses

The pH was measured immediately after sampling, and the main dissolved ions (Na^+, NH_4^+, K^+, Mg^{2+}, Ca^{2+}, Cl^-, NO_3^- and SO_4^{2-}) were examined by ion chromatography from 5 mL of filtered (0.2 μm) subsamples kept at –25°C. Analyses were carried out using either a Dionex DX320 (column AS11) for anions and a Dionex ICS1500 (column CS16) for cations, as in (Deguillaume *et al.* 2014), or an ICS3000 dual channel chromatograph (Thermo Fisher Scientific, Waltham, MA USA) with AS11HC column for anions and CS12 for cations, as in (Jaffrezo, Calas and Bouchet 1998) and (Waked *et al.* 2014).

Cell counts and ATP quantification

Total cells counts were performed by flow cytometry using a BD FacsCalibur instrument (Becton Dickinson, Franklin Lakes, NJ), as in (Amato *et al.* 2017b). Adenosine-5′-triphosphate (ATP) in the samples was quantified by bioluminescence (ATP Biomass Kit HS; BioThema; Handen, Sweden) as in (Vaïtilingom *et al.* 2013).

DNA extraction, amplification and sequencing

Water samples were filtered (0.22 μm porosity) and DNA was extracted from filters using commercial kits. Amplification of the 16S sub-unit of bacterial ribosomal genes was performed from genomic DNA extracts by PCR targeting the V5, V6 and V7 regions, using the universal primers 799f (5′-ACCMGGA TTAGATACCCKG-3′) and 1193r (3′-GAGGAAGGTGGGGATGCGT -5′) and following the conditions specified in (Bulgarelli *et al.* 2012). Amplicons were sequenced on Illumina Miseq 2*250 bp (GenoScreen; Lille, France). Demultiplexed sequencing files were deposited at the European Nucleotide Archive with the accession numbers ERS5445211–ERS5445225.

Bioinformatics and data analysis

Illumina reads were demultiplexed using a custom Python 3.0 script. All the 360 567 reads (average size of 395 bp) were preprocessed using Mothur software v 1.41.3 (Schloss *et al.* 2009) with the Miseq standard operating procedure (Kozich *et al.* 2013). Sequences were filtered for ambiguous bases and to a minimum length of 350 bp. A total of 329 986 sequences with an average length of 394 bp remained for further analysis. Then, the pipeline FROGS (Find Rapidly OTUs with Galaxy Solution; Escudié *et al.* 2018) was used through the Galaxy v 3.1 environment deployed by the AuBi (Auvergne BioInformatique) network and the regional calculation cluster Mesocentre Clermont Auvergne. The tools 'FROGS Clustering swarm' and 'FROGS Remove chimera' were used with default parameters to cluster the sequences at 97% identity and remove possible chimera. This step removed 10.9% of the Operational Taxonomic Units (OTUs), representing 4% of the total sequences, and resulted in 82 267 OTUs. Then, 'FROGS Filters' was used to select only OTUs represented by at least three sequences: 4601 OTUs were kept after this step, corresponding to 6.3% of the clusters and to 22.7% of all sequences. The 'FROGS Clusters stat' tool was used at each step to obtain metrics on the clusters. Taxonomic affiliation of each OTU's seed was carried out using the 'FROGS Affiliation OTU' tool with 'Silva_132_16S' as the reference database (Quast *et al.* 2013). Both BLAST and RDP assignations were performed. Data were rarefied to 11 300 sequences per sample (corresponding to the sample with the lowest number of sequences) using 'FROGS Abundance normalization' tool.

Abundance table was checked manually for accurate OTU affiliations. Multi-affiliations at family level, percentages of identity <95% and percentages of query coverage <98% with BLAST assignations from SILVA 132 database were verified using RDP and EzBioCloud 16S rRNA gene-based ID database (https://www.ezbiocloud.net/, update 2020.02.25; Yoon *et al.* 2017). A total of 38 OTUs affiliated with *Mitochondria* and five unaffiliated OTUs were deleted, leaving 4510 OTUs for analysis.

OTU abundance data were then centered log-ratio (CLR)-transformed, as recommended by Gloor and colleagues to account for their compositional nature (Gloor *et al.* 2017). Data were analysed and represented using the *R* environment 3.6.0 [4]. First, *zCompositions* package v 1.3.4 (Palarea-Albaladejo and Martín-Fernández 2015) was used for imputing zeros in our compositional count data set based on a Bayesian-multiplicative replacement with *cmultRepl* function using CZM method (count zero multiplicative) and an output format in p-counts (pseudo-counts). Then, the abundance table was transformed in centered log-ratio using *clr* function. Principal component analyses (PCA) were done based on total clustered biodiversity (4510 OTUs) using *factoextra* v 1.0.7 (Kassambara and Mundt 2019). ANOSIM test was used to do multivariate comparison on microbial communities (*anosim* function from *vegan* v 2.5–6; Oksanen *et al.* 2020). Heatmaps were done using *pheatmap* v 1.0.12 (Kolde 2019) and *ggdendro* v 0.1–20 (de Vries and Ripley 2020) packages, and Venn diagrams were obtained using *VennDiagram* package v 1.6.0 (Chen and Boutros 2011). Other univariate and multivariate statistics were performed using PAST v 3.07 (Hammer, Ryan and Harper 2001) and SIMCA 16.0 (Sartorius Stedim Biotech).

RESULTS

Meteorological context

The main meteorological characteristics pertaining to sample collection are summarized in Table 1 and Figure S2 (Supporting Information). Sampling was conducted during fall and winter; at three occasions designated as events *a*, *b* and *c*, clouds and precipitation could be collected simultaneously or within 1 day.

Samples were all collected at positive ambient temperature as liquids, at the exception of the snow sample collected at nearly −6°C. Due to the difference of elevation of almost

800 m between the two sampling sites, ambient meteorological conditions were colder and windier during cloud sampling (5.3 ±2.1°C and 9.6 ±6.1 m/s) than during rain collection (10.6°C ±4.8°C and 2.1 ±1.8 m/s). Rainfall events of different intensities were sampled, from light and moderate with average rates < 2.0 mm/h, such as event *a*, to heavy rains with maxima up to 24.0 mm/h, such as events *b* and *c*.

Based on the meteorological model ERA5, the boundary layer (BL) top altitude ranged from 441 to 1815 m a.s.l. on the days of rain collection and from 511 to 1728m a.s.l. during cloud sampling (Figure S2, Supporting Information). All rainfalls were thus collected at least partly from within the BL, while clouds and snow were sampled in the free troposphere.

Backward trajectory plots of the air masses associated with the samples indicate a wide range of geographical origins, with a general predominance of a large Western area (Atlantic Ocean; Table S1 and Figure S3, Supporting Information). The geographical zones over which the air masses traveled at low altitude within the planetary boundary layer can be considered probable source areas of the material collected; these are shown in Figure S3 (Supporting Information). In most cases, the air masses at cloud altitude spent most of the last 72 h preceding sampling in the free troposphere over marine areas (Fig. 1). In event *a*, the air masses traveled almost exclusively at high altitude over the Atlantic Ocean before reaching the sampling sites. Event *b*'s air masses traveled at low altitude over the Channel Sea and south Great Britain Island, 500 km from there. Contrasting with other situations, event *c*'s air masses originated from South-East and were issued from continental regions in northern Africa and the Saharan desert (Fig. 2).

Chemical signature

The main inorganic dissolved ions were present at micromolar concentrations in both clouds and precipitation, within ranges typical for atmospheric water samples at these sites (~1–100 μM; Deguillaume *et al.* 2014; Pouzet *et al.* 2017; Fig. 3A, Table S2, Supporting Information). Some of them varied together (Spearman's correlations, $P < 0.05$; Table S3, Supporting Information), illustrating similar sources: Na^+, Cl^-, Mg^{2+}, SO_4^{2-} and K^+ on one side, reflecting oceanic sources (Deguillaume *et al.* 2014) in good agreement with air masses' history, and to a lesser extent ($P < 0.1$) NH_4^+ and NO_3^- on the other side, indicative of continental and agricultural inputs (Mosier 2001; Almaraz *et al.* 2018).

Oceanic sources-related ions were in average all significantly more concentrated in cloud water than in rain (Mann–Whitney test, $P < 0.05$, Fig. 3A), as observed in the past at the same sampling sites (Pouzet *et al.* 2017), by median factors of ~47, ~26 and ~12 for Na^+, Cl^- and K^+, respectively. However, there were large variations depending on samples, in relation with rainfall intensity: in event *a* (light rainfall), Na^+ concentrations in rain and cloud water were similar, while in events *b* and *c* (heavy and moderate rain), this was diluted in rain compared with cloud water, by factors of 51 and 25, respectively (see Fig. 3B). In turn, rain samples were characterized by relatively high contributions of Ca^{2+}, NH_4^+ and NO_3^- attesting of continental influence. In particular, NH_4^+ and to a lesser extent Ca^{2+} tended to be more concentrated in rain than in clouds by median factors of 3.6 and 1.2 respectively, supporting consequent inputs from below cloud scavenging for these compounds. Here again, event *a*'s cloud and rain samples were more even than in events *b* and *c*, indicating lower scavenging.

In our dataset, pH ranged from 4.62 to 6.82. Acidity increased with decreasing oceanic influence (Spearman's correlations between pH and Na^+ or Cl^-, $P < 0.05$; Table S3A, Supporting Information), so rain water tended to be more acidic than clouds.

Biomass and ATP contents

Cell number concentration in the samples ranged from 2.67×10^3 to 2.81×10^4/mL in clouds and from 4.30×10^3 to 1.51×10^5/mL in precipitation (Table 2); these are typical values at these sites (Vaïtilingom *et al.* 2012; Pouzet *et al.* 2017). The corresponding average wet deposition fluxes of bacteria with rain, as inferred from precipitation rates, span over nearly two orders of magnitude from 1.07×10^7 to 6.63×10^8 cells/m^2/h^1. Although not significant due to high variability, cell number concentration tended to be higher in precipitation than in cloud water (Mann–Whitney test, $P = 0.23$; Fig. 3A); in the events *a*, *b* and *c*, cell numbers in rain were ~4–10 times higher than in clouds. Consistently, cell number concentration was overall negatively correlated with oceanic inputs (Spearman's correlations, *P*-values < 0.05; Table S3A, Supporting Information) and, in precipitation, positively with Ca^{2+} concentration (*P*-value < 0.01; Table S3B, Supporting Information) as observed earlier on other sampling sites (Christner *et al.* 2008). Multivariate analysis (PLS) specified the positive impact of the time spent by air mass over continental areas on cell concentration (see model coefficients in Table S4, Supporting Information).

The raw ATP concentration varied from 340 to 842 pmol/L in cloud water and from 155 to 1405 pmol/L in rain (Table 2), with no statistical difference. These values are consistent with previous reports in clouds at PUY, averaging 410 pmol/L over 28 samples (Vaïtilingom *et al.* 2012). In rainfall, ATP and cell concentrations were strongly positively linked (Spearman's correlation, *P*-value = 0.01, Spearman's $r = 0.89$; Table S3B, Supporting Information), suggesting a large proportion of viable cells. However, the ATP-to-cell ratio was significantly higher in cloud water than in rain, by a factor of ~8.2 [$(4.1–10.6) \times 10^{-6}$ pmol ATP/cell in rain vs. $(12.2–127.2) \times 10^{-6}$ pmol ATP/cell in cloudwater] (Mann–Whitney test, *P*-value < 0.01; Fig. 3A).

Bacterial diversity

A total of 4510 distinct prokaryotic OTUs (4507 bacteria and three archaea) were detected: 246 to 600 in rain, 174 to 1173 in clouds and 154 in snow (Table 2 and see complete list in Table S5, Supporting Information). These were distributed over 23 distinct phyla, 88 orders and 435 genera whose respective distributions among clouds (363 genera) and rain samples (277 genera) are shown in Figure S4 (Supporting Information).

Bacteria species richness (number of distinct OTUs) tended to be higher in clouds than in precipitation (Table 2; Fig. 3; Figure S5, Supporting Information), except for event *a* which exhibited particularly low richness likely in relation with air mass history (Fig. 1). Richness varied independently from biomass (Spearman's correlation; *P*-values > 0.05; Table S3, Supporting Information). Rather, this increased with the time spent by air masses at low altitude (PLS model coefficients, Table S4, Supporting Information) and in rain, this decreased with the time spent by the air mass nearby the sampling sites (<50 km) before sampling, supporting a distant origin of a large fraction of the richness.

The most represented phyla in the dataset (in terms of read numbers) were Proteobacteria (in particular the orders Betaproteobacteriales [eq. Burkholderiales], Pseudomonadales, Sphingomonadales and Rhizobiales), Actinobacteria (Micrococcales, Corynebacteriales and Frankiales), Firmicutes (Bacillales and Lactobacillales), Bacteroidetes and Deinococcus-Thermus.

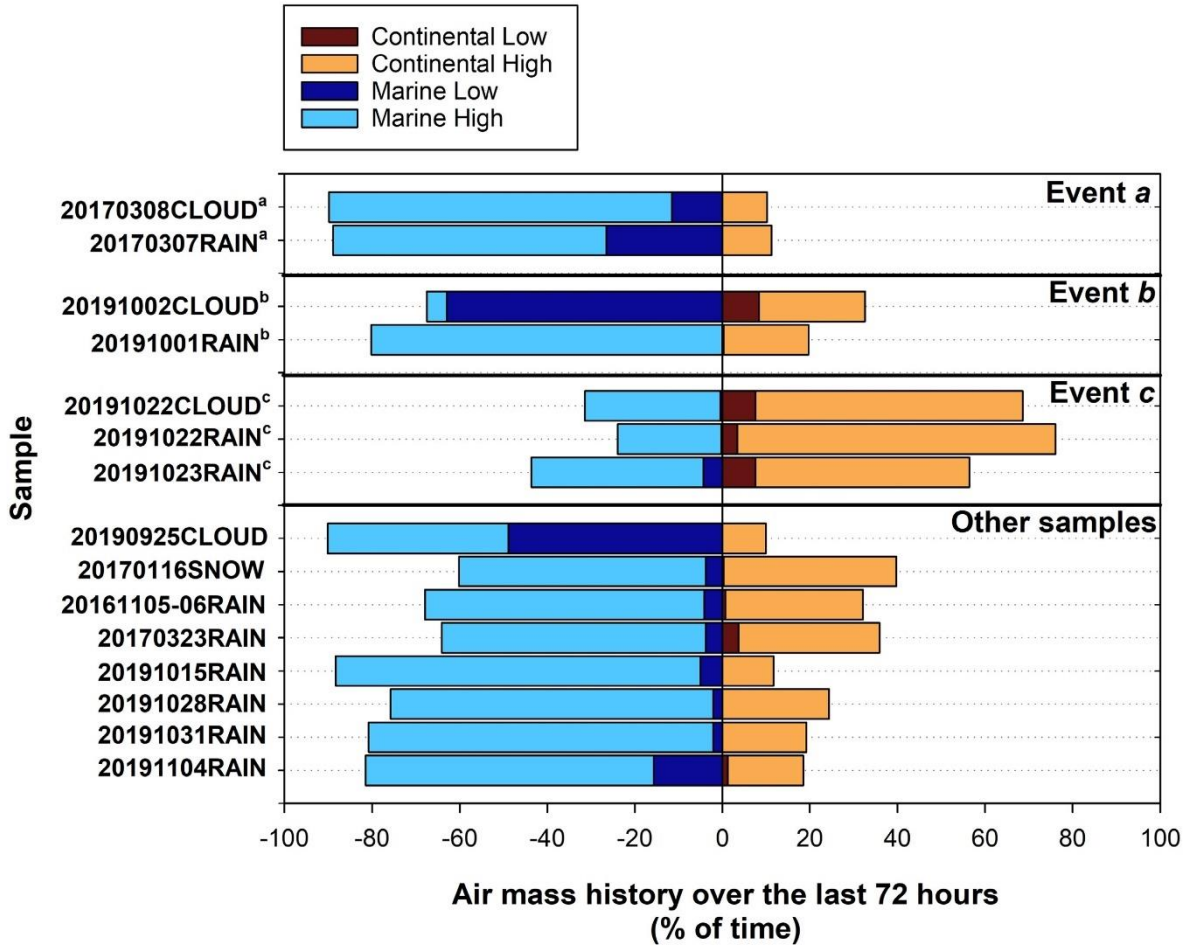


Figure 1. Air masses' history over the 72 h preceding sampling, expressed as % of the time spent over continental (right, brown) or marine (left, blue) surfaces, in the free troposphere (High) or in the planetary boundary layer (Low). Data were extracted from ERA5 backward trajectory plots at cloud attitude for rain samples or PUY altitude for clouds and snow. The % of time spent over marine surfaces is shown as negative values for graphical representation. [a]Superscripted letters indicate chronological associations between cloud and rain samples.

These microorganisms are commonly observed in atmospheric samples (e.g. Amato *et al.* 2017a; and references therein). Bacteria community composition varied from one sample to another, as shown at in Fig. 4 for the whole bacteria community and in Figure S6 (Supporting Information) for Proteobacteria, Firmicutes and Actinobacteria. Hierarchical clustering and multivariate analyses (PCA) based on community structure (Figure S7, Supporting Information) allowed distinguishing clouds from precipitation, although 20170308CLOUD[a] tended to resemble precipitation compared to other cloud samples.

In clouds, the common core of bacteria (i.e. the taxa detected in each cloud) was composed of 20 genera (Table S6, Supporting Information). These comprised a small fraction of total cloud's richness (5% of the 363 genera) but large proportions of the reads in each sample (42–79%). Most taxa (215 genera, i.e. 59%) were common to at least two samples (Figure S4, Supporting Information). There were, logically, more specific taxa in the richest sample (20190925CLOUD, 129 specific genera) than in the poorest (20170308CLOUD[b]; four specific genera; Figure S8, Supporting Information).

In rain, although more samples were collected, still numerous (33%) of the 277 genera detected were sample-specific (Figure S4, Supporting Information). This reflects the high biological variability in the atmospheric environment. As for clouds, the bacterial common core of rain samples was relatively limited in richness, with 15 genera that largely predominated (Table S6, Supporting Information; 59–92% of the reads).

At the genus level, about half of the bacteria detected in the study were common to clouds and rain samples (i.e. these were present in at least one cloud and one rain sample), and 155 (36%) and 69 (16%) were specific of either clouds or rain, respectively (Fig. 5). A large proportion of rain's richness was thus contained in clouds (208 genera out of 277, 75%), whereas cloud's richness was not fully retrieved (57%) in the rain samples (Table S7, Supporting Information).

In the associated cloud-rain samples events *a* and *b*, ~21% of the total genera were common to the two types of environment (Fig. 5). In event *c*, even at the OTU level (i.e. species) cloud and rain samples were remarkably similar, with as high as 42% of rain's OTU comprised in the corresponding cloud, vs. 4–11% for samples collected at 1-d interval from same air masses (events

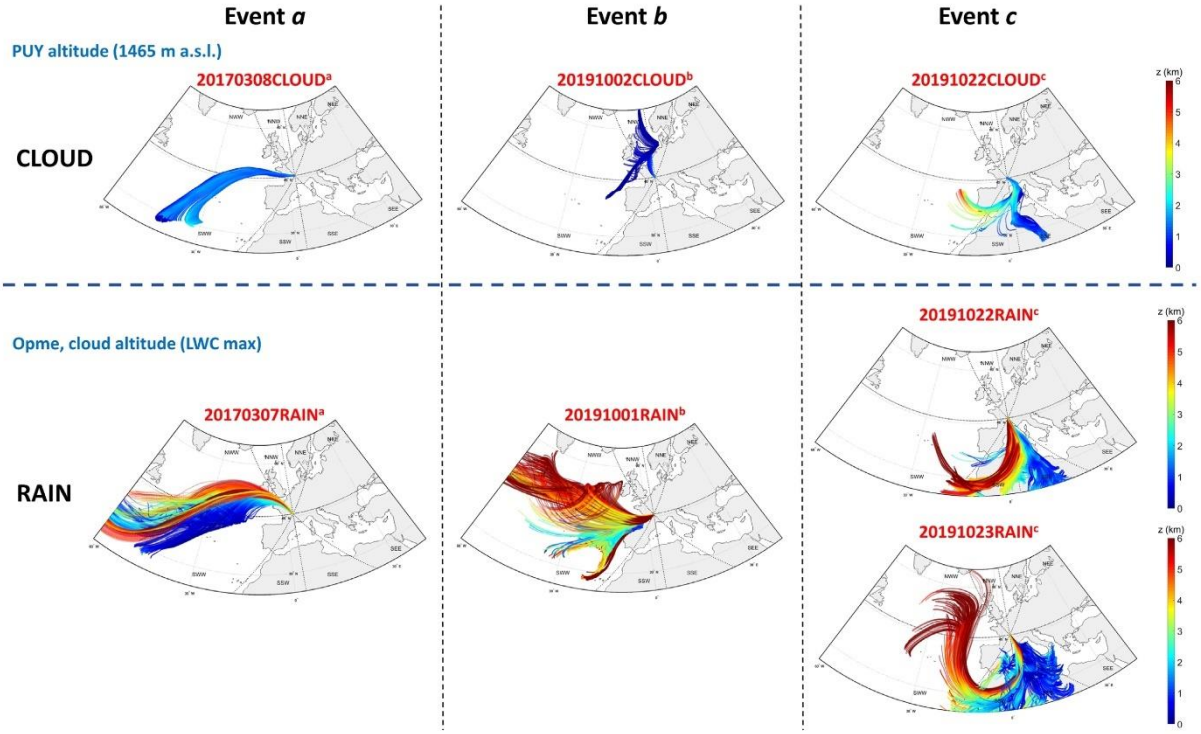


Figure 2. A 72-h backward trajectory plots at cloud level, extracted from ERA5 data reanalysis, for the associated cloud and rain samples in events *a*, *b* and *c*. See Figure S3 (Supporting Information) for other samples and the potential corresponding source areas identified.

Table 2. Biological characteristics of the samples.

SampleID*	Bacteria cell number concentration (cells/mL of water)	Bacteria wet deposition rate (cells/m^2/h)	ATP concentration (pmol/L)	Bacteria richness (number of distinct OTUs)
SNOW				
20170116SNOW	3.29×10^4	-	NA**	154
RAIN				
20 161105–06RAIN	4.30×10^3	1.08×10^7	NA**	246
20170307RAIN[a]	1.19×10^4	1.54×10^7	NA**	582
20170323RAIN	1.12×10^4	3.26×10^7	NA**	600
20191001RAIN[b]	1.51×10^5	6.63×10^8	1405	264
20191015RAIN	1.84×10^4	7.91×10^7	155	380
20191022RAIN[c]	1.41×10^5	2.11×10^8	673	586
20191023RAIN[c]	6.70×10^4	1.81×10^8	271	531
20191028RAIN	8.14×10^4	1.22×10^8	431	433
20191031RAIN	6.64×10^4	8.64×10^7	443	393
20191104RAIN	2.01×10^4	3.83×10^7	214	529
Average rain	5.72×10^4	1.44×10^8	513.1	454
Standard error rain	5.41×10^4	1.95×10^8	430.0	132
CLOUDS				
20170308CLOUD[a]	2.67×10^3	-	340	174
20190925CLOUD	1.36×10^4	-	682	1140
20191002CLOUD[b]	1.43×10^4	-	842	1019
20191022CLOUD[c]	2.81×10^4	-	344	1173
Average clouds	1.46×10^4	-	552.1	877
Standard error clouds	1.04×10^4	-	251.5	473

*Superscripted letters indicate chronological associations between cloud and rain samples;
NA**: No data available.

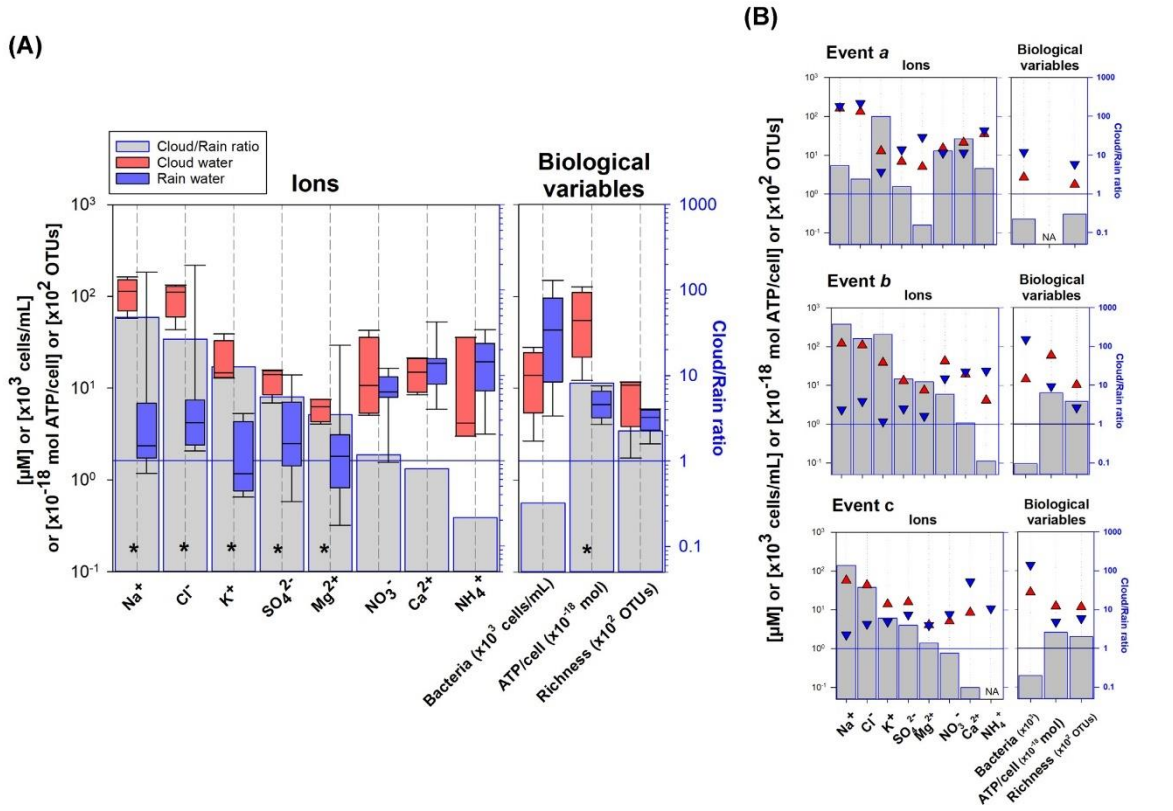


Figure 3. Absolute concentrations (left axes) and median ratios (right axes) of the main ions and biological variables in cloud and rain water samples, (as medians, 10th, 25th, 75th and 90th percentiles). (A) all samples together; (B) events *a*, *b* and *c*. Note the logarithmic scales on both y-axes. Asterisks indicate significant differences between concentrations in clouds and in rain (Mann–Whitney test, $P < 0.05$).

a and *c*(2)), and 3% when the air mass origin in addition slightly differed (event *b*; Table S7, Supporting Information). A total of six identified genera, among the most abundant, were detected in all cloud and rain samples (see Table S6, Supporting Information): *Massilia*, *Noviherbaspirillum*, *Pseudomonas*, *Sphingomonas*, *Undibacterium* (Proteobacteria) and *Rhodococcus* (Actinobacteria); with the exception of the latter, these were also all present in snow and thus composed the common bacterial core of this study. In turn, numerous less abundant or rare taxa were specific of clouds or rains, such as *Aerococcus*, *Oceanobacillus* and *Ammoniphilus* in clouds, and *Hymenobacter*, *Granulicella* and *Variovorax* in rain.

Certain microbial taxa were significantly more represented in clouds than in rain samples, and conversely (Kruskal–Wallis, $P \leq 0.05$). Those included in the common core of clouds and/or of rain samples are indicated in Fig. 6 (complete list in Table S8, Supporting Information). In particular, *Undibacterium* (Proteobacteria), *Bacillus* and *Staphylococcus* (Firmicutes) were more represented in clouds than in precipitation samples, while notably *Massilia*, *Sphingomonas* (Proteobacteria), as well as *Rhodococcus*, *Curtobacterium* and *Frondihabitans* (Actinobacteria) were more represented in rain; *Pseudomonas*, the genus gathering most ice-nucleation active bacteria, tended to be more present in precipitation. In particular, in the events *a*, *b* and *c*, rainfall was mostly characterized by strong enrichments of *Pseudomonas* and *Sphingomonas*, *Massilia* and *Rhodococcus*, and *Massilia* and *Sphingomonas*, respectively, as compared with the corresponding source clouds (Table S5, Supporting Information).

Identified probable marine bacteria, yet at low abundance, were more often present in clouds than in rain, such as GL500.29 marine group, *Demequina* (Park *et al.* 2016), *Marinactinospora* (Tian *et al.* 2009), *Nitratireductor* (Kang, Yang and Lee 2009) or again *Oceanobacillus*, confirming the influence of distant sources.

DISCUSSION

Sources of biological material in atmospheric waters

We report a high bacterial richness with predominant taxa frequently observed in atmospheric samples, in particular members of Alpha, Beta and Gammaproteobacteria, Bacteroidetes, Firmicutes and Actinobacteria (Šantl-Temkiv *et al.* 2018; Tignat-Perrier *et al.* 2020). Bacteria related with vegetation dominated, including Pseudomonadales, Sphingomonadales, Burkholderiales, Rhizobiales and Actinomycetales (e.g. Jeger, Spence and Pathology 2001; Gnanamanickam 2007). Other abundant bacteria included groups often rather frequently associated with soil like Corynebacteriales, Clostridiales and Bacillales. Finally, bacteria frequently found in atmospheric samples but also pointed out as indicators of human sources (Barberán *et al.* 2015) were present, such as *Staphylococcus* and *Streptococcus*.

Modeling and experimental studies have showed that bacteria and other microorganisms can travel thousands of kilometers from their emission source and connect distant ecosystems (e.g. Burrows *et al.* 2009; Griffin *et al.* 2017; Leyronas *et al.* 2018). In our dataset, biomass and biodiversity were overall disconnected

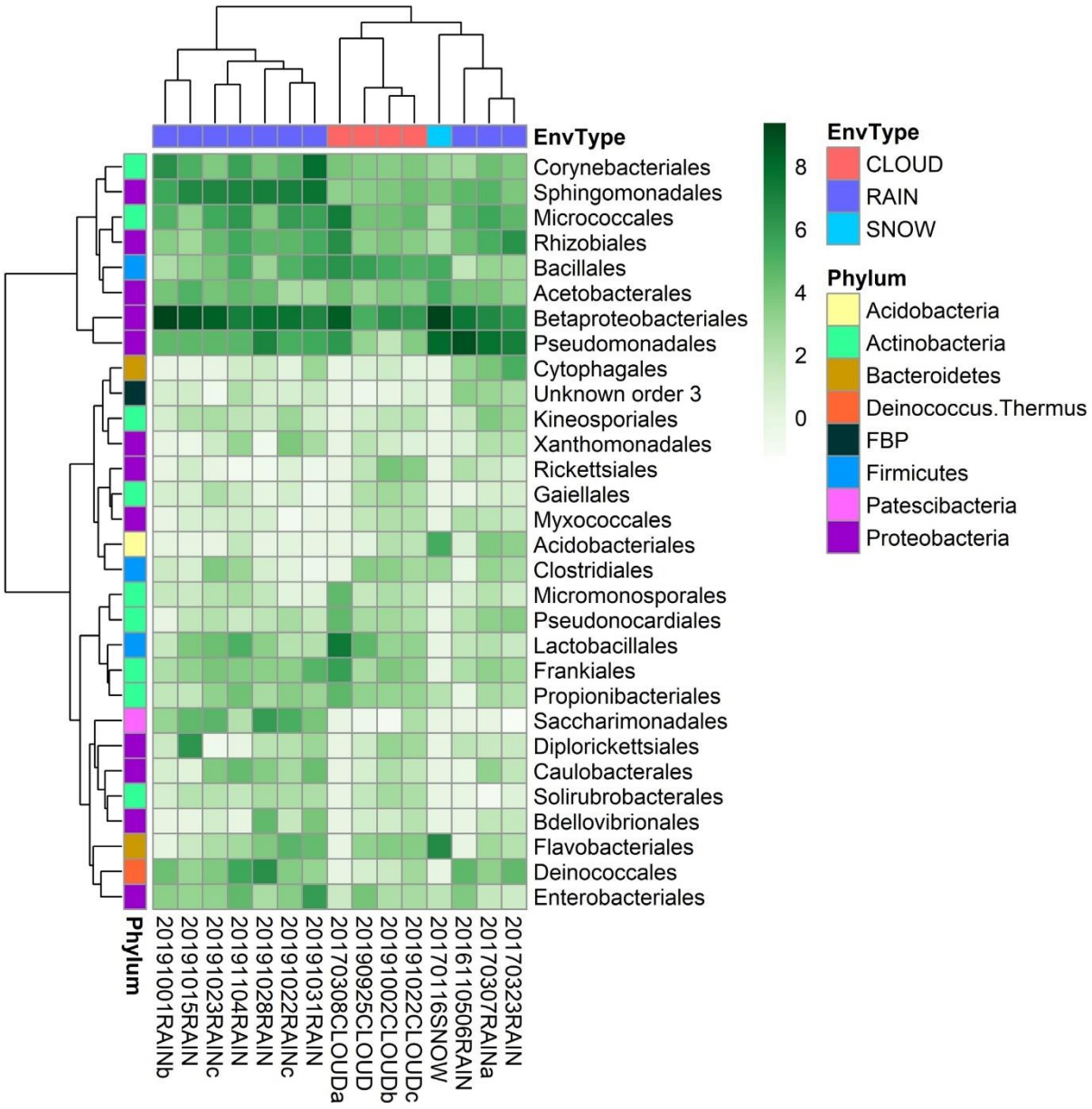


Figure 4. Distribution of the 30 most represented bacteria orders, and corresponding hierarchical clusterings (Ward's method). Intensity scale depicts centered-log ratio abundances; *a*, *b* and *c* letters indicate chronological associations between cloud and rain samples.

from each other, but these could be both explained for large parts by air mass history over the 3 days preceding sampling: the fraction of time spent over continental areas for biomass, and the fraction spent at low attitude for diversity, respectively. This confirms first that continental areas are much stronger sources of airborne bacteria than marine areas, and second that microbial material is recruited by air masses from a variety of surfaces that do not necessarily emit large amounts of material, due to low emission activity and/or small surface areas, but that can influence microbial diversity and contribute spreading rare biodiversity. These trends are well illustrated by the cloud sample in event *a*, which exhibited much lower richness and biomass than any other sample, including the precipitation associated with it. The corresponding air mass had very narrow and almost exclusively oceanic source area, and it traveled at high altitude for longer time than other cloud samples (see Figure S3, Supporting Information).

Clouds host higher metabolic activity than precipitation

ATP content in bacteria is typically in the order of ~10^{-6} pmol/cell (e.g. (Amato and Christner 2009)). Here, we observed higher ATP-per-cell content in clouds than in precipitation. This could indicate higher proportions of viable cells in clouds, and/or relatively higher metabolic activity supporting clouds as microbial habitats (Sattler, Puxbaum and Psenner

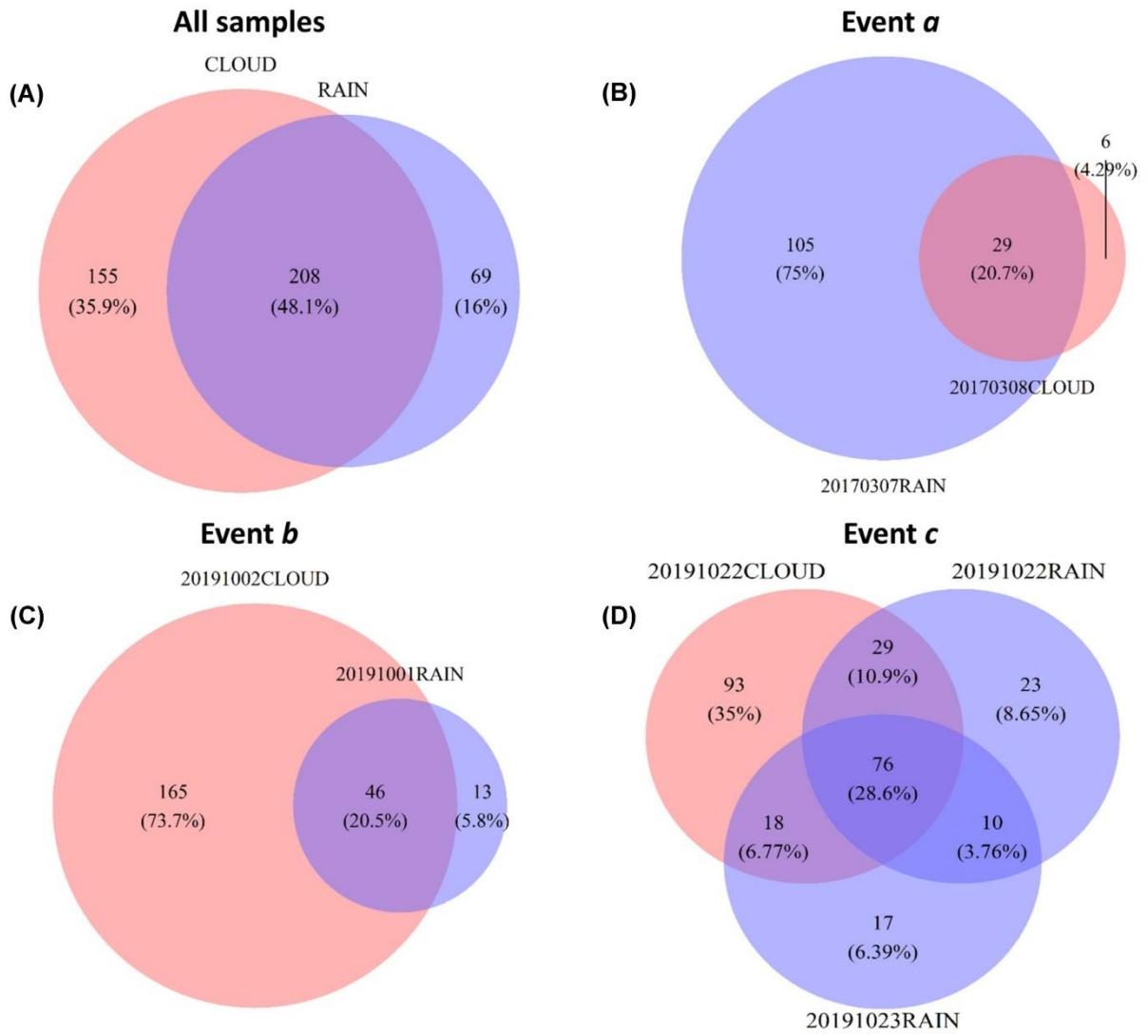


Figure 5. Venn diagrams depicting the distribution of distinct bacteria genera between cloud and rain samples: **(A)** all samples included; **(B)**, **(C)** and **(D)** for the associated cloud-rain samples in events *a*, *b* and *c*.

2001; Ervens and Amato 2020), with potential implication for atmospheric chemistry (Khaled *et al.* 2020). The stressful conditions existing in clouds for microbial cells, such as low temperatures and high H_2O_2 concentrations, could also be responsible for increased ATP contents in cells as reported from laboratory studies (Napolitano and Shain 2004; Amato and Christner 2009; Wirgot *et al.* 2017). Many of the bacteria detected in this study were reported earlier to be active in cloud water by examining their rRNA content (Amato *et al.* 2017b). These included both abundant and rare genera, such as *Acidiphilium*, *Sphingomonas*, *Pseudomonas*, *Rickettsia*, *Curtobacterium*, *Deinococcus* and many others. Although bacteria abundance and metabolic activity are often correlated in communities, rare bacteria can be even more active at the individual level and greatly contribute to the whole microbial activity in the ecosystem (Campbell *et al.* 2011).

Precipitation carries to the ground subsets of cloud microbial diversity and large amounts of biomass from the air column

Wet deposition fluxes of ~10^7 to nearly ~10^9 bacteria cells/m^2/h were quantified during rainfall periods. These are about one order of magnitude higher than the dry deposition fluxes reported on a daily basis above the boundary layer (Reche *et al.* 2018), which clearly confirms rainfall as major routes for airborne bacteria redeposition (Woo and Yamamoto 2020).

The biological similarity between cloud and rain samples was remarkably high, with ~75% of the bacteria genera present in rain also detected in clouds. This was noticeable even at deep taxonomic level in samples collected simultaneously, illustrating the strong connectivity between these consecutive steps of the water cycle. Beta-diversity can, thus, overall be interpreted

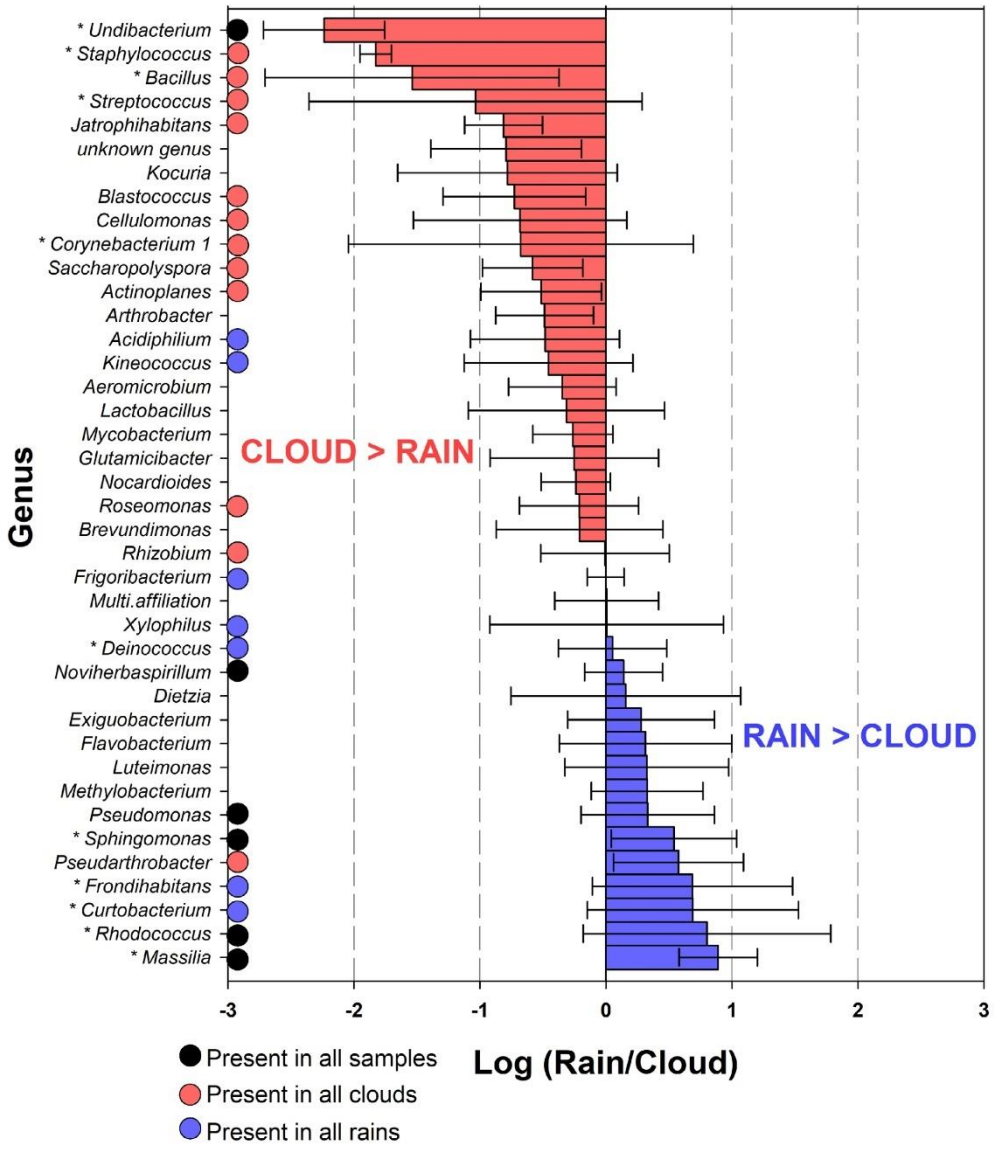


Figure 6. Average rain-to-cloud log-ratio representation in the events *a*, *b* and *c* for the 40 bacteria genera represented by >100 reads, out of the 135 distinct genera detected in total in this set of samples. The genera more represented in clouds than in rains are depicted in red, the genera more represented in rain in blue. Error-bars represent standard deviations from the mean rain-to-cloud ratio. The genera present in all the samples (common core), in all clouds and in all rain samples of the study are indicated by black, red and blue dots respectively. Asterisks on genera names indicate those whose representation was significantly different between clouds and rains, considering all the samples of the study (Kruskal–Wallis test, *P*-value < 0.05).

through the ecological concepts of nestedness, with precipitation carrying a proportion of cloud's richness as one can expect, and spatial turnover of taxa (Baselga 2010; Baselga and Leprieur 2015; Aho *et al.* 2020), illustrated by the numerous sample specific taxa. The high proportion of sample specific taxa in particular in rain is likely related at least in part with the boundary layer/free troposphere localization of the sampling locations, and with the influence of multiple local sources acting alternately in relation with air mass movements. It is well-documented indeed that the planetary boundary layer carries more material per unit volume than the free troposphere, and is more variable at small scale due to proximity with sources (Patton, Sullivan and Moeng 2005; Sasakawa *et al.* 2013).

Na^+ is emitted by marine sources and its relative contribution to the pool of dissolved ions can be used for tracking marine inputs to an air mass (Xiao *et al.* 2018). Assuming an origin in precipitation exclusively sourced in clouds at our sampling sites, we used the $[Na^+]_{cloud}$ to $[Na^+]_{rain}$ concentration ratio to track the proportion of material originating from the source cloud in rain, and normalize biomass and richness data. We infer that 25.5%, 0.2% and 0.8% of the bacteria cells in rain originated from the corresponding source cloud in events *a*, *b* and *c* respectively, so that by far that the largest fraction of the bacterial biomass was scavenged from the air column, in particular during intense rain events (*b* and *c*). In turn, the relative bacteria richness in rain water tended to decrease with the proportion of material

originating from the source cloud, pointing these latter as sources of diversity.

Based on aerosol scavenging efficiencies by raindrops, Moore and colleagues estimated that most (55–73%) of the bacteria cells present in rainfall, in terms of biomass, originate from the source cloud, in particular in light rainfall whose scavenging intensity is low (Moore *et al.* 2020). Our observations indicate higher proportions of scavenged biomass. The environmental context greatly differs between the continental mid-altitude area in France investigated in our study and that in Louisiana in (Moore *et al.* 2020), and parameters such as aerosol and drop numbers and size, rainfall intensity and else could contribute to large differences.

Biodiversity is not evenly distributed between clouds and precipitation

As for ions, bacteria taxa were not evenly distributed between clouds and rain: the bacteria more represented in precipitation were essentially known plant-associated taxa, consistently with previous observation that these are limited in their vertical atmospheric dispersal (Els *et al.* 2019a). Rain drops impacting the surface, i.e. grassland here, are themselves responsible for the emission of large amounts of biological aerosols (Huffman *et al.* 2013; Joung, Ge and Buie 2017), which could have greatly contributed to rain water composition.

Although different environmental situations were examined here, the data converged towards the depletion or enrichment of certain bacteria taxa in precipitation vs. the source clouds. Noteworthy, the bacteria assemblage in the snow sample resembled more precipitation than clouds, although this was collected at the cloud sampling site. This suggests the existence of some extent of environmental determinism in the distribution of taxa. Beside differential influences from the emission sources between clouds and precipitation, it seems thus legitimate to wonder whether specific phenotypic traits in bacteria could also contribute shaping their distribution.

Undibacterium and *Massilia*, the most representative bacteria of clouds and rain, respectively, along with *Noviherbaspirillum*, a member of the common core, are all Oxalobacteraceae. Oxalobacteraceae were also found persistent members of aerosols at high altitude (~10 km) over the tropics (DeLeon-Rodriguez *et al.* 2013). Noteworthy, in Sierra Nevada (Spain), *Massilia* and *Noviherbaspirillum* were also pointed out for their enrichment in rainfall compared with dry deposition (Triadó-Margarit *et al.* 2019). *Undibacterium* comprises oligotrophic bacteria recently described in clean water environments (Kämpfer *et al.* 2007; Eder *et al.* 2011). *Bacillus* and *Staphylococcus*, both dominant in clouds, have interestingly previously been reported viable at high altitudes in the dry stratosphere in several studies (Wainwright *et al.* 2003; Smith *et al.* 2018). The capacity of *Bacillus* to form spores undoubtedly improves its atmospheric persistence and dispersal (Smith *et al.* 2011). Additionally, the low-GC content in the genomes of these bacteria (Firmicutes) could favor their tolerance to such demanding environments as the high atmosphere and clouds (Foerstner *et al.* 2005; Mann and Chen 2010).

The possibility that bacteria could have avoided precipitation from cloud due to particular unidentified trait seems unreasonable. On the contrary, certain bacteria enriched in precipitation may have properties that could have favored their wet deposition. Phenotypic traits potentially related with bacteria's fate in the atmosphere have been proposed in the past. Proteobacteria in particular were found to be effective producers of biosurfactants, a factor mentioned as potentially favoring the integration of bacteria into cloud droplets (Renard *et al.* 2016). Besides, ice nucleation is probably the most cited biological process that could lead to a selective partitioning of bacteria in the atmosphere; this can initiate precipitation and so participate to the preferential wet deposition of plant associated ice-nucleation active bacteria (bioprecipitation), in particular Gammaproteobacteria such as *Pseudomonas syringae* (Morris, Georgakopoulos and Sands 2004). It was reported earlier that the most efficient ice nuclei were indeed enriched in rain compared with clouds and aerosols (Pouzet *et al.* 2017). Additionally, bacteria taxa known to include ice-nucleation active members were found overrepresented in the wet phases of the atmosphere compared with dry aerosols (Els *et al.* 2019b). Our observations are in good agreement, with Pseudomonadales also tending to be more represented in rain than in clouds and supporting the special relationship of these bacteria with the atmospheric water cycle.

CONCLUDING REMARKS

The data demonstrates that undoubtedly the atmosphere acts as a bacterial seed bank vehicle from high altitudes and probably distant environments to receptacle environments through the water cycle (Lennon and Jones 2011; Caporaso *et al.* 2012), thereby contributing to ecosystem microbial dynamics. Precipitation are in this regard increasingly prospected for novel potential biotechnologies (Sarmiento-Vizcaíno *et al.* 2018). The immigrant microorganisms can interact and compete with existing communities, and eventually colonize their new environment (Morris and Sands 2017). The constant carriage of new taxa, taxonomically close, to the surface via precipitation contributes to the spatial and temporal stability of ecosystems and tends to improve microbial fitness by spreading potentially beneficial and compatible biological innovations (Jalasvuori 2020).

The atmosphere is probably one of the most challenging environments to sample and analyse. As cloud altitude and the occurrence and localization of precipitation varied, not all precipitation samples could be associated with their source cloud and conversely. Inevitably, different sampling procedures have been deployed for prospecting clouds and precipitation, and the methods evolved within the time frame of this study which have contributed to differences in the datasets. Nevertheless, constant and meaningful trends emerged that could be related with emission sources and biological traits. We examined biodiversity in regard to chemical and meteorological contexts, and setup a coordinated sampling along the altitude gradient that allowed deciphering bacteria's fate along the first steps of the atmospheric water cycle, connecting high altitudes to surface environments. Prospecting environmental gradients in the highly variable atmospheric ecotone appears beneficial, if not necessary, to understand the dynamics and trends involving its microbiota.

SUPPLEMENTARY DATA

Supplementary data are available at FEMSEC online.

FUNDINGS

This work was supported by the French National Research Agency MOBIDIC project (ANR-17-MPGA-0013) to PA and the

French National Research Agency—German Research Foundation CHLOROFILTER project (DFG KE 884/10-1, DFG KO 2912/10-1, ANR-14-CE35-0005-01) to PA.

ACKNOWLEDGMENTS

We thank M. Brissy and C. Ghaffar for help in the field, L. Nauton for managing computer environment and T. Mas and V. Darbot for help with bioinformatics. We are grateful to the Mésocentre Clermont Auvergne University and AuBi platform for providing support, computing and storage resources and to ECMWF's computing and archive facilities. We also thank F. Conen for the loan of the automated rain collector. This study has been performed using CO-PDD instrumented site of the OPGC observatory and LaMP laboratory, under support of Université Clermont Auvergne, CNRS-INSU and CNES. Cloud sampling was performed in the frame of PUYCLOUD observation service.

Conflicts of interest. None declared.

REFERENCES

Aho KA, Weber CF, Christner BC *et al.* Spatiotemporal patterns of microbial composition and diversity in precipitation. *Ecol Monogr* 2020;**90**. DOI: 10.1002/ecm.1394.

Almaraz M, Bai E, Wang C *et al.* Agriculture is a major source of NOx pollution in California. *Sci Adv* 2018;**4**:eaao3477.

Amato P, Besaury L, Joly M *et al.* Metatranscriptomic exploration of microbial functioning in clouds. *Sci Rep* 2019;**9**:4383.

Amato P, Brisebois E, Draghi M *et al.* Main biological aerosols, specificities, abundance, and diversity. In: Delort A-M, Amato P (eds). *Microbiology of Aerosols*. John Wiley & Sons, Inc., 2017a, 1–21.

Amato P, Christner BC. Energy metabolism response to low-temperature and frozen conditions in *Psychrobacter cryohalolentis*. *Appl Environ Microbiol* 2009;**75**:711–8.

Amato P, Joly M, Besaury L *et al.* Active microorganisms thrive among extremely diverse communities in cloud water. *PLoS ONE* 2017b;**12**:e0182869.

Amato P, Joly M, Schaupp C *et al.* Survival and ice nucleation activity of bacteria as aerosols in a cloud simulation chamber. *Atmos Chem Phys* 2015;**15**:6455–65.

Amato P, Parazols M, Sancelme M *et al.* Microorganisms isolated from the water phase of tropospheric clouds at the Puy de Dôme: major groups and growth abilities at low temperatures. *FEMS Microbiol Ecol* 2007;**59**:242–54.

Baray J-L, Deguillaume L, Colomb A *et al.* Cézeaux-Aulnat-Opme-Puy De Dôme: a multi-site for the long-term survey of the tropospheric composition and climate change. *Atmos Meas Tech* 2020;**13**:3413–45.

Barberán A, Ladau J, Leff JW *et al.* Continental-scale distributions of dust-associated bacteria and fungi. *Proc Natl Acad Sci* 2015;**112**:5756–61.

Baselga A, Leprieur F. Comparing methods to separate components of beta diversity. *Method Ecol Evol* 2015;**6**:1069–79.

Baselga A. Partitioning the turnover and nestedness components of beta diversity. *Global Ecol Biogeogr* 2010;**19**:134–43.

Bauer H, Kasper-Giebl A, Löflund M *et al.* The contribution of bacteria and fungal spores to the organic carbon content of cloud water, precipitation and aerosols. *Atmos Res* 2002;**64**:109–19.

Blanco-Alegre C, Castro A, Calvo AI *et al.* Below-cloud scavenging of fine and coarse aerosol particles by rain: the role of raindrop size. *Q J R Meteorolog Soc* 2018;**144**:2715–26.

Bourcier L, Masson O, Laj P *et al.* A new method for assessing the aerosol to rain chemical composition relationships. *Atmos Res* 2012;**118**:295–303.

Bulgarelli D, Rott M, Schlaeppi K *et al.* Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* 2012;**488**:91–5.

Burrows SM, Butler T, Jöckel P *et al.* Bacteria in the global atmosphere – Part 2: modeling of emissions and transport between different ecosystems. *Atmos Chem Phys* 2009;**9**:9281–97.

Camacho-Sanchez M, Burraco P, Gomez-Mestre I *et al.* Preservation of RNA and DNA from mammal samples under field conditions. *Mol Ecol Resour* 2013;**13**:663–73.

Campbell BJ, Yu L, Heidelberg JF *et al.* Activity of abundant and rare bacteria in a coastal ocean. *Proc Natl Acad Sci* 2011;**108**:12776–81.

Caporaso JG, Paszkiewicz K, Field D *et al.* The Western English Channel contains a persistent microbial seed bank. *ISME J* 2012;**6**:1089–93.

Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 2011;**12**:35.

Christner BC, Cai R, Morris CE *et al.* Geographic, seasonal, and precipitation chemistry influence on the abundance and activity of biological ice nucleators in rain and snow. *Proc Natl Acad Sci* 2008;**105**:18854–9.

de Vries A, Ripley BD. Ggdendro: Create Dendrograms and Tree Diagrams Using "Ggplot2. https://cran.r-project.org/web/packages/ggdendro/index.html 2020.

Deguillaume L, Charbouillot T, Joly M *et al.* Classification of clouds sampled at the puy de Dôme (France) based on 10 yr of monitoring of their physicochemical properties. *Atmos Chem Phys* 2014;**14**:1485–506.

DeLeon-Rodriguez N, Lathem TL, Rodriguez-R LM *et al.* Microbiome of the upper troposphere: species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proc Natl Acad Sci* 2013;**110**:2575–80.

Després VR, Hufffman JA, Burrows SM *et al.* Primary biological aerosol particles in the atmosphere: a review. *Tellus B Chem Phys Meteorol* 2012;**64**. DOI: 10.3402/tellusb.v64i0.15598.

Eder W, Wanner G, Ludwig W *et al.* Description of *Undibacterium oligocarboniphilum* sp. nov., isolated from purified water, and *Undibacterium pigrum* strain CCUG 49012 as the type strain of *Undibacterium parvum* sp. nov., and emended descriptions of the genus Undibacterium and the species *Undibacterium pigrum*. *Int J Syst Evol Microbiol* 2011;**61**:384–91.

Els N, Baumann-Stanzer K, Larose C *et al.* Beyond the planetary boundary layer: bacterial and fungal vertical biogeography at Mount Sonnblick, Austria. *Geo Geogr Environ* 2019a;**6**. DOI: 10.1002/geo2.69.

Els N, Larose C, Baumann-Stanzer K *et al.* Microbial composition in seasonal time series of free tropospheric air and precipitation reveals community separation. *Aerobiologia* 2019b;**35**. DOI: 10.1007/s10453-019-09606-x.

Ervens B, Amato P. The global impact of bacterial processes on carbon mass. *Atmos Chem Phys* 2020;**20**:1777–94.

Escudié F, Auer L, Bernard M *et al.* FROGS: find, Rapidly, OTUs with Galaxy Solution. *Bioinformatics* 2018;**34**:1287–94.

Evans SE, Dueker ME, Logan JR *et al.* The biology of fog: results from coastal Maine and Namib Desert reveal common drivers of fog microbial composition. *Sci Total Environ* 2019;**647**:1547–56.

Foerstner KU, von Mering C, Hooper SD *et al.* Environments shape the nucleotide composition of genomes. *EMBO Rep* 2005;**6**:1208–13.

Gloor GB, Macklaim JM, Pawlowsky-Glahn V *et al.* Microbiome datasets are compositional: and this is not optional. *Front Microbiol* 2017;**8**:1–6.

Gnanamanickam SS. *Plant-Associated Bacteria*. Springer Science & Business Media, 2007.

Griffin D w., Gonzalez-Martin C, Hoose C *et al.* Global-scale atmospheric dispersion of microorganisms. In: Delort A-M, Amato P (eds). *Microbiology of Aerosols*. John Wiley & Sons, Inc., 2017, 155–94.

Hammer Ø, Ryan P, Harper D. PAST: paleontological Statistics software package for education and data analysis. *Palaeontol Electron* 2001;**4**:9.

Hervàs A, Camarero L, Reche I *et al.* Viability and potential for immigration of airborne bacteria from Africa that reach high mountain lakes in Europe. *Environ Microbiol* 2009;**11**:1612–23.

Hoffmann L, Günther G, Li D *et al.* From ERA-Interim to ERA5: the considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. *Atmos Chem Phys* 2019;**19**:3097–124.

Hou P, Wu S, McCarty JL *et al.* Sensitivity of atmospheric aerosol scavenging to precipitation intensity and frequency in the context of global climate change. *Atmos Chem Phys* 2018;**18**:8173–82.

Huffman JA, Pöhlker C, Prenni AJ *et al.* High concentrations of biological aerosol particles and ice nuclei during and after rain. *Atmos Chem Phys Discuss* 2013;**13**:1767–93.

Jaffrezo J-L, Colin J-L. Rain-aerosol coupling in urban area: scavenging ratio measurement and identification of some transfer processes. *Atmos Environ* 1988;**22**:929–35.

Jaffrezo JL, Calas N, Bouchet M. Carboxylic acids measurements with ionic chromatography. *Atmos Environ* 1998;**32**:2705–8.

Jalasvuori M. Silent rain: does the atmosphere-mediated connectivity between microbiomes influence bacterial evolutionary rates? *FEMS Microbiol Ecol* 2020;**96**. DOI: 10.1093/femsec/fiaa096.

Jeger MJ, Spence NJ. *Pathology BS for P. Biotic Interactions in Plant-Pathogen Associations*. CAB International, CABI, 2001.

Joung YS, Ge Z, Buie CR. Bioaerosol generation by raindrops on soil. *Nat Commun* 2017;**8**:14668.

Kämpfer P, Rosselló-Mora R, Hermansson M *et al.* Undibacterium pigrum gen. nov., sp. nov., isolated from drinking water. *Int J Syst Evol Microbiol*, 2007;**57**:1510–5.

Kang HS, Yang HL, Lee SD. Nitratireductor kimnyeongensis sp. nov., isolated from seaweed. *Int J Syst Evol Microbiol* 2009;**59**:1036–9.

Kassambara A, Mundt F. *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses Version 1.0.7 from CRAN*. 2019.

Khaled A, Zhang M, Amato P *et al.* Biodegradation by bacteria in clouds: an underestimated sink for some organics in the atmospheric multiphase system. *Atmos Chem Phys Discuss* 2020;**21**:1–32.

Kolde R. *Pheatmap: Pretty Heatmaps*. 2019.

Kozich JJ, Westcott SL, Baxter NT *et al.* Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl Environ Microbiol* 2013;**79**:5112–20.

Ladino L, Stetzer O, Hattendorf B *et al.* Experimental study of collection efficiencies between submicron aerosols and cloud droplets. *J Atmospheric Sci* 2011;**68**:1853–64.

Lennon JT, Jones SE. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat Rev Microbiol* 2011;**9**:119–30.

Leyronas C, Morris CE, Choufany M *et al.* Assessing the aerial interconnectivity of distant reservoirs of *Sclerotinia sclerotiorum Front Microbiol* 2018;**9**. DOI: 10.3389/fmicb.2018.02257.

Lindow SE, Arny DC, Upper CD. Distribution of ice nucleation-active bacteria on plants in nature. *Appl Environ Microbiol* 1978;**36**:831–8.

Luan T, Guo X, Zhang T *et al.* Below-cloud aerosol scavenging by different-intensity rains in Beijing city. *J Meteorol Res* 2019;**33**:126–37.

Mann S, Chen Y-PP. Bacterial genomic G+C composition-eliciting environmental adaptation. *Genomics* 2010;**95**:7–15.

Menke S, Gillingham MAF, Wilhelm K *et al.* Home-made cost effective preservation buffer is a better alternative to commercial preservation methods for microbiome research. *Front Microbiol* 2017;**8**. DOI: 10.3389/fmicb.2017.00102.

Mircea M, Stefan S, Fuzzi S. Precipitation scavenging coefficient: influence of measured aerosol and raindrop size distributions. *Atmos Environ* 2000;**34**:5169–74.

Möhler O, DeMott PJ, Vali G *et al.* Microbiology and atmospheric processes: the role of biological particles in cloud physics. *Biogeosciences* 2007;**4**:1059–71.

Moore RA, Hanlon R, Powers C *et al.* Scavenging of sub-micron to micron-sized microbial aerosols during simulated rainfall. *Atmosphere* 2020;**11**:80.

Morris CE, Georgakopoulos DG, Sands DC. Ice nucleation active bacteria and their potential role in precipitation. *Journal de Physique IV (Proceedings)* 2004;**121**:87–103.

Morris CE, Sands DC, Vinatzer BA *et al.* The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle. *ISME J* 2008;**2**:321–34.

Morris CE, Sands DC. Impacts of microbial aerosols on natural and agro-ecosystems: immigration, invasions and their consequences. In: *Microbiology of Aerosols*. Hoboken, NJ: John Wiley and Sons Inc, Delort, A.M. and Amato, P., 2017.

Mosier AR. Exchange of gaseous nitrogen compounds between agricultural systems and the atmosphere. *Plant Soil* 2001;**228**:17–27.

Napolitano MJ, Shain DH. Four kingdoms on glacier ice: convergent energetic processes boost energy levels as temperatures fall. *Proc Biol Sci* 2004;**271 Suppl 5**:S273–276.

Oksanen J, Blanchet FG, Friendly M *et al. Vegan: Community Ecology Package*. 2020.

Palarea-Albaladejo J, Martín-Fernández JA. zCompositions — R package for multivariate imputation of left-censored data under a compositional approach. *Chemom Intell Lab Syst* 2015;**143**:85–96.

Park S, Jung Y-T, Won S-M *et al.* Demequina litorisediminis sp. nov., isolated from a tidal flat, and emended description of the genus Demequina. *Int J Syst Evol Microbiol* 2016;**66**:4197–203.

Patton EG, Sullivan PP, Moeng C-H. The influence of idealized heterogeneity on wet and dry planetary boundary layers coupled to the land surface. *J Atmospheric Sci* 2005;**62**:2078–97.

Pouzet G, Peghaire E, Aguès M *et al.* Atmospheric processing and variability of biological ice nucleating particles in precipitation at Opme, France. *Atmosphere* 2017;**8**:229.

Quast C, Pruesse E, Yilmaz P *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;**41**:590–6.

Radke LF, Hobbs PV, Eltgroth MW. Scavenging of aerosol particles by precipitation. *J Appl Meteorol* 1980;**19**:715–22.

Reche I, D'Orta G, Mladenov N *et al.* Deposition rates of viruses and bacteria above the atmospheric boundary layer. *ISME J* 2018;**12**:1154–62.

Renard P, Bianco A, Baray J-L *et al.* Classification of clouds sampled at the Puy de Dôme Station (France) based on chemical measurements and air mass history matrices. *Atmosphere* 2020;**11**:732.

Renard P, Canet I, Sancelme M *et al.* Screening of cloud microorganisms isolated at the Puy de Dôme (France) station for the production of biosurfactants. *Atmos Chem Phys* 2016;**16**:12347–58.

Šantl-Temkiv T, Amato P, Gosewinkel U *et al.* High-flow-rate impinger for the study of concentration, viability, metabolic activity, and ice-nucleation activity of airborne bacteria. *Environ Sci Technol* 2017;**51**:11224–34.

Šantl-Temkiv T, Gosewinkel U, Starnawski P *et al.* Aeolian dispersal of bacteria in southwest Greenland: their sources, abundance, diversity and physiological states. *FEMS Microbiol Ecol* 2018;**94**. DOI: 10.1093/femsec/fiy031.

Sarmiento-Vizcaíno A, Espadas J, Martín J *et al.* Atmospheric precipitations, hailstone and rainwater, as a novel source of Streptomyces producing bioactive natural products. *Front Microbiol* 2018;**9**. DOI: 10.3389/fmicb.2018.00773.

Sasakawa M, Machida T, Tsuda N *et al.* Aircraft and tower measurements of CO2 concentration in the planetary boundary layer and the lower free troposphere over southern taiga in West Siberia: long-term records from 2002 to 2011. *J Geophys Res Atmos* 2013;**118**:9489–98.

Sattler B, Puxbaum H, Psenner R. Bacterial growth in supercooled cloud droplets. *Geophys Res Lett* 2001;**28**:239–42.

Schloss PD, Westcott SL, Ryabin T *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;**75**:7537–41.

Smith DJ, Griffin DW, McPeters RD *et al.* Microbial survival in the stratosphere and implications for global dispersal. *Aerobiologia* 2011;**27**:319–32.

Smith DJ, Ravichandar JD, Jain S *et al.* Airborne bacteria in Earth's lower stratosphere resemble taxa detected in the troposphere: results from a New NASA Aircraft Bioaerosol Collector (ABC). *Front Microbiol* 2018;**9**. DOI: 10.3389/fmicb.2018.01752.

Smith DJ, Timonen HJ, Jaffe DA *et al.* Intercontinental dispersal of bacteria and archaea by transpacific winds. *Appl Environ Microbiol* 2013;**79**:1134–9.

Sonwani S, Kulshrestha UC. PM10 carbonaceous aerosols and their real-time wet scavenging during monsoon and non-monsoon seasons at Delhi, India. *J Atmos Chem* 2019;**76**:171–200.

Tian X-P, Tang S-K, Dong J-D *et al.* Marinactinospora thermotolerans gen. nov., sp. nov., a marine actinomycete isolated from a sediment in the northern South China Sea. *Int J Syst Evol Microbiol* 2009;**59**:948–52.

Tignat-Perrier R, Dommergue A, Thollot A *et al.* Seasonal shift in airborne microbial communities. *Sci Total Environ* 2020;**716**:137129.

Triadó-Margarit X, Caliz J, Reche I *et al.* High similarity in bacterial bioaerosol compositions between the free troposphere and atmospheric depositions collected at high-elevation mountains. *Atmos Environ* 2019;**203**:79–86.

Vaïtilingom M, Attard E, Gaiani N *et al.* Long-term features of cloud microbiology at the puy de Dôme (France). *Atmos Environ* 2012;**56**:88–100.

Vaïtilingom M, Deguillaume L, Vinatier V *et al.* Potential impact of microbial activity on the oxidant capacity and organic carbon budget in clouds. *Proc Natl Acad Sci* 2013;**110**:559–64.

Wainwright M, Wickramasinghe NC, Narlikar JV *et al.* Microorganisms cultured from stratospheric air samples obtained at 41 km. *FEMS Microbiol Lett* 2003;**218**:161–5.

Waked A, Favez O, Alleman LY *et al.* Source apportionment of PM_{10} in a north-western Europe regional urban background site (Lens, France) using positive matrix factorization and including primary biogenic emissions. *Atmos Chem Phys* 2014;**14**:3325–46.

Weil T, Filippo CD, Albanese D *et al.* Legal immigrants: invasion of alien microbial communities during winter occurring desert dust storms. *Microbiome* 2017;**5**:32.

Willis PT, Tattelman P. Drop-size distributions associated with intense rainfall. *J Appl Meteorol* 1989;**28**:3–15.

Wirgot N, Vinatier V, Deguillaume L *et al.* H2O2 modulates the energetic metabolism of the cloud microbiome. *Atmos Chem Phys* 2017;**17**:14841–51.

Woo C, Yamamoto N. Falling bacterial communities from the atmosphere. *Environ Microbiome*, 2020;**15**:22.

Xiao H-W, Xiao H-Y, Shen C-Y *et al.* Chemical composition and sources of marine aerosol over the western North Pacific Ocean in winter. *Atmosphere* 2018;**9**:298.

Yoon SH, Ha SM, Kwon S *et al.* Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* 2017;**67**:1613–7.

SUPPLEMENTARY INFORMATION to:

Rainfalls sprinkle cloud bacterial diversity while scavenging biomass

Raphaëlle Péguilhan[1*], Ludovic Besaury[1], Florent Rossi[1], François Enault[2], Jean-Luc Baray[3,4], Laurent Deguillaume[3,4], Pierre Amato[1]

[1] Université Clermont Auvergne, CNRS, SIGMA Clermont, ICCF, F-63000 CLERMONT-FERRAND, France.
[2] Université Clermont Auvergne, CNRS, Laboratoire Microorganismes : Genome et Environnement, F-63000 CLERMONT-FERRAND, France.
[3] Université Clermont Auvergne, CNRS, Observatoire de Physique du Globe de Clermont-Ferrand, UMS 833, F-63000 CLERMONT-FERRAND, France.
[4] Université Clermont Auvergne, CNRS, Laboratoire de Météorologie Physique, UMR 6016, F-63000 CLERMONT-FERRAND, France.

List of supplements:

- **Supplementary information to Material and Methods:**

Cell counts
ATP quantification
DNA extraction and amplification

- **Supplementary Tables and Figures:**

Supplementary Table 1: Geographical origin of the air masses sampled.

Supplementary Table 2: Chemical characteristics of the samples.

Supplementary Table 3 (electronic .xlsx file): Spearman's correlation matrices including air mass history, chemical and biological variables (p-values at upper right/Spearman's r at lower left): **A-** all samples together; **B-** precipitation samples (including snow) only. Correlations with p-values <0.1 are bolded, and correlation with p-values < 0.05 are indicated in red.

Supplementary Table 4 (electronic .xlsx file): PLS coefficient list of chemical and environmental explicatory variables in the prediction of biological variables (biomass, richness, relative taxa abundances).

Supplementary Table 5 (electronic .xlsx file): Taxonomic affiliation of clustered sequences and corresponding read numbers in each sample.

Supplementary Table 6: Bacteria genera composing the common cores of clouds and rain samples. The genera present in all samples are bolded.

Supplementary Table 7: Distribution of bacteria richness at different taxonomic depths between cloud and rain samples, and in the particular cases of events *a*, *b* and *c*.

Supplementary Table 8 (electronic .xlsx file): Statistical comparison (Kruskal-Wallis test) of bacteria genera read abundances in clouds vs rain samples.

Supplementary Figure 1: Geographical configuration of the sampling sites. **A-** map of France; **B-** altitude profile between puy de Dôme Mountain' summit and Opme station (This figure was created using © Google Earth); **C-** Cloud sampling operations at PUY station, with high-flow-rate impingers and cloud droplet impactor pointed by the blue arrow; **D-** Rain collection at Opme station, with puy de Dôme Mountain visible at the background.

Supplementary Figure 2 (electronic .pdf file): Daily meteorological data for the sampling dates at puy de Dôme station (PUY, clouds and fresh snow sampling) and Opme station (OPME, rain sampling) organized by chronological order: **(a)** altitude profiles of cloud ice and liquid water contents (ERA5 ECMWF model); **(b)** Boundary layer height (ERA5 ECMWF model from Cézeaux station, 410 a.s.l.); **(c)** Main meteorological data at the sampling site (5 min intervals; T, RH and wind speed and direction at PUY for clouds and snow sampling, T and RH at OPME for rain sampling); **(d)** total precipitation from disdrometer measurements at OPME station (30 s intervals). Cloud sampling periods are framed in red.

Supplementary Figure 3 (electronic .pdf file): Seventy-two-hours backward trajectory plots and associated sources areas, extracted from ERA5 data reanalysis.

Supplementary Figure 4: Frequency distribution of bacteria genera in clouds (363 genera) and rain (277 genera).

Supplementary Figure 5: Rarefaction curves.

Supplementary Figure 6: Distribution of the 30 most represented bacteria genera in **a)** Proteobacteria, **b)** Firmicutes and **c)** Actinobacteria among the samples, and corresponding hierarchical clustering (Ward's method). Intensity scale depicts centered-log ratio abundances; *a*, *b* and *c* letters after sample names indicate chronological associations between cloud and rain samples.

Supplementary Figure 7: Principal component analysis plot based on OTU composition in the samples (4,510 OTUS).

Supplementary Figure 8: Venn diagrams depicting the distribution of distinct bacteria genera among cloud samples.

Supplement to Material and Methods

Cell counts

Briefly, triplicate subsamples of 450 µL were added with 50 µL of 5% glutaraldehyde (0.5% final concentration; Sigma-Aldrich G7651, St-Louis, MO, USA) and kept at 4°C before analysis, within one week from sampling. Just before analysis, samples were mixed with 1 vol. (500 µL) of 0.02 µm filtered Tris-EDTA pH 8.0 (40 mM Tris-Base, 1 mM EDTA, acetic acid to pH 8.0) and stained with SYBRGreen I (Molecular Probes Inc., Eugene, OR, USA) from a 100X solution. Cells counts were performed at excitation and emission wavelengths of λexc = 488 nm and λem = 530 nm, respectively, at a flow rate of ~70 µL min-1 further determined by weighting. The median standard deviation of cell number concentration in the samples was 4.8% of the mean concentration from triplicates.

ATP quantification

Triplicate subsamples of 50 µl were fixed with 50 µl of extractant B/S (ATP Biomass Kit HS; BioThema; Handen, Sweden). Luminescence was measured from 50 µl of the previous mix using a GloMax 20/20 luminometer (Promega, Madison, WI, USA), immediately after the injection of 200 µl of reconstituted ATP Reagent HS. Following manufacturer's recommendations, luminescence was recorded after a complementary injection of 10 µl of 10^{-7} mol/L ATP internal standard to account for the presence of inhibitors, and the actual ATP concentration in samples was calculated. The median standard deviation of ATP concentration from measurements on triplicate subsamples was 10.1% of the mean.

DNA extraction and amplification

Immediately after sampling, all samples were subsampled for complementary biological and chemical analyses, then filtered for DNA extraction. Samples collected in 2016-2017 were filtered on sterile 0.22 µm polyethersulfone (PES) filters (0.22 µm porosity, 47 mm diameter; MoBio 14880), using sterile 500 mL Nalgene filtration units. The filters were then cut in quarters using sterile scalpels and ¾ of these were transferred into the bead-beating tubes of the MoBio PowerWater® DNA Isolation Kit used for DNA extraction, then stored at -80°C until being further processed. Upon extraction following manufacturer's instructions, DNA was finally eluted into a final volume of 100 µL.

In 2019, DNA was extracted from mixed cellulose esters (MCE)-filtered samples (0.22 µm porosity, 47mm diameter; ClearLine 0421A00023) using Macherey-Nagel NucleoMag® DNA/RNA Water Kit and NucleoSpin Bead Tubes 5 mL Type A (Macherey-Nagel, 740799.50) added with 1 200 µL (cloud samples) or 900 µL (precipitation samples) of lysis buffer MWA1. For DNA extraction, bead-beating (10 min at maximum vortex speed) lysate volumes of 900 µL and 600 µL of precipitation and cloud sample, respectively, were processed following a protocol adapted for 47 mm filter membranes (section 7, p. 25 of the manufacturer's protocol). Lysates were then RNase-treated by adding 1:50 volume of a 12 mg/mL stock solution, and DNA was finally eluted into 50 µL RNase-free H_2O. DNA was quantified in the extracts using Quant-iT™ PicoGreen® dsDNA kit (Invitrogen).

For PCR amplification of the 16S sub-unit of bacterial ribosomal gene, each reaction of 50 µL contained ~1 ng of genomic DNA, 5 µL of 10X HiFi Buffer, 2 mM $MgSO_4$, 0.2 or 0.4 µM of

each primer, 200 µM dNTPs, and 1 unit of HiFi Taq DNA Polymerase Platinum. The conditions were as indicated in (Bulgarelli et al., 2012). Amplicons were purified using QIAquick gel extraction kit or QIAquick PCR Purification kit (Qiagen, Hilden, Germany) following manufacturer's instructions, quantified and equimolarly pooled at the total concentration of 5 pM for Illumina Miseq 2*250 bp by a subcontracting company (GenoScreen; Lille, France).

Tables and Figures

Supplementary Table 1: Geographical origin of the air masses sampled.

SampleID*	Air mass geographical origin (% of sector contribution over the last 72 hours)[¤]							
	NNE	**ENE**	**ESE**	**SSE**	**SSW**	**WSW**	**WNW**	**NNW**
SNOW								
20170116SNOW	29.5	0	0	0	0	0	30.4	38.4
RAIN								
20161105-06RAIN	1.9 - 0.5	0.2 - 0	0.1 - 0	1.8 - 0	8.6 - 3.1	10.0 - 32.3	57.7 - 53.4	14.5 - 8.7
20170307RAIN[a]	0 - 0	0 - 0	0 - 0	0 - 0	0 - 0	1.4 - 49.1	93.9 - 46.6	0 - 3.2
20170323RAIN	0 - 0	0 - 0	0 - 0	9.5 - 10.5	38.6 - 28.9	32.0 - 48.1	12.4 - 11.6	0 - 0
20191001RAIN[b]	0 - 0	0 - 0	0 - 0	0 - 0	0.9 - 0	86.2 - 64.8	10.1 - 34.1	0 - 0
20191015RAIN	0 - 0	0 - 0	0 - 0	0 - 0	0 - 0.3	19.0 - 15.0	78.6 - 83.1	0 - 0
20191022RAIN[c]	0 - 0	8.7 - 0	27.0 - 0	27.6 - 54.2	22.1 - 41.6	10.2 - 3.1	1.2 - 0	0 - 0
20191023RAIN[c]	0 - 0	0 - 0	93.0 - 6.8	3.8 - 42.9	0 - 31.2	0 - 10.3	0 - 7.9	0 - 0.2
20191028RAIN	0 - 0	0.3 - 0	1.0 - 0	1.4 - 0	5.4 - 0	73.3 - 83.4	0 - 14.0	0 - 0
20191031RAIN	0 - 0	0 - 0	0 - 0	0 - 0	0 - 0	47.4 - 52.4	49.8 - 44.4	0 - 0.8
20191104RAIN	0 - 0	0 - 0	0 - 0	0 - 0	0.2 - 0	59.5 - 26.4	38 - 72.5	0 - 0.2
CLOUDS								
20170308CLOUD[a]	0	0	0	0	0	57.6	41.1	0
20190925CLOUD	0	0	0	0	0	19.2	79.5	0
20191002CLOUD[b]	24.2	0	0	0	0	7.6	10.4	55.4
20191022CLOUD[c]	0	0	21.4	65.2	8.6	2.9	0	0

[¤] Based on CAT backward trajectory plots starting at the sampling site: for rain (ground level - cloud level).

Supplementary Table 2: Chemical characteristics of the samples.

SampleID*	pH	Cations (µM)					Anions (µM)		
		Na^+	NH_4^+	K^+	Mg^{2+}	Ca^{2+}	Cl^-	NO_3^-	SO_4^{2-}
SNOW									
20170116SNOW	6.82	35.68	6.90	1.28	0.33	17.23	6.64	12.07	0.55
RAIN									
20161105-06RAIN	6.75	3.28	19.43	0.82	0.32	5.88	4.20	9.11	1.54
20170307RAIN[a]	5.95	183.84	43.49	3.76	29.32	11.60	217.91	11.48	13.97
20170323RAIN	5.57	9.81	37.94	0.65	2.21	10.60	9.93	16.56	5.07
20191001RAIN[b]	4.75	2.36	23.74	1.17	1.61	22.58	3.90	15.13	2.50
20191015RAIN	4.62	3.89	22.44	5.31	2.18	19.23	4.97	8.61	6.61
20191022RAIN[c]	5.23	2.26	10.82	4.90	4.01	52.56	4.30	7.63	7.46
20191023RAIN[c]	5.45	NA**	NA**	NA**	NA**	NA**	NA**	NA**	NA**
20191028RAIN	4.98	1.89	10.87	1.79	1.81	18.82	2.17	9.65	1.96
20191031RAIN	4.66	1.18	7.94	0.70	1.02	18.62	2.06	8.36	1.31
20191104RAIN	4.70	1.58	3.15	0.90	0.62	12.27	2.62	1.55	0.58
Average rain	**5.27**	**23.34**	**19.98**	**2.22**	**4.79**	**19.13**	**28.01**	**9.79**	**4.56**
Standard error rain	**0.69**	**60.24**	**13.65**	**1.90**	**9.26**	**13.61**	**71.25**	**4.38**	**4.30**
CLOUDS									
20170308CLOUD[a]	6.22	162.42	35.64	12.88	5.05	21.43	133.45	15.21	6.92
20190925CLOUD	5.52	105.35	2.99	15.38	7.56	10.58	114.23	6.14	14.79
20191002CLOUD[b]	5.15	120.49	4.18	38.92	7.48	19.32	109.02	42.60	13.22
20191022CLOUD[c]	5.80	57.43	NA**	13.98	4.07	8.51	43.59	5.07	15.78
Average clouds	**5.67**	**111.42**	**14.27**	**20.29**	**6.04**	**14.96**	**100.07**	**17.25**	**12.68**
Standard error clouds	**0.45**	**43.34**	**18.52**	**12.46**	**1.76**	**6.37**	**39.10**	**17.50**	**3.98**
cloud/rain average concentration ratio	1.08	4.77[s]	0.71	9.13[s]	1.26[s]	0.78	3.57[s]	1.76	2.78[s]

* Superscripted letter indicate chronological associations between clouds and precipitation (events *a*, *b* and *c*); NA** No data available; [s] Significant difference between clouds and rain (Mann-Whitney test, $p<0.05$).

Supplementary Table 3 (electronic .xlsx file): Spearman's correlation matrices including air mass history, chemical and biological variables (p-values at upper right/Spearman's r at lower left): **A-** all samples together; **B-** precipitation samples (including snow) only. Correlations with p-values <0.1 are bolded, and correlation with p-values < 0.05 are indicated in red. The variable "Near" relates to the fraction of time spent by the air mass within 50 km around the sampling site over the 72 hours preceding sampling.

Supplementary Table 4 (electronic .xlsx file): Scaled and centered PLS coefficient list of chemical and environmental explicatory variables in the prediction of biological variables (biomass, richness, relative taxa abundances).

Supplementary Table 5 (electronic .xlsx file): Taxonomic affiliation of clustered sequences and corresponding read numbers in each sample.

Supplementary Table 6: Bacteria genera composing the common cores of clouds and rain samples (checked in black). The genera present in all samples are bolded.

Genus	Cloud core	Rain core
Acidiphilium		■
Acidovorax		■
Actinoplanes	■	
Bacillus	■	
Blastococcus	■	
Cellulomonas	■	
Corynebacterium	■	
Corynebacterium 1	■	
Curtobacterium		■
Cutibacterium	■	
Deinococcus		■
Frigoribacterium		■
Frondihabitans		■
Geodermatophilus		■
Jatrophihabitans	■	
Kineococcus		■
Massilia	■	■
Noviherbaspirillum	■	■
Pantoea	■	
Pseudarthrobacter	■	
Pseudomonas	■	■
Rhizobium	■	
Rhodococcus	■	■
Roseomonas	■	
Saccharopolyspora	■	
Sphingomonas	■	■
Staphylococcus	■	
Streptococcus	■	
Undibacterium	■	■
Xylophilus		■

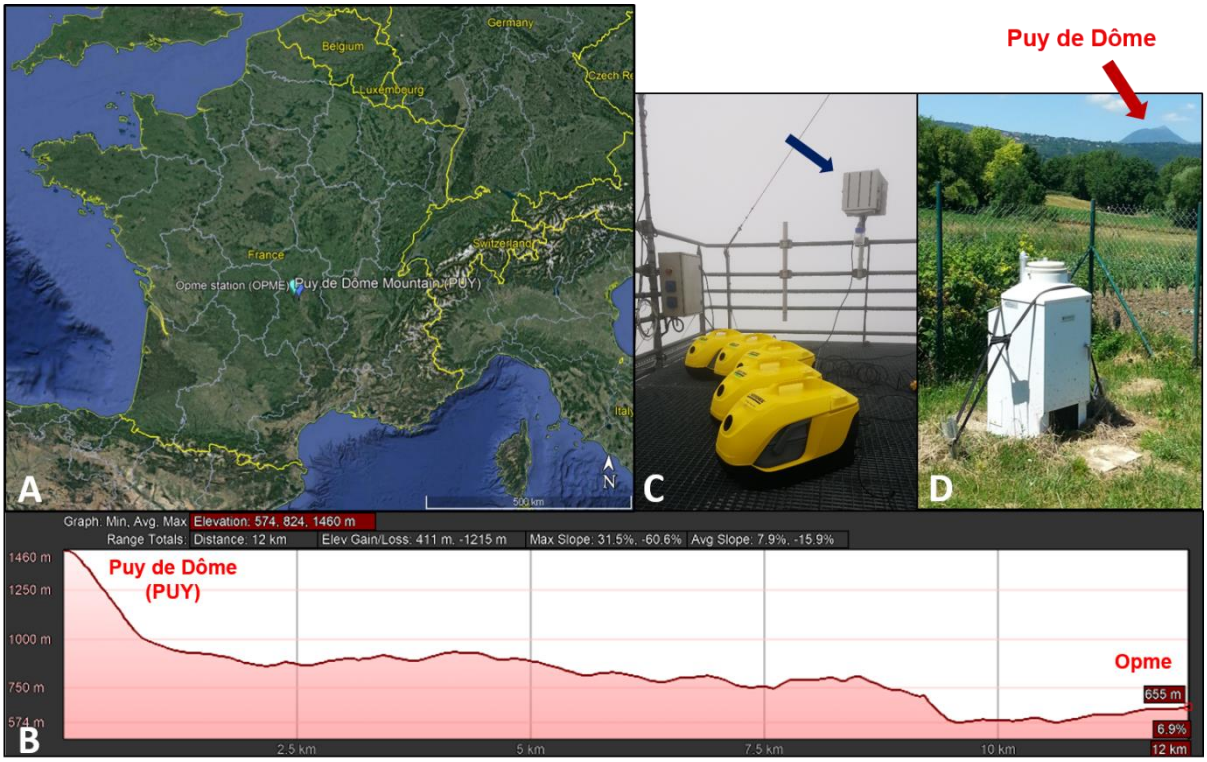
Supplementary Table 7: Distribution of bacteria richness at different taxonomic depths between cloud and rain samples, and in the particular cases of events *a*, *b* and *c*.

	All samples				**Event *a*** (20170307RAIN & 20170308CLOUD) (1-day interval. same air mass)				**Event *b*** (20191001RAIN & 20191002CLOUD) (1-day interval. distinct air mass)				**Event *c (1)*** (20191022RAIN & 20191022CLOUD) (Same day. same air mass)				**Event *c (2)*** (20191023RAIN & 20191022CLOUD) (1-day interval. same air mass)			
	Cloud	**Rain**	**Core**	**Total**	**Cloud**	**Rain**	**Core**	**Total**	**Cloud**	**Rain**	**Core**	**Total**	**Cloud**	**Rain**	**Core**	**Total**	**Cloud**	**Rain**	**Core**	**Total**
Species*	1784	1792	851	**4427**	107	515	67	**689**	1012	257	7	**1276**	924	337	249	**1510**	1151	509	22	**1682**
Proportion of total OTUs (%)	40.30	40.48	19.22	**100**	15.53	74.75	9.72	**100**	79.31	20.14	0.55	**100**	61.19	22.32	16.49	**100**	68.43	30.26	1.31	**100**
% in common with clouds or rain	**32.30**	**32.20**			**38.51**	**11.51**			**0.69**	**2.65**			**21.23**	**42.49**			**1.88**	**4.14**		
Genus*	155	69	208	**432**	6	105	29	**140**	165	13	46	**224**	111	33	105	**249**	122	27	94	**243**
Proportion of total OTUs (%)	35.88	15.97	48.15	**100**	4.29	75.00	20.71	**100**	73.66	5.80	20.54	**100**	44.58	13.25	42.17	**100**	50.21	11.11	38.68	**100**
% in common with clouds or rain	**57.30**	**75.09**			**82.86**	**21.64**			**21.80**	**77.97**			**48.61**	**76.09**			**43.52**	**77.69**		
Family*	47	10	119	**176**	4	55	25	**84**	78	0	41	**119**	54	4	67	**125**	56	6	65	**127**
Proportion of total OTUs (%)	26.70	5.68	67.61	**100**	4.76	65.48	29.76	**100**	65.55	0	34.45	**100**	43.20	3.20	53.60	**100**	44.09	4.72	51.18	**100**
% in common with clouds or rain	**71.69**	**92.25**			**86.21**	**31.25**			**34.45**	**100**			**55.37**	**94.37**			**53.72**	**91.55**		
Order*	22	7	59	**88**	2	27	17	**46**	33	0	28	**61**	22	3	41	**66**	25	4	38	**67**
Proportion of total OTUs (%)	25.00	7.95	67.05	**100**	4.35	58.70	36.96	**100**	54.10	0	45.90	**100**	33.33	4.55	62.12	**100**	37.31	5.97	56.72	**100**
% in common with clouds or rain	**72.84**	**89.39**			**89.47**	**38.64**			**45.90**	**100**			**65.08**	**93.18**			**60.32**	**90.48**		
Class*	12	4	27	**43**	0	13	5	**18**	14	1	12	**27**	8	3	18	**29**	7	4	19	**30**
Proportion of total OTUs (%)	27.91	9.30	62.79	**100**	0	72.22	27.78	**100**	51.85	3.70	44.44	**100**	27.59	10.34	62.07	**100**	23.33	13.33	63.33	**100**
% in common with clouds or rain	**69.23**	**87.10**			**100**	**27.78**			**46.15**	**92.31**			**69.23**	**85.71**			**73.08**	**82.61**		
Phylum*	7	1	15	**23**	0	4	4	**8**	5	1	8	**14**	3	0	11	**14**	3	1	11	**15**
Proportion of total OTUs (%)	30.43	4.35	65.22	**100**	0	50.00	50.00	**100**	35.71	7.14	57.14	**100**	21.43	0	78.57	**100**	20	6.67	73.33	**100**
% in common with clouds or rain	**68.18**	**93.75**			**100**	**50.00**			**61.54**	**88.89**			**78.57**	**100**			**78.57**	**91.67**		

* Number of OTUs

Supplementary Table 8 (electronic .xlsx file): Kruskal-Wallis test results of the comparison of read abundances in clouds *vs* rain samples at the genus level.

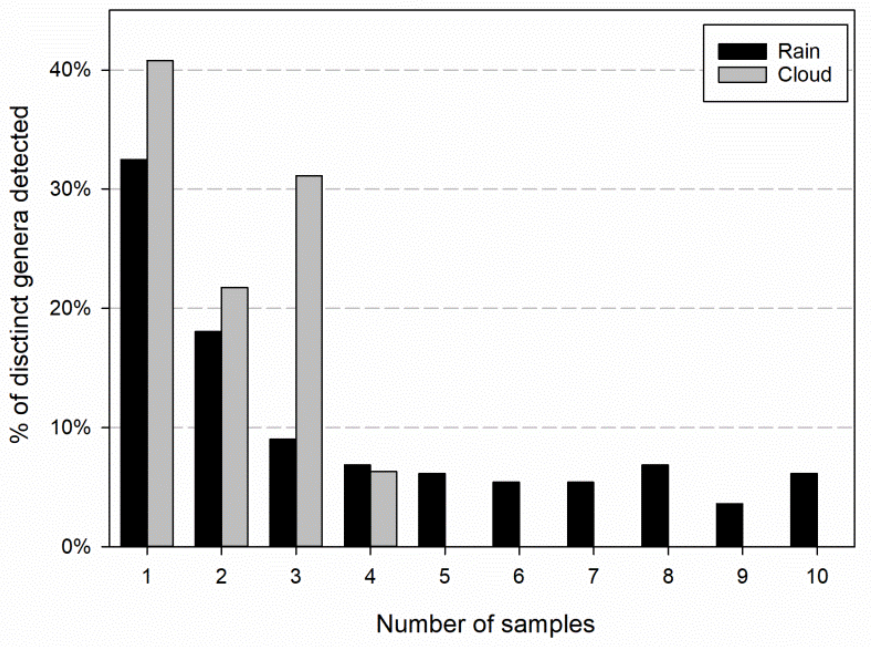
Supplementary Figure 1: Geographical configuration of the sampling sites. **A-** map of France; **B-** altitude profile between puy de Dôme Mountain' summit and Opme station (This figure was created using © Google Earth); **C-** Cloud sampling operations at PUY station, with high-flow-rate impingers and cloud droplet impactor pointed by the blue arrow; **D-** Rain collection at Opme station, with puy de Dôme Mountain visible at the background.



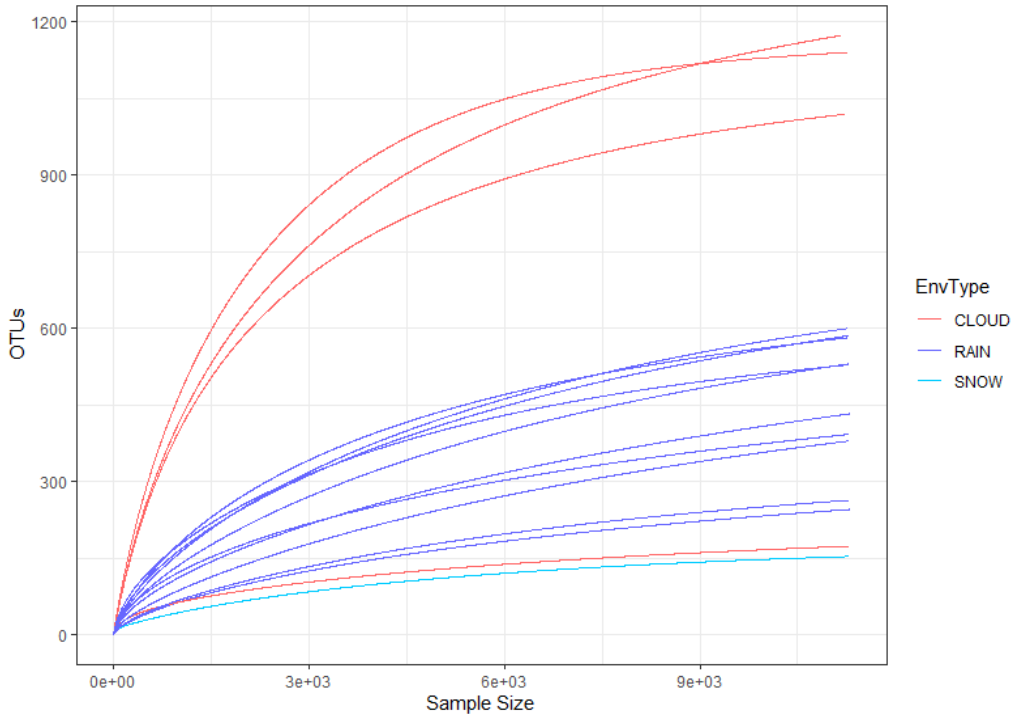
Supplementary Figure 2 (electronic .pdf file): Daily meteorological data for the sampling dates at puy de Dôme station (PUY, clouds and fresh snow sampling) and Opme station (OPME, rain sampling) organized by chronological order: **(a)** altitude profiles of cloud ice and liquid water contents (ERA5 ECMWF model); **(b)** Boundary layer height (ERA5 ECMWF model from Cézeaux station, 410 a.s.l.); **(c)** Main meteorological data at the sampling site (5 min intervals; T, RH and wind speed and direction at PUY for clouds and snow sampling, T and RH at OPME for rain sampling); **(d)** total precipitation from disdrometer measurements at OPME station (30 s intervals). Cloud sampling periods are framed in red.

Supplementary Figure 3 (electronic .pdf file): Seventy-two-hours backward trajectory plots and associated sources areas, extracted from ERA5 data reanalysis.

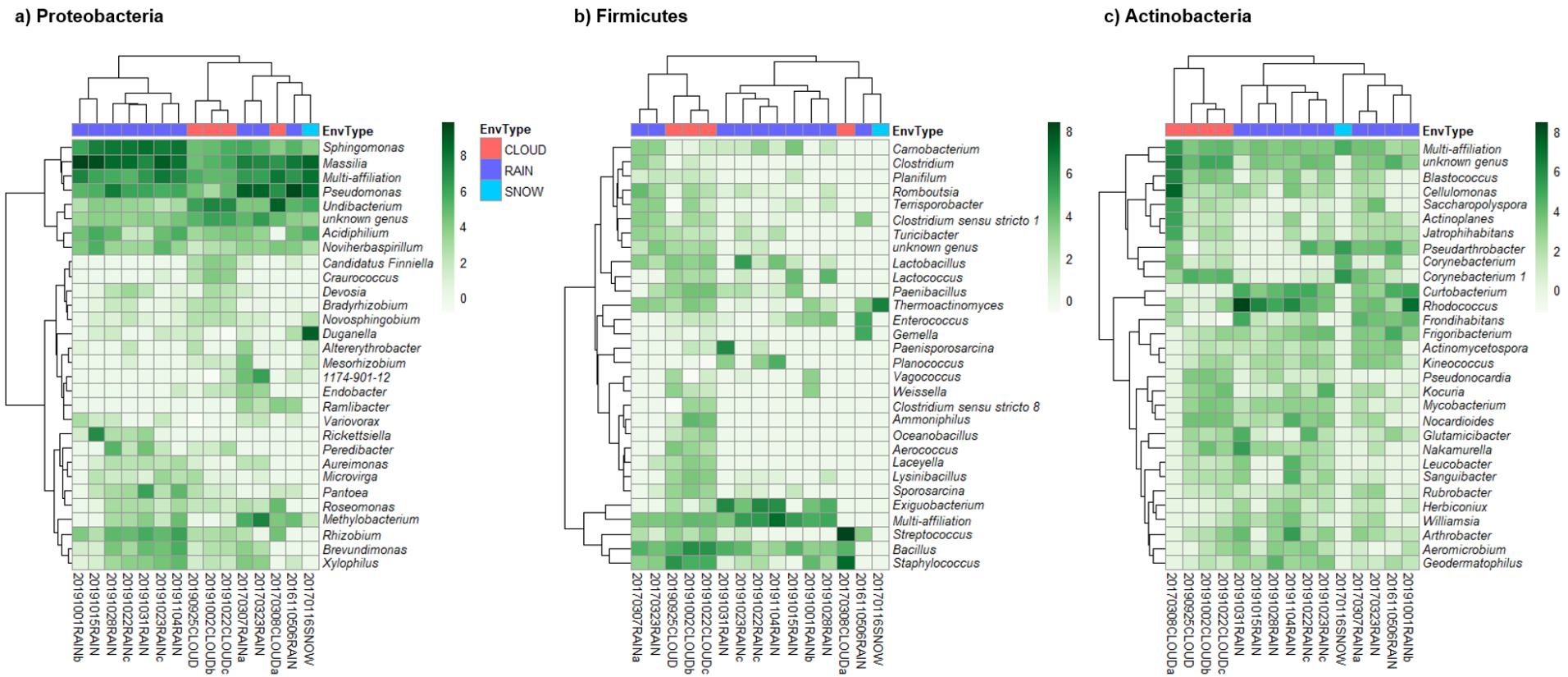
Supplementary Figure 4: Frequency distribution of distinct bacteria genera in clouds (363 genera) and rain samples (277 genera).



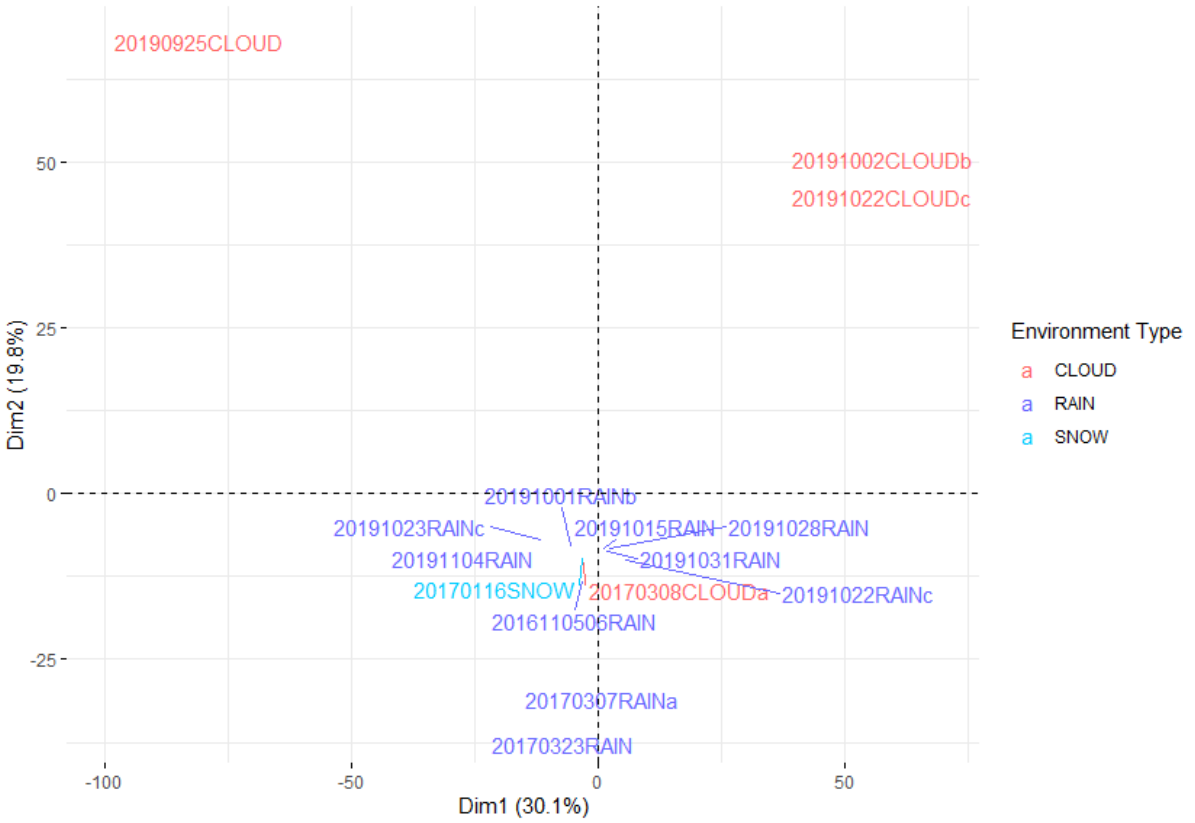
Supplementary Figure 5: Rarefaction curves.



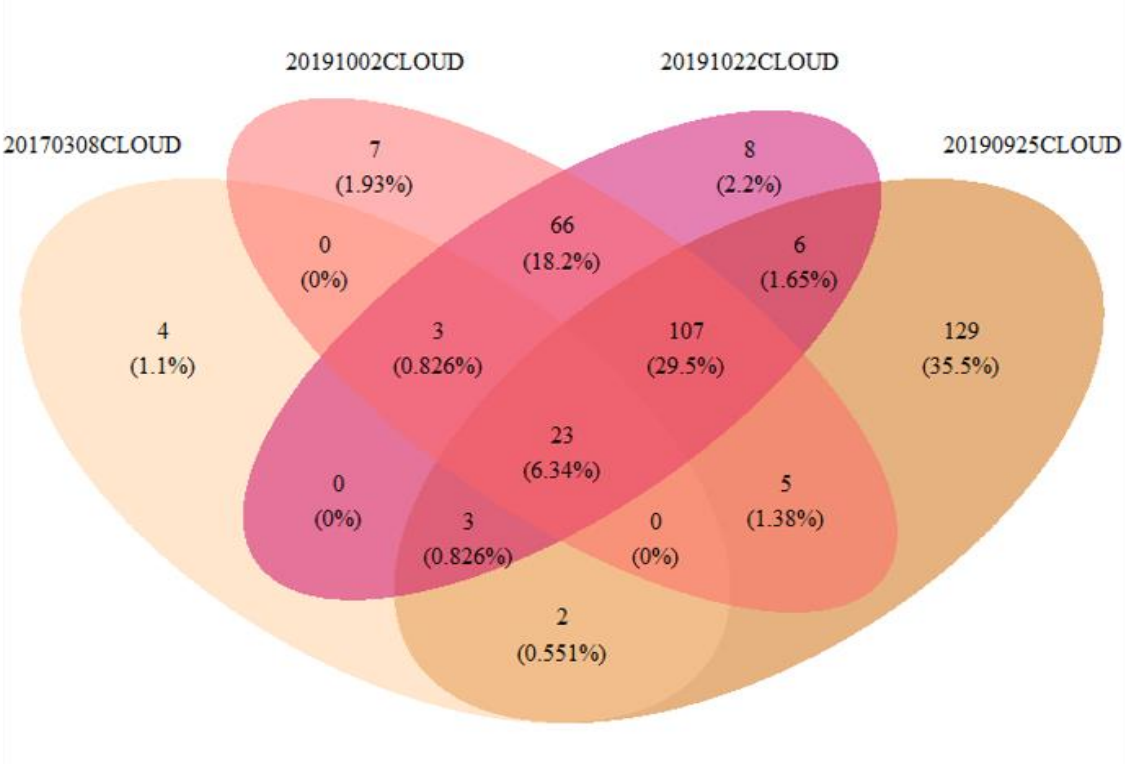
Supplementary Figure 6: Distribution of the 30 most represented bacteria genera in **a)** Proteobacteria, **b)** Firmicutes and **c)** Actinobacteria among the samples, and corresponding hierarchical clustering (Ward's method). Intensity scale depicts centered-log ratio abundances; *a*, *b* and *c* letters after sample names indicate chronological associations between cloud and rain samples.



Supplementary Figure 7: Principal component analysis plot based on OTU composition in the samples (4,510 OTUS).



Supplementary Figure 8: Venn diagrams depicting the distribution of distinct bacteria genera among cloud samples.



204 # References:

205 Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, et al.
206 Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature*
207 2012; **488**: 91–95.

208

3. Conclusion

This work highlighted the high biological similarity between clouds and rain with ~75% bacterial genera in common, but also the presence of significantly more abundant genera in clouds (e.g., Undibacterium, Staphylococcus) or precipitation (e.g., Massilia, Rhodococcus). The publication also concludes that clouds contain bacterial richness, while precipitation scatter this richness and scavenged the air column, picking up biomass in the process. To investigate this last point further, cloud:rain ratios were calculated for the three events a, b and c (where clouds and associated rain were collected on the same day) regarding the biomass (number of cells by mL) and the richness (number of OTUs). These ratios were related to the dilution factor (estimated relative to [Na+]) between clouds and rain and demonstrate highly linear relationships between dilution factor and biomass and richness (Figure 1). First, biomass increases proportionally in rain with increasing dilution factor (dilution of cloud influence in rain with scavenging), supporting scavenging as a source of biomass. Second, the dilution of richness raises with increasing dilution factor, i.e., the richness in rain decreases with increasing scavenging, making clouds the source of the richness.

However, these results were not presented in the article due to the few events represented (only three points on the graphs, Figure 1), which do not allow for true correlations. It would be very interesting in the future to collect more associated cloudy and rainy events to complete this dataset and confirm or establish correlations between these phenomena. It would also be relevant to link these cloud:rain ratios of biomass and richness with precipitation density and duration. For now, no direct relationship could be established with these three events alone. Finally, the bacterial communities of clouds and precipitation have been studied here and it would now be interesting to look at aerosols to get a more complete picture of what is happening during the scavenging and the sources contributed.

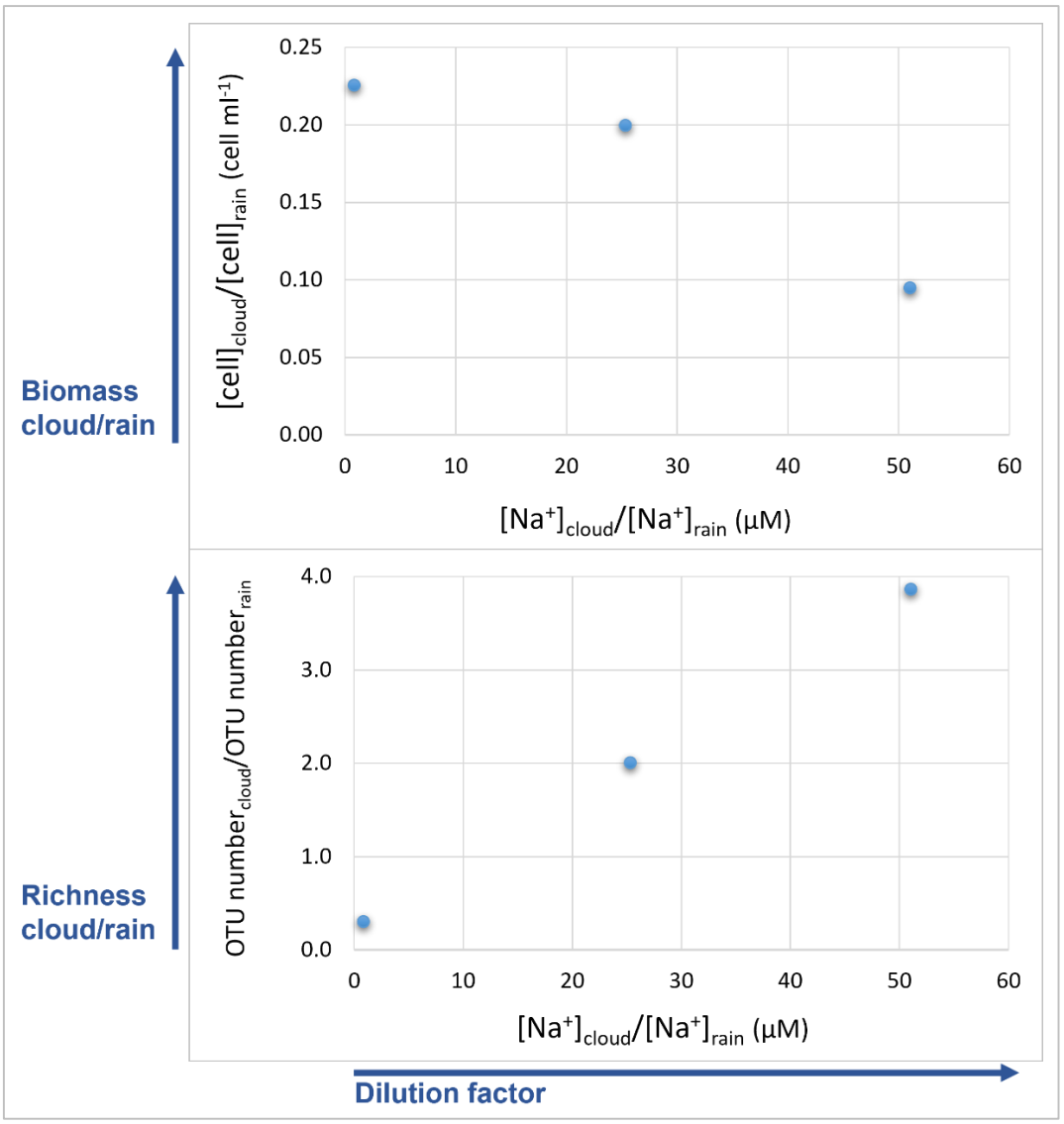


Figure 1: Graphical representations of the cloud:rain ratios for cell concentration and number of OTUs as a function of the dilution factor between the rain and the cloud source, for the events *a*, *b* and *c.* These events are composed of rain and the associated cloud collected on the same day. The dilution factor is estimated based on the concentrations of Na+ in clouds relative to the rain.

Chapter 4: Clouds as atmospheric oases

The introduction section below is a summary and introduces the general context for the following study section 3.

1. Introduction

It has long been established that viable microorganisms are present in the atmosphere up to high altitude in clouds (Fuzzi et al., 1996; Lighthart and Shaffer, 1995; Sattler et al., 2001). Their activity has, more recently, been studied through rRNA and rRNA gene sequencing (Amato et al., 2017; Klein et al., 2016; Šantl-Temkiv et al., 2018; Womack et al., 2010) revealing which part of the communities were potentially active. Nevertheless, the functional profile of airborne microbial communities is still completely unknown in outdoor aerosols and poorly understood in clouds, with only one recent study investigating biological functioning in three clouds (Amato et al., 2019).

In this context, the main objective of this thesis was to investigate whether clouds could be a specific habitat for airborne microorganisms, and thus harbor a peculiar microbial functional profile. This study continues what was initiated in Amato et al. (2019) on clouds, and was expanded to include investigations on aerosols (i.e., “dry atmosphere”) as a point of comparison. Clouds were therefore compared to aerosols in terms of microbial biodiversity and functional profile with untargeted high-throughput sequencing techniques: metagenomics (MG) coupled with metatranscriptomics (MT). The quantitative aspects have, furthermore, been strengthened compared to this previous work, by avoiding the amplification step such as MDA and thus allowing a more accurate relative quantification of gene expression and functions.

This part of the study was highly challenging (as emphasized in **Chapter 2**, **Article 1**):

- The goal of direct sequencing from such low-biomass environment implied sampling large volumes, over short periods of time to avoid “smoothing out” possible environmental variations.
- The high turnover of RNA required sampling with the use of a fixative agent.
- The lack of reference atmospheric MG, along with the known elevated variability of atmospheric microbiota, created the need to generate both MG and MT from each sample.
- The processing of MT data is not yet well standardized in the literature and requires development and adaptation, especially for such environmental data.
- Interpretation of these large datasets generated by modern sequencing techniques (e.g., Illumina) is challenging due to the enormous amounts of information provided.

Based on these last challenges, a workflow adapted to our specific needs had to be developed, based on existing tools. A collaboration was initiated with the Galaxy team of the Freiburg University in Germany (Dr Bérénice Batut and PhD student Engy Nasr) in order to improve the current workflow and make it publicly available on the Galaxy Europe platform. Similarly, a collaboration has been started with the MEDIS team in Clermont-Ferrand (Dr Pierre Peyret and Dr Sophie Marre) to improve the accuracy of the taxonomic affiliation from metagenomes and metatranscriptomes by reconstructing rRNA sequences (see **Chapter 2** - **section 4** for collaborations).

The computing center "Mésocentre Clermont-Auvergne" was used, as well as the associated Galaxy AuBi (Auvergne BioInformatique) platform, to store and process the large datasets generated by metagenomes and metatranscriptomes sequencing (~60-200 M reads per sample).

I presented preliminary results (first MG et MT samples from a cloud) at the World Microbe Forum in 2021 (I-Poster), and the bioinformatics workflow built at the AuBI bioinformatics platform day for NGS (oral presentation, 2021) and at the JOBIM bioinformatics congress (online poster, 2021) (see **Annexes 2**).

The following section elaborates on the main bioinformatics steps of the metatranscriptomes processing and the procedures chosen for the analysis of our data. Sampling and sample processing were also optimized, and positive and negative controls were performed to support this part of the work. These methodological improvements are presented in **Chapter 2**. The results section will describe the preliminary results of this study (**section 3**). Here, 6 aerosol samples and 8 cloud samples (out of 6 and 9) were common with the amplicon study **Chapter 2**.

2. Bioinformatics workflow for processing environmental metagenomes (MGs) and metatranscriptomes (MTs)

The analysis of MT data consists of several main steps similar to those of MG: preprocessing (quality filtering, etc.), assembly, taxonomic and functional annotations with adapted databases. In addition, differential expression analyses must be performed by comparing MTs and MGs, when MG data are available. Many bioinformatics tools are available for each step and their choice is guided by the type of dataset, and the purpose of the experiment (Shakya et al., 2019).

Public bioinformatics workflows are already available for processing metatranscriptomic data. However, these do not address our two main needs regarding current knowledge in atmospheric microbiology:

- A database adapted to the biodiversity of the environmental microbiota and containing at least bacteria and fungi.
- The possibility of parallel processing of metagenomes to standardize the number of metatranscriptomic read.

We developed our own bioinformatics workflow on the Galaxy AuBI platform with existing tools (**Figure 1 of the article section 3**), inspired by in the work of Salazar *et al.* (2019) on the marine environment.

The following sections will be separated in two parts: a state of the art on current uses (called “In the Literature”) and then specific choices in our bioinformatics workflow (BiW) to process metatranscriptomic and metagenomic data (called “In our BiW”).

2.1. Preprocessing

In the literature

The first step of preprocessing is common to all NGS datasets. It consists of filtering out poor quality sequences (quality control, QC). Many QC tools can be used for Illumina sequencer-derived short reads, such as: FastQC (Andrews, 2010), FaQCs (Lo and Chain, 2014), fastp (Chen et al., 2018) and Trimmomatic (Bolger et al., 2014).

Another important step is the removal or depletion of ribosomal RNA (rRNA) sequences that are highly abundant in RNA-Seq datasets, up to 90% of the transcripts. Although informative as proxies for metabolic activity, rRNAs are useless for downstream functional analyses to qualify activity, such as characterization of active metabolic pathways. They are therefore often physically removed prior to sequencing, but this adds an experimental step that can alter samples, and rRNAs depletion is often not complete. Alternatively, rRNA sequences can also be discriminated in RNA-Seq datasets through

bioinformatics, using tools and ribosomal databases such as SortMeRNA (Kopylova et al., 2012) with Silva databases (Quast et al., 2013).

Additionally, unwanted organisms can be selectively filtered and removed from the datasets, such as human reads for example. These reads are typically detected and removed using common mapping tools.

In our BiW

In the case of our bioinformatics workflow (BiW), the FastQC and Trimmomatic tools were used for QC and trimming of raw reads. The Bowtie2 (Langmead and Salzberg, 2012) tool was used to filter human reads from the dataset by mapping them to the human genome. Bowtie2 is a common tool well-suited for quickly aligning single or paired short reads to long reference sequences and is available on Galaxy. Finally, the SortMeRNA tool was used to filter rRNA reads and retrieve them in separate files.

2.2.Taxonomic affiliation

In the literature

The taxonomic affiliation of transcripts is similar to that of metagenomic data and uses the same tools. The taxonomy to which transcripts are affiliated indicates which taxonomic groups have transcriptional activity in a sample and are therefore potentially active. Taxonomic assignment is often performed on ribosomes alone, for example from rRNA amplicon data. Another approach is to take advantage of the availability of entire metagenomes and metatranscriptomes to use all potentially accessible information, not just rRNA sequences.

Common read-based taxonomic classification tools include Kraken2 (Wood and Salzberg, 2014), MetaPhlan2 (Truong et al., 2015) or GOTTCHA (Freitas et al., 2015). For contigs or full-length transcripts, tools such as Centrifuge have been designed (Kim et al., 2016). The main limitation of taxonomic classification tools is the available reference databases. Most bioinformatics tools only use subsets of the available genomes and focus on certain microorganisms or microbiota. This makes it even more difficult to process environmental samples with bioinformatics workflows that are unusable in our case due to the databases used for taxonomic affiliation.

In our BiW

The taxonomic affiliation was performed with Kraken2 using the whole metagenomes and metatranscriptomes as inputs and the “PlusPF” database (index size of 53.2 GB; 31,580 taxonomy nodes) which contains: archaea, bacteria, viral plasmid, humans, protozoa and fungi. Kraken2 is an ultrafast tool that uses k-mers to assign taxonomy from metagenomic sequences. It can be used for short reads or contigs and has the advantage of being available on Galaxy. It also allows to select an existing database, or to create your own.

2.3.Assembly

In the literature

In many cases, no reference genomes are available for environmental microbiomes. Therefore, preprocessed reads must be assembled using *de novo* assemblers. Assembled contigs provide longer genome or expressed genome fragments to perform more accurate functional annotation (compared to short reads) and a set of reference genes for reads mapping to obtain a count table and perform expression analysis. In the case of metatrancriptomics, in the absence of reference MG, transcript reads are mapped to the assembled transcripts to obtain the expression count table. In the best case, the associated MG is used as a reference for transcript mapping, allowing the derivation of relative expression levels.

Several metagenomic assemblers exist to handle complex metagenomes such as MEGAHIT (Li et al., 2015), IDBA-UD (Peng et al., 2012) or metaSPAdes (Nurk et al., 2017). All of these tools can handle sequence similarities between genomes in highly conserved regions, but their efficiency in managing transcript reconstruction is not well characterized. Transcripts have certain peculiarities such as introns/exons, different isoforms, and shorter non-coding RNAs that complicate their assembly. Specific assemblers such as Trinity (Grabherr et al., 2011), rnaSPAdes (Bushmanova et al., 2019) or IDBA-tran (Peng et al., 2013) among others, are designed to manage transcriptome sequences in single organisms. Finally, the *de novo* assemblers IDBA-MT (Leung et al., 2013), IDBA-MTP (Leung et al., 2014) and Transcript Assembly graph (TAG) (Ye and Tang, 2016) are designed for metatranscriptomic data from communities, and they take into account both unique specificities of transcripts and the complexity of microbial assemblies. *De novo* assembly of metatranscriptomic reads is still being improved and only few assemblers have been developed for these particular datasets (Shakya et al., 2019).

In our BiW: construction of a non-redundant catalog of genes

In order to have a set of reference genes that could be used throughout our study, and given that no reference metagenome is available for atmospheric samples, we constructed an exhaustive non-redundant catalog of genes including all generated cloud and aerosol metagenomes.

Each metagenome was individually *de novo* assembled using MEGAHIT from non-rRNA reads filtered with SortMeRNA. MEGAHIT has the advantages of being fast, available on Galaxy, and has been ranked as one of the most efficient *de novo* assemblers (Forouzan et al., 2018). Next, all assembled MGs were merged to predict the genes likely present. Gene prediction was performed using MetaGeneAnnotator (Noguchi et al., 2008), which is specifically fitted for prokaryotic gene detection. This tool can detect eukaryotic genes but this requires improvements, including coupling with another tool such as MetaEuk (Karin et al., 2020) to obtain a more complete detection of eukaryotic

(particularly fungal) genes. The nucleotide sequences of the genes in our new catalog were clustered at 95% similarity and only representative sequences from each cluster were retained using CD-Hit (Fu et al., 2012; Li and Godzik, 2006) to remove redundancies.

2.4.Functional Annotation

In the literature

Transcripts are products of gene expression and approximations of the actual metabolic functioning of a community (in the case of metatranscriptomics).

The first step of functional annotation is to predict gene positions with tools such as FragGeneScan (Rho et al., 2010), Prodigal (Hyatt et al., 2010) or MetaGeneAnnotator (Noguchi et al., 2008). Gene prediction is not an easy step, as many tools are not suitable for complex multi-organism's community datasets that require annotated reference genomes. Moreover, the above mentioned tools, although suitable for metagenomics, focus on prokaryotic genes as these are not yet able to handle eukaryotic genetic structures including exons and introns. MetaEuk (Karin et al., 2020) is one of the few tools developed to detect eukaryotic genes in metagenomic datasets.

After gene prediction, functional assignment is performed based on similarity searches using tools such as DIAMOND (Buchfink et al., 2014). The most complete reference databases existing so far can be used such as KEGG (Kanehisa and Goto, 2000), NCBI RefSeq (O'Leary et al., 2016) or UniProt (The UniProt Consortium, 2019).

Once functional annotations are performed, multiple gene identifiers can be used to gather information such as gene ontology (GO) (Ashburner et al., 2000; Carbon et al., 2021), KEGG orthology (KO), orthologous group (OG) from EggNOG database (Huerta-Cepas et al., 2019), etc. It is possible in most cases to convert gene identifiers from one type to another, but these multiple and complex trees of identifiers make it difficult to standardize of functional information across studies.

In our BiW

Functional annotation of the genes present in the catalog was performed using DIAMOND (Buchfink et al., 2014) and the UniProtKB (The UniProt Consortium, 2019) functional database. UniProt IDs are linked to information including full gene name, phylogeny, gene ontology (GO), gene length, etc.

DIAMOND software was selected here for its computational efficiency and accuracy in aligning of large datasets: it is 2,500 to 20,000 times faster than BLASTx for Illumina sequences (100-150bp) depending on mode of use ("sensitive" or "fast").

2.5. Data normalization and differential expression analyses

In the literature

First, because sequencing depth can vary considerably between samples, regardless of actual functional differences, normalization of RNA-Seq counts is a crucial step before performing differential expression analysis (DEA) for reliable detection of true transcriptional differences between samples (Klingenberg and Meinicke, 2017). Several normalization methods can be applied each with their weaknesses and approximations and can also be coupled with different data transformations (log, centered-log ratio, etc.) (Gloor et al., 2017; Klingenberg and Meinicke, 2017; Salazar et al., 2019). Second, it is important to consider that transcript copy number depends not only on the level of gene expression but also on the number of gene copies in the genome. Therefore, it is also necessary to normalize MT counts to the associated counts in MG (Klingenberg and Meinicke, 2017), although this was not always done in metatranscriptomics studies because MG data were not available.

DEA is now an essential step in the analysis of MT data that allows for comparison of different environmental conditions and parameters and their effect on community function or to observe community dynamics over time (Shakya et al., 2019). Some of the most widely used DEA tools include the R packages EdgeR (Robinson et al., 2010), DeSeq2 (Love et al., 2014) and limma (Ritchie et al., 2015). These take into account quantitative (abundance) information to identify significantly differentially expressed genes among multiple samples. These tools include options for normalization (for sequencing depth) and/or transformation prior to DEA. However, because most of these were designed primarily for single genome analyses, they generally do not offer to include reference MG count data in MT normalization. A brand new tool, the R package MTXmodel (Zhang et al., 2021b), was developed to address this particular need in comparative MG/MT analyses and proposes to consider DNA abundance for MT DEA.

In our BiW

For each sample, MG and MT non-rRNA reads were mapped against our gene catalog to obtain the number of specific gene reads using BWA-MEM (Li and Durbin, 2009, 2010). The BWA-MEM mapping tool was preferred over Bowtie2 because it is faster at this step, and is designed to map medium and long reads (>100 bp) against a large reference genome.

Count data were normalized by centered-log ratio (clr) transformation and DEA was performed with the MTXmodel R package.

3. Results, Article 4: “Clouds as atmospheric habitats for microorganisms”, written as a preliminary article

Clouds as atmospheric habitats for microorganisms

Raphaëlle Péguilhan[1], Florent Rossi[1], Engy Nasr[2], Bérénice Batut[2], François Enault[3], and Pierre Amato[1]

[1] Université Clermont Auvergne, CNRS, SIGMA Clermont, ICCF, F-63000 CLERMONT-FERRAND, France.

[2] University of Freiburg, Department of Computer Science, Bioinformatics group, 79110 FREIBURG, Germany.

[3] Université Clermont Auvergne, CNRS, Laboratoire Microorganismes : Génome et Environnement, F-63000 CLERMONT-FERRAND, France.

Keywords: Clouds, aerosols, metagenomics, metatranscriptomics, microbial activity, airborne microorganisms

Abstract

Microorganisms were reported to be present and active in the atmosphere up to the clouds. However, little is known about their function and potential differences between atmospheric compartments. Clouds may provide an aqueous microenvironment for cells in contrast to the drastic desiccation conditions encountered in dry air (i.e., the aerosols). Our question was therefore whether clouds can be specific atmospheric habitats (i.e., oases) for airborne microorganisms. For this purpose, clouds and aerosols were collected at puy de Dôme mountain (1,465 m a.s.l.) and metagenomics coupled to metatranscriptomics was used to compare the two atmospheric situations in terms of biodiversity and, in particular, functional profile. Microbial diversity was not distinct between clouds and aerosols. Central metabolism was overexpressed in both aerosols and clouds, pointing out aerosols as a player as well in atmospheric microbial activity. Nevertheless, a significantly higher functional potential was found in clouds, with strong overexpression of energy metabolism and specific responses to the cloud environment. This constitutes a major advance in the understanding of microbial functioning and cycling in the atmosphere with a potential increase in activity in clouds before falling back to surface ecosystems with precipitations.

Introduction

The outdoor atmosphere harbours a wide variety of airborne microbial communities. Commonly observed bacterial phyla are the Proteobacteria, Actinobacteria, Firmicutes and Bacteroidetes, and fungal phyla are the Basidiomycota and Ascomycota (Amato et al., 2017; Bowers et al., 2013; Fröhlich-Nowoisky et al., 2012). The concentration of bioaerosols in the atmosphere is generally between 102 to 106 microbial cells by m3 of air, depending on the proximity to the emission source (Bauer et al., 2002; Burrows et al., 2009b).

Potentially active taxa have been reported in the atmosphere. Bacteria were demonstrated to remain active when aerosolized based on laboratory experiment (Krumins et al., 2014). Bacterial activity in aerosols has also been investigated on environmental samples at high altitudes (USA) (Klein et al., 2016) and in the Artic (Šantl-Temkiv et al., 2018) using rRNA and rRNA gene sequencing. In Klein et al. (2016), the most potentially active bacterial orders were Rhodospirillales (Proteobacteria), Actinomycetales (Actinobacteria), Saprospirales and Cytophagales (Bacteroidetes). In Šantl-Temkiv et al. (2018) the bacterial phyla Cyanobacteria, the order Clostridiales (Firmicutes) and the family Rubrobacteridae (Actinobacteria) were the most potentially active, while Proteobacteria had a low activity potential in Artic aerosols. Concerning fungi, Womack et al. (2015) studied the active taxa in aerosols over the Amazon rainforest and reported that the Ascomycota were the dominant potentially active phyla (Sordariomycetes and Lecanoromycetes). The Basidiomycota were also active along with the class Agaricomycetes.

Microbial activity has also been reported in clouds (Amato et al., 2017, 2019; Sattler et al., 2001), with bacteria being the main potentially active kingdom. Bacteroidetes, Alpha and Beta-Proteobacteria were the most potentially active bacterial phyla, and it was Ascomycota for fungi (Amato et al., 2017, 2019). Moreover, the functional profile was characterized on three clouds with pre-amplified (MDA) metagenomes and metatranscriptomes (Amato et al., 2019).

Biological functions indicative of a harsh environment were overexpressed (e.g., maintenance of homeostasis, responses to oxidative stress, radicals and oxidants) as well as central metabolic pathways (e.g., tricarboxylic acids cycle, pentose phosphate cycle and glucose metabolic processes). These metabolic functions expressed in clouds confirm that microorganisms are potential players in cloud chemistry, including through processes such as hydrogen peroxide catabolism. H2O2 has been demonstrated to modulate the energy metabolism of cloud microbiota and consequently affect the biotransformation rates of carbon compounds, which impacts cloud chemistry (Wirgot et al., 2017). Microorganisms (mainly bacteria) are now considered in atmospheric models (Ervens and Amato, 2020; Fankhauser et al., 2019; Khaled et al., 2021; Zhang et al., 2021a). In addition, microorganisms in clouds fall back to the Earth's surface with precipitation and potentially impact local ecosystems (e.g., lake, vegetation) if they are still active or viable when reaching the surface (Hervàs et al., 2009; Morris et al., 2008; Noirmain et al., 2022). It is therefore important to investigate the functioning of atmospheric microbial communities to better predict the potential impact of microorganisms on cloud chemistry and on local ecosystems.

Clouds may provide an aqueous microenvironment for cells, protection from direct sunlight, and perhaps better access to nutrients, in contrast to conditions encountered in air with no condensed water (i.e. aerosols). The hypothesis here is that clouds may therefore be specific habitats, like atmospheric oases, for airborne microbial communities. The delineation between clouds and dry aerosols is unclear as relative humidity is more of a continuum for airborne particles (Ge et al., 1998), however clouds are defined by the presence of condensed water, incorporating aerosol particles as cloud condensation nuclei (CCN), when the atmosphere is supersaturated with water (Koehler theory). Thus, biodiversity is not expected to be significantly different between clouds and aerosols, as most airborne microorganisms are considered efficient CCNs due to their size (~1 µm) and shape (Bauer et al., 2003; Lazaridis,

2019). However, microbial communities are expected to be more active in clouds due to the presence of condensed water and more tolerable atmospheric conditions.

To this end, metagenomics (MG) and metatranscriptomics (MT) have been used to compare the functioning of microbial communities in clouds and aerosols. Both of these non-targeted exploratory approaches are increasingly used in environmental contexts, although they remain challenging to implement with field sampling (Carvalhais et al., 2012; Pascault et al., 2015; Salazar et al., 2019; Shakya et al., 2019). Samples were collected at high altitude (puy de Dôme summit, France; 1,465 m a.s.l.) and methodological improvements were made to achieve metagenomes and metatranscriptomes sequencing from the low biomass encountered in the atmosphere. A new bioinformatics workflow for processing MTs and MGs was built to meet the specific needs of the dataset. Additional analyses were performed, such as ATP quantification, total cell counts, and ion quantification, as well as retrieval of meteorological data and backward trajectory plots of the air masses.

This is the first study, to our knowledge, to investigate and compare potential microbial activity in clouds and aerosols, and the first (non-amplified) MG and MG recovered from cloud and aerosol samples in the literature.

Material and methods

Sample collection

Samples were collected at puy de Dôme Mountain's summit (PUY; 1,465 m a.s.l., 45.772° N, 2.9655° E), France, located ~400 km East from the Atlantic Ocean and ~300 km North from the Mediterranean Sea. This rural area is mainly composed of deciduous forests, pastures landscapes, and the city of Clermont-Ferrand ~8 km from the puy de Dôme Mountain. The PUY station is mainly exposed to winds coming from the North and West (Deguillaume et al., 2014; Renard et al., 2020). This station is part of the Cézeaux-Aulnat-Opme-Puy de Dôme (CO-PDD) instrumented platform for atmospheric research (Baray et al., 2020) and is fully equipped for the processing and biological analysis of samples on site.

Cloud water was collected mainly during falls 2019 and 2020 over periods of ~2.5 to ~6.5 consecutive hours (**Table 1**). Aerosols were collected during summer and fall 2020 over periods of ~5.8 to 6.5 consecutive hours. A total of 9 cloud water and 6 aerosol samples were collected. For both clouds and aerosols sampling, several high-flow-rate impingers (HFRi) (DS6, Kärcher SAS, Bonneuil sur Marne, France) (air-flow rate of 2 m3/min) (Šantl-Temkiv et al., 2017) were used. In the case of aerosol sampling, HFRi were filled with 1,7 L of 0.5 X GF/F-filtered autoclaved nucleic acid preservation (NAP) buffer solution (Camacho-Sanchez et al., 2013; Menke et al., 2017) as collection liquid, specifically for nucleic acid analyses. Another HFRi was filled with 1,7 L of autoclaved ultrapure water for other analyses requiring no saline contamination (biological and chemical analyses). The evaporation of the collection liquid during sampling was compensated with autoclaved ultrapure water every hour by weight. For cloud sampling, HFRi were filled with 850 mL of 1 X NAP buffer solution (as the collection liquid will gain in volume and become diluted with the recovery of cloud water) and used only for nucleic acid analyses. Other biological and chemical analyses were carried out from a cloud droplets impactor sterilized by autoclave and sampling in parallel as in Amato et al. (2019).

Sampling blanks with NAP buffer were performed at each sampling occasion. A volume of NAP buffer was left in the impactor tank for 10 min, or during the sampling time with the sampler off and the inlet closed. For other analyses, the water blank was sampled while it was in the impinger tank just before sampling. Cloud and aerosol collected samples and blanks were processed immediately after sampling using the PUY station's microbiology facility, within a laminar flow hood previously exposed to UV for 15 min.

Nucleic acid preservation (NAP) buffer

NAP buffer was prepared in high quantities according to the recipe in Camacho-Sanchez et al. (2013): 0.019 M of ethylenediaminetetra-acetic acid (EDTA) disodium salt dihydrate, 0.018 M of sodium citrate trisodium alt dihydrate, 3.8 M of ammonium sulfate, and H2SO4 (adjust pH at 5.2). NAP buffer was GF/F filtered to remove dirt particles and then autoclaved for sterilization.

Meteorological data and backward trajectory plots

Meteorological variables were measured at the PUY meteorological station: temperature, relative humidity, liquid water content (LWC), wind speed and direction (https://www.opgc.fr/data-center/public/data/copdd/pdd). The boundary layer height (BLH) was extracted from the ECMWF ERA5 global reanalysis (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5) (Hoffmann et al., 2019). The geographical origin of the air masses were recovered from 72-hours backward trajectory plots computed using the CAT trajectory model (Baray et al., 2020), which uses dynamical fields extracted from the ERA-5 meteorological data archive with a spatial resolution of 0.125° in latitude and longitude (for the present work). This tool allowed to estimate: (i) air mass backward trajectories starting from PUY summit; (ii) air masses history, as the density of

trajectory points below the BLH and the percentage of trajectory points above and below the BLH, over land and seas; (iii) the percentage of trajectory points near the CO-PDD observatory (distance < 50 km) and in each of 8 direction sectors (Renard et al., 2020).

Chemical analyses

The pH was measured directly after sampling. The main dissolved ions (Na+, NH4+, K+, Mg2+, Ca2+, Cl-, NO3- and SO42-) were quantified by ion chromatography from 5 mL filtered (0.2 µm) subsamples stored at -25°C until the analysis. Quantifications were performed using a Dionex DX320 (column AS11) for anions and a Dionex ICS1500 (column CS16) for cations, as in Deguillaume et al. (2014).

Total cell counts and ATP quantification

Cells were count by flow cytometry using a BD FacsCalibur instrument (Becton Dickinson, Franklin lakes, NJ) as in Amato et al. (2017). Adenosine-5'-triphosphate (ATP) was quantified in samples by bioluminescence (ATP Biomass Kit HS; BioThema; Handen, Sweden) using a GloMax 20/20 luminometer (Promega, Madison, WI, USA) with the same protocol as in Péguilhan et al. (2021).

Cell and ATP concentrations per m3 of air were extrapolated from the LWC for clouds and from the time of sampling and the sampler air-flow rate for aerosols.

Nucleic acid extraction and shotgun sequencing

DNA and RNA extraction was performed with the NucleoMag® DNA/RNA Water kit (Macherey-Nagel, Hoerdt, France) from same samples. All facilities were previously treated with RNase away spray solution (Thermo Scientific; Waltham, USA) to avoid possible deterioration of RNA. Immediately after sampling, the NAP buffer from each impingers was filtered on 0.22 µm porosity mixed cellulose esters (MCE) filters (47 mm diameter; ClearLine,

ref. 0421A00023), the bead-beating step of the kit protocol was performed in 5 mL Beads Tubes Type A (Macherey-Nagel, ref. 740799.50) with 1,200 µL of lysis buffer MWA1 (Macherey-Nagel). Filters and lysates were stored at -80°C in beads tubes until further processing. For DNA extraction, ~600 µL of lysate (split in two for RNA) were processing following the extraction kit protocol adapted for 47 mm filter membranes. Lysate for DNA extraction was RNA-treated by adding 1:50 volumes of RNase A (12 mg/mL, stock solution from Macherey-Nagel) after the lysis and centrifugal steps. DNA was eluted into 50 µL of RNase-free H2O with 5 min incubation at 56°C. DNA was quantified by fluorescence using the Quant-iT™ PicoGreen® dsDNA kit (Invitrogen; Thermo Fisher Scientific, Waltham, MA USA). For RNA extraction, the remaining ~600 µL of lysate were processed following the protocols for 47 mm filter membrane and for the isolation of RNA on 25 mm filter. RNA was DNA-treated adding 1:7 volumes of reconstructed rDNase (cf kit standard protocol). RNA was eluted into 30 µL of RNase-free H2O with 10 min incubation at room temperature.

DNA and RNA extracts were pooled for a same sample and sent (~30 µl) to GenoScreen (Lille, France) for further processing of RNAs (quantification, reverse-transcription) and shotgun sequencing of the metagenomes and metatranscriptomes by Illumina HiSeq 2*150 bp. A first sample (CLOUD20191022) was sent as a proof of concept and to adjust sequencing depth (~200 M reads per sample) and bioinformatics treatment. The other samples were sequenced with a lower sequencing depth (40 - 60 M reads per sample). Due to too low concentrations in DNA or RNA after libraries preparation, several samples were pre-concentrated from additional extract volumes (20 µl eluted in 8 µl).

DNA and RNA concentrations per m3 of air were extrapolated from the time of sampling and the sampler air-flow rate for both aerosols and clouds.

Bioinformatics workflow for environmental metagenome and metatranscriptome analysis

A bioinformatics workflow (BiW) was built for the processing of metatranscriptomes (MTs) and associated metagenomes (MGs) (**Figure 1**) to meet our specific needs: taxonomic database adapted to very diverse environmental samples and parallel processing of MGs to standardize MT count data. The BiW was constructed through the Galaxy instance deployed by the AuBI (Auvergne BioInformatique) network and the regional calculation cluster Mesocentre Clermont Auvergne (until obtaining count tables). The BiW contains the following steps: preprocess, taxonomic affiliation, construction of a catalog of genes, functional annotation and obtaining of count tables for the differential expression analysis (DEA).

Raw MG files (R1 and R2) contained approximately 30-260 M reads, while raw MT files are about 65-195 M reads (**Supplementary Table 1**; **Supplementary Table 2**), with an average read size of 150 bp. The preprocess step of our BiW used: FastQC (Galaxy v 0.72) (Andrews, 2010) for quality control (QC) analysis; Trimmomatic (Galaxy v 0.36.6) (Bolger et al., 2014) to filter and trim erroneous reads, Bowtie2 (Galaxy v 2.4.2) (Langmead and Salzberg, 2012) to map and filter human reads, and SortMeRNA (Galaxy v 2.1b.6) (Kopylova et al., 2012) to filter and recover in separated files the rRNA reads. Trimmomatic was used with an initial ILLUMINACLIP step for remaining Nextera (paired-reads) adapters, and with a sliding window of 10:30, a reads min length of 100 bp and the leading and trailing parameters with a quality threshold at 30. This trimming step removed about 26 to 37 % of the reads in MG files and about 20 to 49 % of the reads in MT files. Ribosomal reads were filtered with SortMeRNA. The default parameters were used with the “paired-out” option and all the available databases. The rejected files (non-rRNA reads) were kept for downstream functional analyses. Proportions of rRNA reads in MG datasets were typically between 1 and 2 % and between 80 and 94 % for MT datasets, except for the AIR20201124 aerosol sample which contained only 12 % of rRNA reads in MT and was removed from the downstream analysis (for MG and MT datasets). One

MT sample (CLOUD20201103) contained only 40 % of rRNA reads but, with further analysis, this was attributed to heavy contamination with human sequences (90% in non-rRNA reads). The filtering of human sequences was done on rejected files from SortMeRNA (non-rRNA reads) using Bowtie2 against the NCBI Homo sapiens genome "hg38_2021-5-18" with default parameters. Human sequences represented 0.01-0.71 % of the total trim MG sequences and 0.2-1.95 % (up to 90% for one sample) of the total trimmed MT sequences. This filtering step was done after SortMeRNA, but could be done directly after trimming to save computing time and a second human reads filtering for taxonomy analysis as described in **Figure 1**.

Taxonomic affiliation was done on the whole MG and MT datasets (including rRNA and non-rRNA reads) using Kraken2 (Galaxy v 2.1.1) (Wood and Salzberg, 2014) against the "PlusPF" kraken database (from 2021-1-27) including archaeal, bacterial, viral plasmid, human, protozoan, and fungal genomes. Here, this step was done using trimmed files and therefore before SortMeRNA and human sequence filtering but, in a better case to save computing time, the human reads filtering could be done before as represented **Figure 1**. Kraken2 was used with a confidence score threshold of 0.1 and with the "report" and "report-zero-counts" options. In total, 1.2-4 % of the MG reads and 63-90% of the MT reads were affiliated, with 2.3-28.8 % and 0.02-0.7 % of the reads affiliated to human genome respectively. The human reads were removed from the taxonomy dataset.

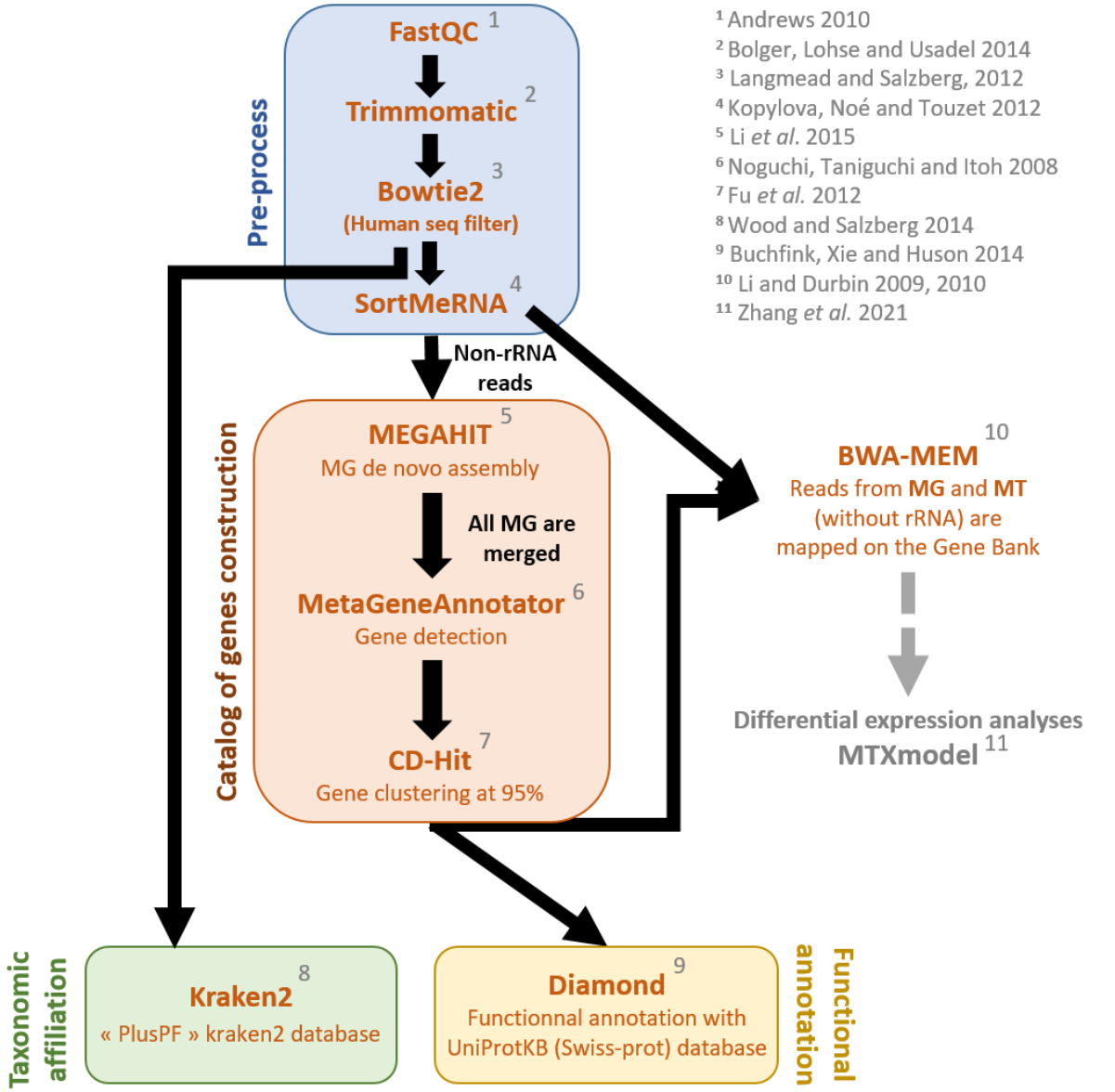


Figure 1: Main steps of the bioinformatics workflow for the analyses of metatranscriptomics and metagenomics data. MG: Metagenome; MT: Metatranscriptome.

A non-redundant gene catalog was constructed from all MG data agglomerated together, to be used as a reference MG common to all sample, as no reference metagenome is available for such environmental samples. The de novo assembler MEGAHIT (Galaxy v 1.1.3.5) (Li et al., 2015) was used to assemble the non-rRNA reads from each MG separately. Default parameters were used with a minimum length for contigs of 500 bp. Depending on the sample,

the number of assembled contigs ranged from ~43,000 to ~495,000 with a total of 2,832,534 contigs all MG together. The maximum size of the contigs obtained, depending on the sample, was between 21,000 and 200,000 bp with a mean size of the contigs between ~750 and 1,010 bp. All the MG assemblies were then merged into a single file for the gene prediction on contigs with MetaGeneAnnotator (Galaxy v 1.0.0). This tool was developed to detect prokaryotic genes but it can also predict certain eukaryotic genes. The "MetaGenomic" option was used and the output file in BED format was chosen. From the 2,832,534 contigs, a total of 3,168,750 genes were predicted. Finally, the nucleic sequences of the genes were clustered at 95% with CD-HIT (Galaxy v 4.8.1) (Fu et al., 2012; Li and Godzik, 2006) and only the representative sequence of each cluster was kept to avoid redundant genes. CD-HIT parameters were set to keep only sequences of at 100 bp minimum and matching at least to 90 % to the reference sequence. A total of 1,067,351 non-redundant genes were kept with an average length of 330 bp (from 100 to 22,065 bp).

Functional annotation was performed on the genes catalog using DIAMOND (Galaxy v 2.0.8.0) (Buchfink et al., 2014) and the UniProtKB swiss-prot functional database (from 2021_03) (The UniProt Consortium, 2019). Diamond was run with the blastx mode and default parameters. A total of 163,057 genes were annotated representing 40,264 unique UniProtKB entries.

The last step was to map non-rRNA MG and MT reads against the gene catalog to obtain the count values of each gene for each samples, using BWA-MEM (Galaxy v 0.7.17.1) (Li and Durbin, 2009). Default parameters were used with a mean insert length for paired reads of 250. The percentage of properly paired reads was between ~4 and ~17 % for MG reads mapped against the gene catalog and between ~3 and ~10 % for MT reads (**Supplementary Table 1**; **Supplementary Table 2**).

The gene count table for MG and MT was filtered to remove gene IDs affiliated to "Embryophytes" and "Metazoa" to keep only microbial genes. Also, only genes with a minimum of 10 mapped sequences for MG samples were kept. Finally, 21,046 gene IDs remained (over the 40,264) for downstream analyses.

Data normalization and differential functional expression analysis

The normalization and differential expression analysis (DEA) for MTs functional and taxonomic count data were performed with the recent R package MTXmodel (R v 4.0.3; MTXmodel v 1.5.1) (Zhang et al., 2021b). MTXmodel was run with the following options: no transformation, clr (centered-log ratio) normalization, LM analysis method, BH correction method, "EnvType" as fixed effect (i.e., cloud or air environment), min abundance at 0.0001, min prevalence at 0.5, max significance at 0.25 and input of DNA data (MGs count table). DEA gives coefficient of relative expression based on the fixed effect. Here, the effect is the environment type with "cloud" as reference, therefore positive coefficient (coeff) indicates that the feature (taxonomic group or gene) is significantly more expressed in clouds, while negative coeff indicates overexpression in aerosols. The absolute highest value illustrates highest overexpression. DEA was also performed to differentiate functional and taxonomic counts of MTs from MGs in order to have the taxa and functions significant expressed in general in aerosol and cloud samples. The tool settings were the same except that the input file contained all MT and MG samples, the fixed effect was "MT or MG" and there was not "DNA data" input for standardization. Here, only positive DEA coeff were considered to look only to features significantly more present in MT data, so overexpressed features.

Finally, for heatmap representations, RNA:DNA log ratios were done with data normalized as relative counts. RNA:DNA ratios are commonly used as an appraisal of the relative level of metabolic activity, with higher ratios indicating potentially higher metabolic activity (Baldrian et al., 2012; Zhang et al., 2014).

Gene ontology

Recovered genes in samples were grouped by ontology (GO) (Ashburner et al., 2000; Carbon et al., 2021). REVIGO (Supek et al., 2011) was used to summarize the GOs list and regroup them among Cellular Component, Molecular Function and Biological Process categories.

Results

Meteorological context and chemical signature

The main meteorological characteristics and sample information are summarized in **Table 1**. The samples were collected primarily in summer and fall at positive ambient temperature. The clouds were, necessarily, collected under colder and windier conditions than the aerosols. Relative humidity was logically oversaturated (> 100 %) in cloudy situations, whereas it was more of a continuum for aerosols (~41-78 %). Both cloud and aerosol events had a predominant origin from the western side (Atlantic Ocean) (**Table 1**; **Supplementary Figure 1**), with clouds being predominantly of marine origin (based on ion amounts and pH) (**Supplementary Table 3**). Indeed, the clouds were variable in terms of dissolved major inorganic ions, although the ranges were typical for atmospheric samples at these sites (~1 to hundreds µM) (Renard et al., 2020) (**Supplementary Table 3**), allowing an estimate of a primary marine or a continental origin for these. However, there was no correlation between meteorological context (temperature, wind speed, etc.), cloud water chemistry and biological content (biomass, ATP and nucleic acid) (**Supplementary Table 4B**). There was also no correlation between biological contents and meteorological context in aerosols (**Supplementary Table 4A**).

Table 1: Main sample information and meteorological conditions.

SampleID	Sampling date (dd/mm/yyyy)	Sampling duration (hour)	Season	Geographical origin[±]	Boundary layer height (min-max [average]) (m)[#§]	In/Out the boundary layer	Average temperature (°C)[#]	Relative humidity (%)[#]	Wind speed (average [max]) (m s^{-1})[#]	Liquid water content (LWC) (g/m3)[#]
AEROSOLS										
20200707AIR	07/07/2020	6.5	Summer	NW	1268-1834 [1626]	In	11.1	61	3.6	0
20200708AIR	08/07/2020	6.1	Summer	NW	623-1675 [1253]	In	14.2	53	3.1	0
20200709AIR	09/07/2020	6.0	Summer	N	651-2377 [1487]	In	20.3	48	3.4	0
20200922AIR	22/09/2020	5.9	Fall	W	665-1334 [972]	Out	12.4	78	1.0	0
20201118AIR	18/11/2020	5.8	Fall	W	680-1142 [870]	Out	14.1	41	6.4	0
20201124AIR	24/11/2020	6.0	Fall	W	644-740 [699]	Out	8.6	50	3.4	0
Minimum	-	**5.8**	**-**	**-**	**-**	**-**	**8.6**	**41**	**1.0**	**0**
Maximum	-	**6.5**	**-**	**-**	**-**	**-**	**20.3**	**78**	**6.4**	**0**
Median	-	**6.0**	**-**	**-**	**-**	**-**	**13.3**	**52**	**3.4**	**0**
Mean	-	**6.1**	**-**	**-**	**-**	**-**	**13.5**	**55**	**3.5**	**0**
Standard error	-	**0.2**	**-**	**-**	**-**	**-**	**4.0**	**13**	**1.7**	**0**
CLOUDS										
20191002CLOUD	02/10/2019	2.4	Fall	NW	1422-1505 [1465]	In	6.5	100	3.0	NA*
20191022CLOUD	22/10/2019	6.4	Fall	S	698-957 [813]	Out	5.7	100	8.7	NA*
20200311CLOUD	11/03/2020	4.1	Winter	W	964-1145 [1060]	Out	5.0	100	7.4	NA*
20200717CLOUD	17/07/2020	3.3	Summer	NW	1271-1437 [1343]	Out	10.1	100	1.6	0.08
20201016CLOUD	16/10/2020	4.7	Fall	NE	917-1034 [958]	Out	1.1	100	1.8	0.35
20201028CLOUD	28/10/2020	6.0	Fall	W	1026-1529 [1269]	Out	5.2	100	11.0	0.23
20201103CLOUD	03/11/2020	3.5	Fall	W	1126-1593 [1390]	In	2.2	100	8.7	0.06
20201110CLOUD	10/11/2020	3.1	Fall	SW	691-1276 [1016]	Out	5.9	100	2.5	0.07
20201119CLOUD	19/11/2020	2.8	Fall	W	1207-1234 [1215]	Out	0.3	100	7.7	0.11
Minimum	-	**2.4**	**-**	**-**	**-**	**-**	**0.3**	**100**	**1.6**	**0.06**
Maximum	-	**6.4**	**-**	**-**	**-**	**-**	**10.1**	**100**	**11.0**	**0.35**
Median	-	**3.5**	**-**	**-**	**-**	**-**	**5.2**	**100**	**7.4**	**0.10**
Mean	-	**4.0**	**-**	**-**	**-**	**-**	**4.7**	**100**	**5.8**	**0.15**
Standard error	-	**1.4**	**-**	**-**	**-**	**-**	**3.0**	**0**	**3.6**	**0.11**
P-value (Mann-Whitney test for clouds vs aerosols)	-	**0.04***	**-**	**-**	**-**	**-**	**0.003****	**0.001****	**0.44**	**0.003****

[±] From the backward trajectory plots detailed in Supplementary figure 1;
[#] Over the sampling period;
[§] Data extracted from ECMWF ERA5 model;
NA*: No data available;
* p-value significant (0.05 > p-value ≥ 0.01); ** p-value highly significant (0.01 > p-value).

Biomass, ATP and nucleic acid contents

Total cell number concentrations were not significantly different between cloud and aerosol samples with ~103 cells m-3 of air, but cell concentration were much more variables within clouds (coefficient of variation [cv] = 108) than for aerosols (cv = 57) (**Table 2**; **Figure 2**). ATP concentrations were significantly higher in aerosols with an average of 4.34 pmol m-3 of air (against 0.15 pmol m-3 in clouds).

Regarding nucleic acid concentrations, more DNA tended to be extracted from aerosol samples than from clouds and more RNA tended to be extracted from cloud samples. RNA:DNA content ratios were therefore significantly higher for clouds (3.92±1.69) compared with aerosols (1.24±1.10) (**Table 2**; **Figure 2**).

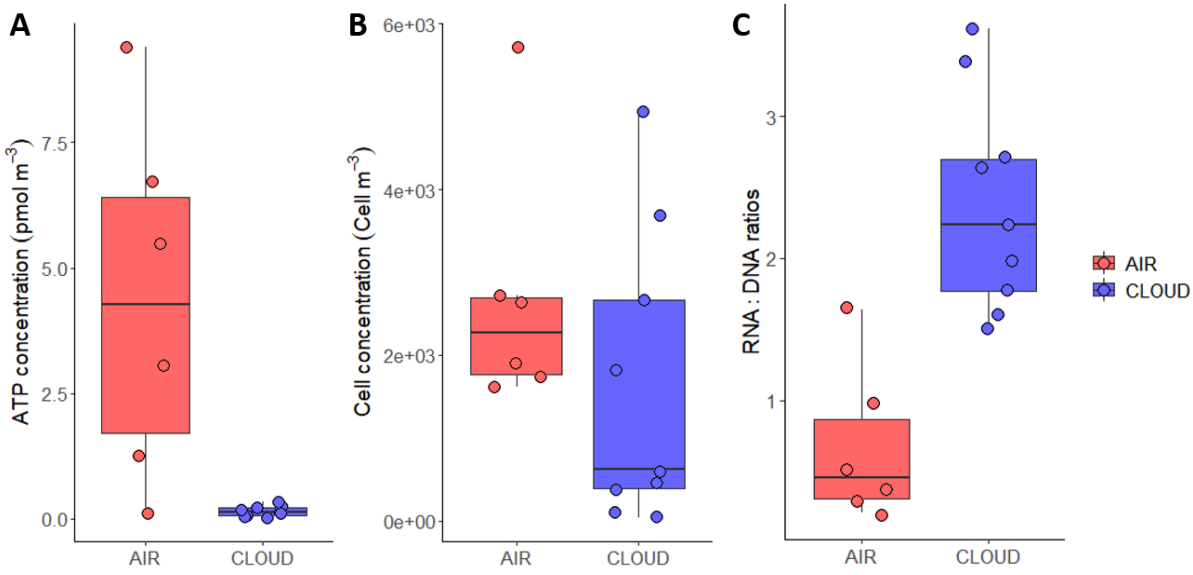


Figure 2: Biological data in aerosols (AIR, red) and clouds (blue). **A:** ATP concentration; **B:** cell concentration; **C:** RNA:DNA content ratios.

Table 2: Biological characteristics, nucleic acid concentrations, and gene number in samples

Sample ID	[Microbial cell] (cells m^{-3} of air)[#]	[ATP] (pmol m^{-3} of air)[#]	[DNA] (ng m^{-3} of air)[±]	[RNA] (ng m^{-3} of air)[±]	RNA:DNA ratio	Nb of annotated gene in MG data	Nb of annotated gene in MT data	MT/MG ratio for annotated gene
AEROSOLS								
20200707AIR	1.62×10^{3}	6.72	2.19	0.47	0.21	5,463	2,237	0.41
20200708AIR	1.72×10^{3}	3.06	1.35	0.38	0.28	3,287	2,095	0.64
20200709AIR	2.62×10^{3}	9.40	0.87	0.33	0.38	6,649	1,676	0.25
20200922AIR	1.91×10^{3}	5.47	0.70	0.68	0.98	11,690	2,248	0.19
20201118AIR	2.72×10^{3}	1.27	0.14	0.23	1.64	15,431	4,471	0.29
20201124AIR	5.72×10^{3}	0.12	0.12	0.06	0.53	NA[$]	NA[$]	NA[$]
Minimum	**1.62×10^{3}**	**0.12**	**0.12**	**0.06**	**0.21**	**3,287**	**1,676**	**0.19**
Maximum	**5.72×10^{3}**	**9.40**	**2.19**	**0.68**	**1.64**	**15,431**	**4,471**	**0.64**
Median	**2.26×10^{3}**	**4.26**	**0.78**	**0.36**	**0.46**	**6,649**	**2,237**	**0.29**
Mean	**2.72×10^{3}**	**4.34**	**0.89**	**0.36**	**0.67**	**8,504**	**2,545**	**0.36**
Standard error	**1.54×10^{3}**	**3.51**	**0.78**	**0.21**	**0.55**	**4,951**	**1,101**	**0.18**
CLOUDS								
20191002CLOUD	1.80×10^{3}	0.13	0.18	0.32	1.77	19,010	2,219	0.12
20191022CLOUD	4.95×10^{3}	0.06	0.24	0.65	2.70	14,804	3,855	0.26
20200311CLOUD	4.75×10^{2}	0.04	0.31	0.62	1.98	14,220	985	0.07
20200717CLOUD	1.20×10^{2}	0.18	0.33	0.52	1.59	12,491	1,477	0.12
20201016CLOUD	2.67×10^{3}	0.24	0.16	0.23	1.49	17,912	6,134	0.34
20201028CLOUD	3.67×10^{3}	0.33	0.16	0.36	2.24	16,271	3,368	0.21
20201103CLOUD	6.17×10^{2}	0.22	0.37	0.97	2.65	15,958	2,737	0.17
20201110CLOUD	3.90×10^{2}	0.12	0.38	1.27	3.38	16,527	3,110	0.19
20201119CLOUD	3.66×10^{1}	0.02	0.28	1.00	3.62	16,064	3,406	0.21
Minimum	**3.66×10^{1}**	**0.02**	**0.16**	**0.23**	**1.49**	**12,491**	**985**	**0.07**
Maximum	**4.95×10^{3}**	**0.33**	**0.38**	**1.27**	**3.62**	**19,010**	**6,134**	**0.34**
Median	**6.17×10^{2}**	**0.13**	**0.28**	**0.62**	**2.24**	**16,064**	**3,110**	**0.19**
Mean	**1.64×10^{3}**	**0.15**	**0.27**	**0.66**	**2.38**	**15,917**	**3,032**	**0.19**
Standard error	**1.77×10^{3}**	**0.10**	**0.09**	**0.35**	**0.76**	**1,934**	**1,496**	**0.08**
P-value (Mann-Whitney test)	**0.21**	**0.01***	**0.32**	**0.16**	**0.004****	**0.01***	**0.59**	**0.05**

[ ] concentration; MG: Metagenomic; MT: Metatranscriptomic;

[#] concentrations per m^3 of air were calculated based on the time of sampling and the sampler air-flow rate for aerosols, and based on liquid water content (LWC) for clouds;

[±] concentrations per m^3 of air were calculated based on the time of sampling and the HFR impinger air-flow rate for both aerosols and clouds;

NA[$]: No data available,

* p-value significant ($0.05 > \text{p-value} \geq 0.01$); ** p-value highly significant ($0.01 > \text{p-value}$).

352 **Microbial diversity in clouds and aerosols**

353 Taxonomic affiliation was performed using whole MGs data (including both rRNA and
354 non-rRNA reads). A majority of bacterial taxa were found (~50 % of the affiliated sequences
355 in clouds and ~90 % in aerosols; **Supplementary Figure 2**), distributed over 33 phyla and 160
356 orders (**Supplementary Table 5**). The most abundant phyla were Proteobacteria (Alpha,
357 Gamma and Beta-Proteobacteria), Actinobacteria, Bacteroidetes, and Firmicutes
358 (**Supplementary Figure 3**). The bacterial genera commonly reported in the atmosphere were
359 found abundant in both aerosols and clouds in our study, such as: Hymenobacter,
360 Staphylococcus, Streptococcus, Acinetobacter, Streptomyces; and several phyllosphere
361 associated genera, Pseudomonas, Sphingomonas, Methylobacterium and Massilia
362 (**Supplementary Figure 4**).

363 Eukaryotic taxa represented ~50 % of the number of affiliated sequences in clouds and
364 only ~10 % of the sequences in aerosols (**Supplementary Figure 2**), and were distributed over
365 9 phyla and 22 orders (**Supplementary Table 6**). Fungi were dominant with the most abundant
366 eukaryotic phylum Ascomycota (e.g., orders Helotiales, Hypocreales, Mycosphaerellales and
367 Saccharomycetales) and the second most abundant phylum Basidiomycota (e.g., orders
368 Tremellales, Ustilaginales and Malasseziales). There were also the three phyla Euglenozoa,
369 Evosea and Cercozoa regrouping single-celled eukaryotes with mainly amoeboids and
370 flagellates, the phylum Bacillariophyta regrouping microalgae, and the phylum Apicomplexa
371 composed only of unicellular animal parasites (**Supplementary Figure 5** and **6**).

372 Three archaeal phyla were also present at low abundance (0.04% of the number of affiliated
373 sequences) with the dominant phylum Euryarchaeota and in a lesser extent Thaumarchaeota
374 and Crenarchaeota. The main archaeal orders were Methanosarcinales, Methanomicrobiales
375 and Methanobacteriales (Euryarchaeota phylum) which contain exclusively methanogen
376 archaea. The order Nitrososphaerales (Thaumarchaeota phylum) was also abundant and

encompasses the ammonia-oxidizing archaea (**Supplementary Figure 7**; **Supplementary Table 7).**

Finally, viral sequences were observed (0.02 % of affiliated sequences) with mainly bacteriophages (Caudovirales order) and some eukaryotic viruses (Lefavirales and Bunyavirales orders) (**Supplementary Table 8**).

Regarding hierarchical clustering (**Supplementary Figure 4** and **6**), PCA (**Supplementary Figure 8**) and NMDS (no solution reached) results, the microbial diversity in clouds and aerosols was not significantly different. The three aerosol samples collected in summer were very similar likely related with their temporal proximity (3 consecutive days).

Potentially active microbial taxa

In contrast to the MG data, there was a majority of eukaryote-affiliated sequences in the MT data (~85 %) compared to bacterial sequences in the cloud samples, and also slightly more eukaryote-affiliated sequences compared to MG data (~30%) in aerosol samples (**Supplementary Figure 2**). However, richness of active taxa was greater in bacteria (**Supplementary Figure 9**) with 84 families (22 % of total families; correspond to 211 bacterial genera) versus 18 eukaryotic families (51 % of total families; correspond to 30 genera) (according to DEA of MT data versus MG data; **Supplementary Table 9**). Furthermore, bacterial families were overall more potentially active than eukaryotic families with DEA coefficients between 0.3 to 3.2 for bacteria significantly more represented in MT data compared to MG and up to 1.7 only for eukaryotes. Among the most potentially active bacterial families were Halomonadaceae (DEA coefficient: 3.23) (Gamma-proteobacteria), Rickettsiaceae (2.43) (Alpha-proteobacteria), Mycoplasmataceae (3.03), Clostridiaceae (3.02), Peptoniphilaceae (2.58), Peptostreptococcaceae (2.04) (Firmicutes), Amoebophilaceae (3.13) (Bacteroidetes), Chroococcaceae (2.47) (Cyanobacteria) and Treponemataceae (2.14) (Spirochaetota). Main abundant genera such as *Hymenobacter*, *Cutibacterium*, *Streptococcus*, *Sphingomonas* or

403 *Massilia* were not found significantly active (not represented in significant results from DEA).
404 However, *Acinetobacter* (1.69), *Clostridium* (3.40), *Pasteurella* (2.34) and *Pseudomonas* (0.51)
405 were found significantly overrepresented, with the two last one active only in clouds (log
406 RNA:DNA ratio >0) (**Figure 3**). Concerning eukaryotes, the algae phylum Cryptophyta was
407 the most represented in potentially active eukaryotes with the three families Hemiselmidaceae
408 (1.69), Geminigeraceae (1.60) and Cryptomonadaceae (1.29). The families Phaeodactylaceae
409 (1.31) (Bacillariophyta, diatoms) and Plasmodiidae (1.06) (Apicomplexa) were also among the
410 most overrepresented in eukaryotes (**Supplementary Table 9**). Among the most abundant
411 eukaryotic genera, it was mainly fungi which were potentially active, with also the diatom
412 *Phaeodactylum* found active in both clouds and aerosols (**Figure 3**).

413 In terms of overall active taxa, clouds and aerosols were not different according to the
414 hierarchical clustering (**Supplementary Figure 9**) and PCA results (**Supplementary Figure**
415 **8**). Still, some bacterial and eukaryotic families were significantly more active in clouds or
416 aerosols (6 bacterial families and 5 eukaryotic families; **Supplementary Figure 10**). In clouds,
417 Parachlamydiaceae (1.54), Halomonadaceae (1.19), Legionellaceae (0.58) (bacterial families),
418 Glomerellaceae (1.26), Cryptomonadaceae (0.94) and Trichomonadaceae (0.89) (eukaryotic
419 families) were significantly expressing more functions. On the contrary, Bryobacteraceae (-
420 1.73), Pirellulaceae (-1.37), Ornithinimicrobiaceae (-1.07) (bacterial families), Dipodascaceae
421 (-1.58) and Theileriidae (-1.41) (eukaryotic families) were expressing more functions in
422 aerosols.
423

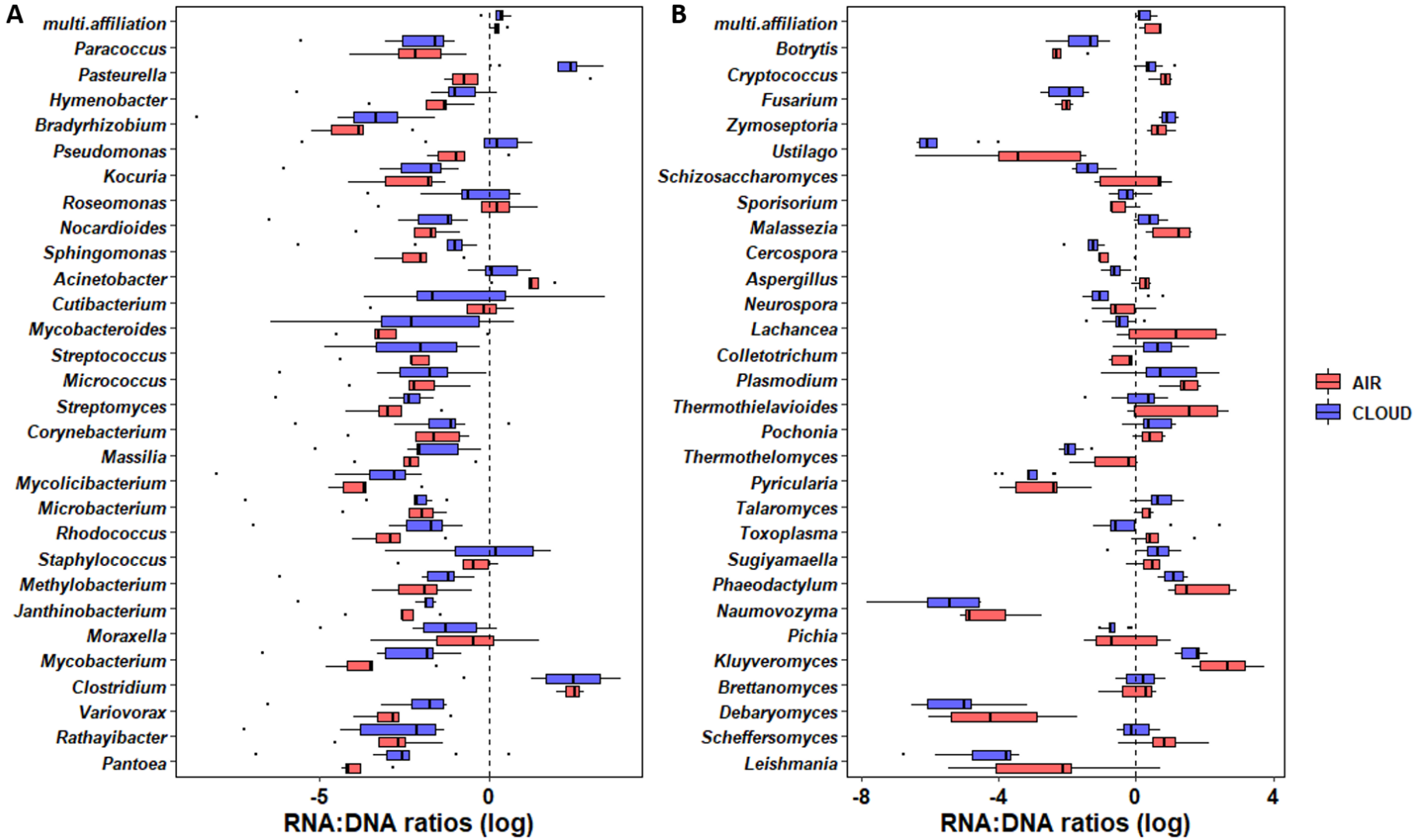


Figure 3: RNA:DNA content ratios for (A) the 50 most abundant bacterial genera and (B) the 53 total eukaryotic genera. Taxa are ranked in descending order of number of reads in the metagenomic data (top to bottom).

427 **Gene catalog**

428 A catalog of genes was obtained from cloud and aerosol MGs (total of 2,832,534 contigs).
429 Approximately 15% of the genes in the catalog were annotated, represented ~5% of the total
430 predicted genes, indicating a high proportion of potentially undiscovered genes. The 40,264 unique
431 gene entries recovered from UniProt was equal to ~7% of the total UniProtKB swiss-prot database
432 (566,996 entries in March 2022) which illustrates the high diversity in the atmosphere. A few
433 UniprotKB entries were overrepresented in the gene catalog (mainly retroviruses and transposons
434 from eukaryotes). The vast majority (91.5 %) of the annotated genes were from eukaryotes, with
435 ~60 % affiliated with fungi, 23.6 % with viridiplantae and 14.3 % with metazoa. The remaining
436 were represented by bacterial genes (7.6 %), viral genes (0.6 %) and archaeal genes (0.3 %)
437 (**Figure 4**).
438

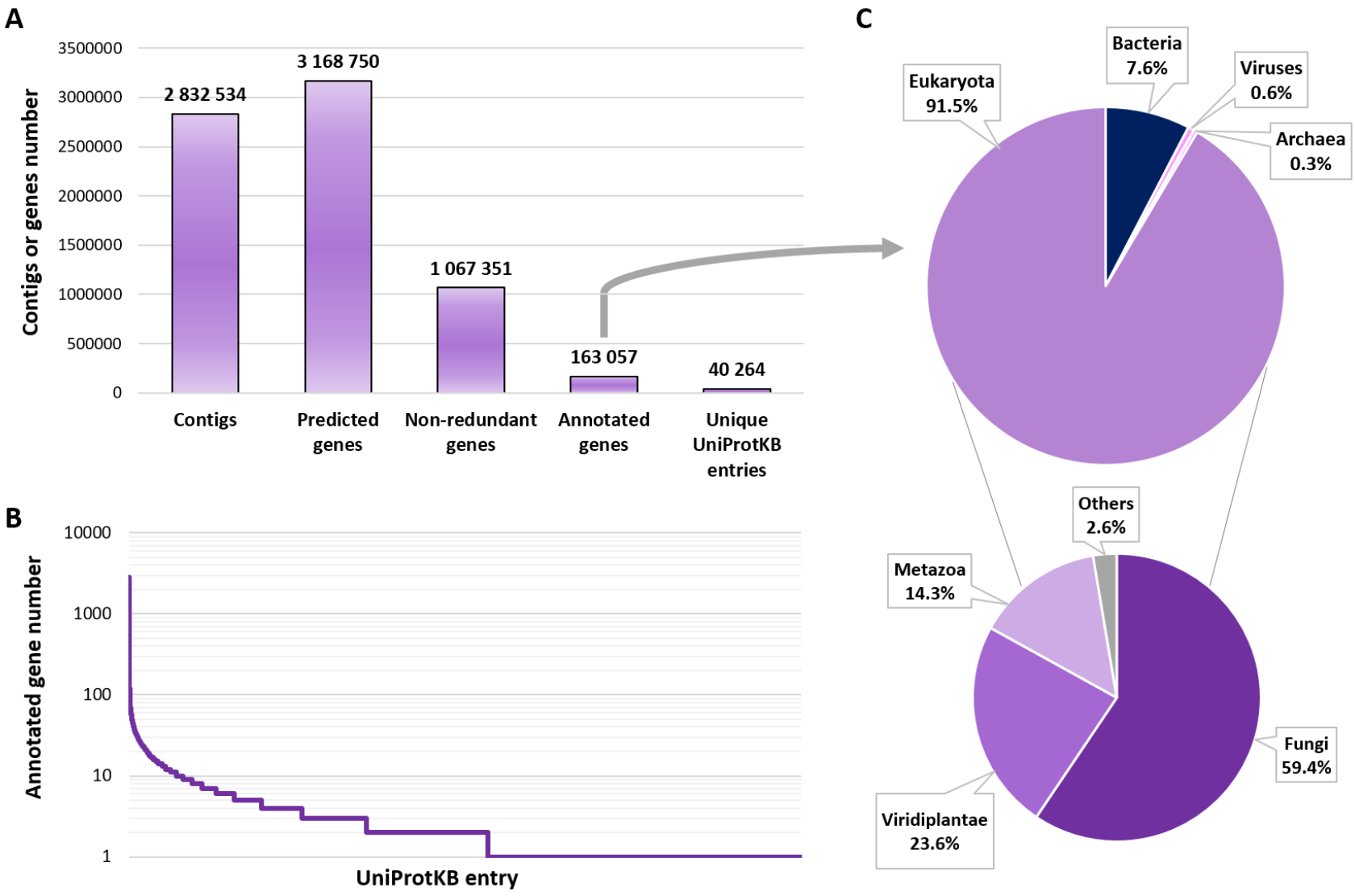


Figure 4: Gene catalog statistics. **A**: Contigs or genes numbers at each step of the gene catalog construction; **B**: rank-abundance curve for UniProtKB entries; **C**: percentages of annotated genes for each taxonomic kingdom and eukaryotic sub-kingdom.

Functioning of microorganisms in both clouds and aerosols

A total of 21,046 gene entries were detected, with 488 genes overexpressed (according to DEA for MT vs MG data). These latter genes were mainly from eukaryotes (~80%), in particular from fungi (~93% of the eukaryotic genes). For the domain of bacteria, overexpressed genes were mostly affiliated with Gamma-proteobacteria (~47%) and Actinobacteria (~25%) (**Supplementary Figure 11**). A total of 1,005 Gene Ontologies (GOs) were recovered from these 488 genes. The 20 most overrepresented GOs for each category (Cellular Components, Biological Processes and Molecular Functions) are presented in **Figure 5** (and the 40 most overrepresented in **Supplementary Table 10**).

First, several biological processes related to central metabolic pathways (e.g., carbohydrate metabolic process, glycolytic process, tricarboxylic acid cycle) and protein biosynthesis processes (translation and cytoplasmic translation) were overrepresented in MT data (**Figure 6-A,E**). Metabolic, biosynthetic and catabolic processes were active also in both clouds and aerosols (**Figure 7-A,C,E**). Interestingly, several processes associated to compounds such as mannitol (metabolic and catabolic processes), glycine and glycerol (biosynthesis) were active as well. Finally, GOs associated with energy metabolism (e.g., ATP synthesis, carbon utilization, respiratory electron transport chain) as well as the cell cycle (e.g., cell division, DNA replication, nuclear division) were also overrepresented in the MT data (**Figure 6-C,E**), all together testifying to microbial survival and activity in the atmosphere.

Second, components linked to the cell surface and outer membrane, as well as biological processes related to transport and transmembrane transport (e.g., carbohydrate, phospholipid, protein, peptide transport) (**Figure 8-A**) were strongly represented in the MT data. Membranes are key components of cellular integrity and exchange and responses to the cellular environment.

466 This leads us to the overexpression of genes in clouds and aerosols testifying of potential
467 responses to stressful atmospheric conditions such as hydrogen peroxide catabolic processes
468 and response, responses to oxidative stress and to osmotic stress, regulation of intracellular pH
469 (GO:0051453) (DEA coefficient: 6.10), DNA repair (GO:0006281) (5.60), autophagy and
470 macropexophagy, stress-activated MAPK cascade, responses to starvation (glucose and amino
471 acid), and the use of the pentose phosphate shunt (GO:0006098) (1.01) (**Figure 8-C**). Isocitrate
472 lyase activity (GO:0004451) (1.27) was interestingly also overexpressed in both aerosol and
473 cloud samples. This enzyme is the key to enter the glyoxylate cycle for microorganisms
474 adapting to the lack of complex sugar.

475 As a last point, the top overrepresented GOs (over 169 GOs) were linked with fungi (e.g.,
476 yeast-form cell wall components, hyphal cell wall components) and micro-algae (e.g.,
477 chloroplast components, and photosystem I), in accordance with taxonomic observation. The
478 components of the photosystem II and of the thylakoid in cyanobacteria were as well among
479 the most overrepresented. Indeed, biological processes and molecular functions related to
480 photosynthesis (e.g., response to herbicide, photosynthetic electron transport in photosystem II;
481 and oxygen evolving activity and electron transport pathway of photosynthesis activity) were
482 among the most overrepresented (**Figure 5**), as well as GOs related to responses to light (e.g.,
483 response to light stimulus, phototransduction). This could indicate significant photosynthetic
484 activity in the atmosphere, with the presence of many airborne chlorophyll organisms.
485

Figure 5: The 20 most overrepresented Cellular Components (among 169 GOs), Biological Processes (among 506 GOs), and Molecular Functions (among 329 GOs) in clouds and aerosols. DEA: differential expression analysis; GO: gene ontology.

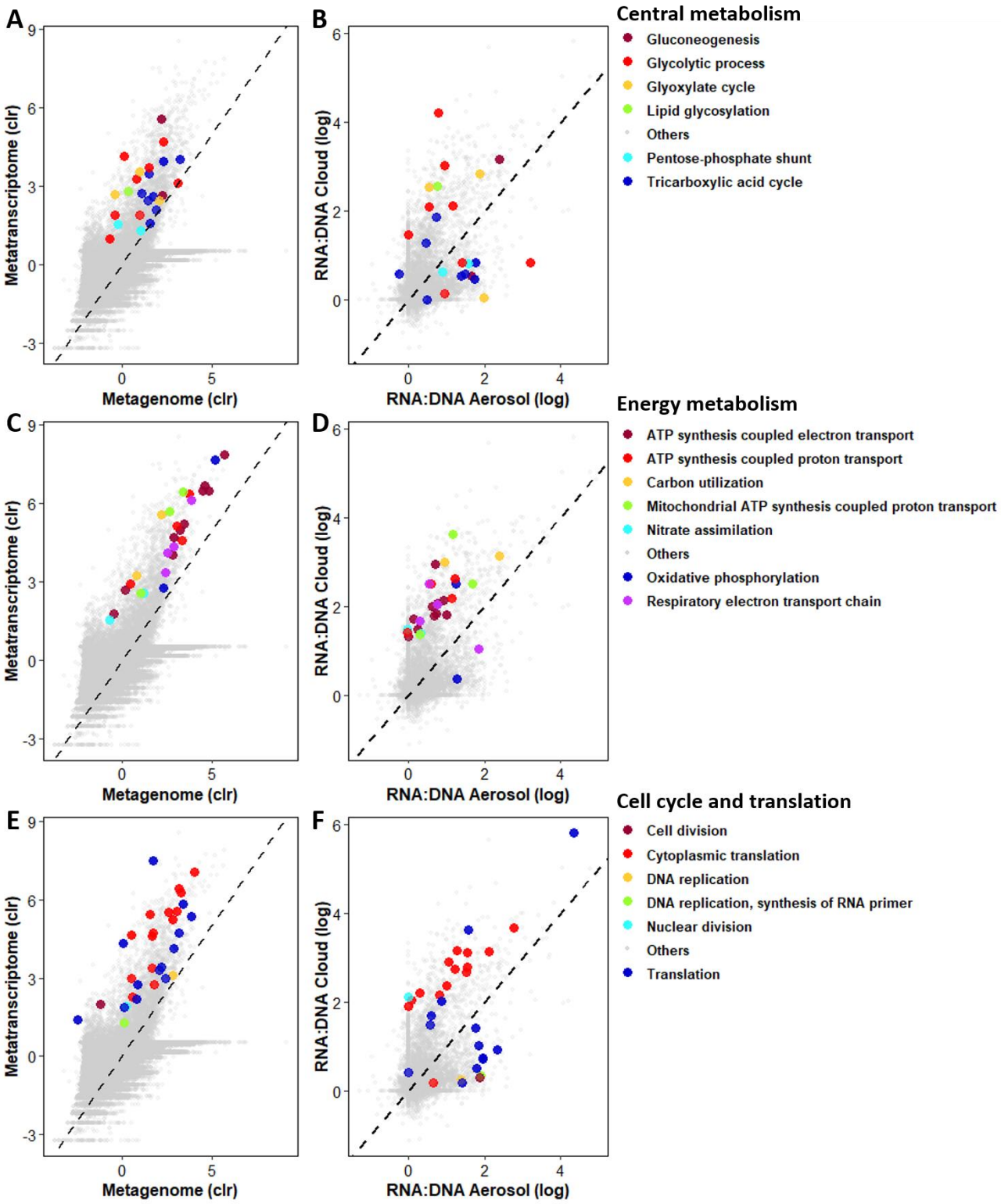


Figure 6: Genes expression regarding (A; C; E) metatranscriptome versus metagenome data; (B; D; F) and cloud versus air RNA:DNA content ratios for Biological Processes (Gene Ontologies category) related to (A; B) central metabolism, (C; D) energy metabolism, and (E; F) cell cycle and translation. clr: centered log ratio transformation.

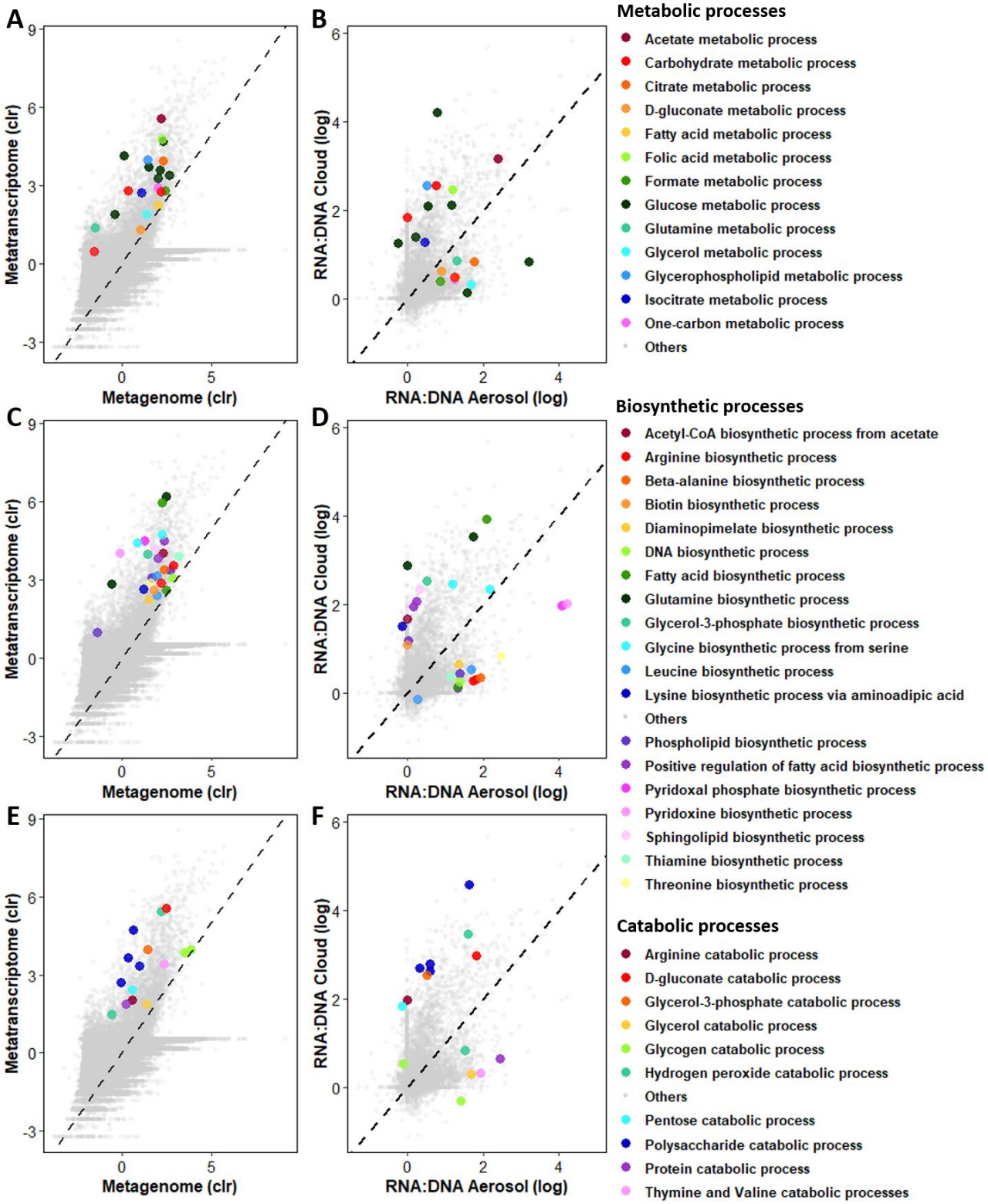


Figure 7: Genes expression regarding (A; C; E) metatranscriptome versus metagenome data; (B; D; F) and cloud versus air RNA:DNA content ratios for Biological Processes (Gene Ontologies category) related to (A; B) metabolic, (C; D) biosynthetic, and (E; F) catabolic processes. clr: centered log ratio transformation.

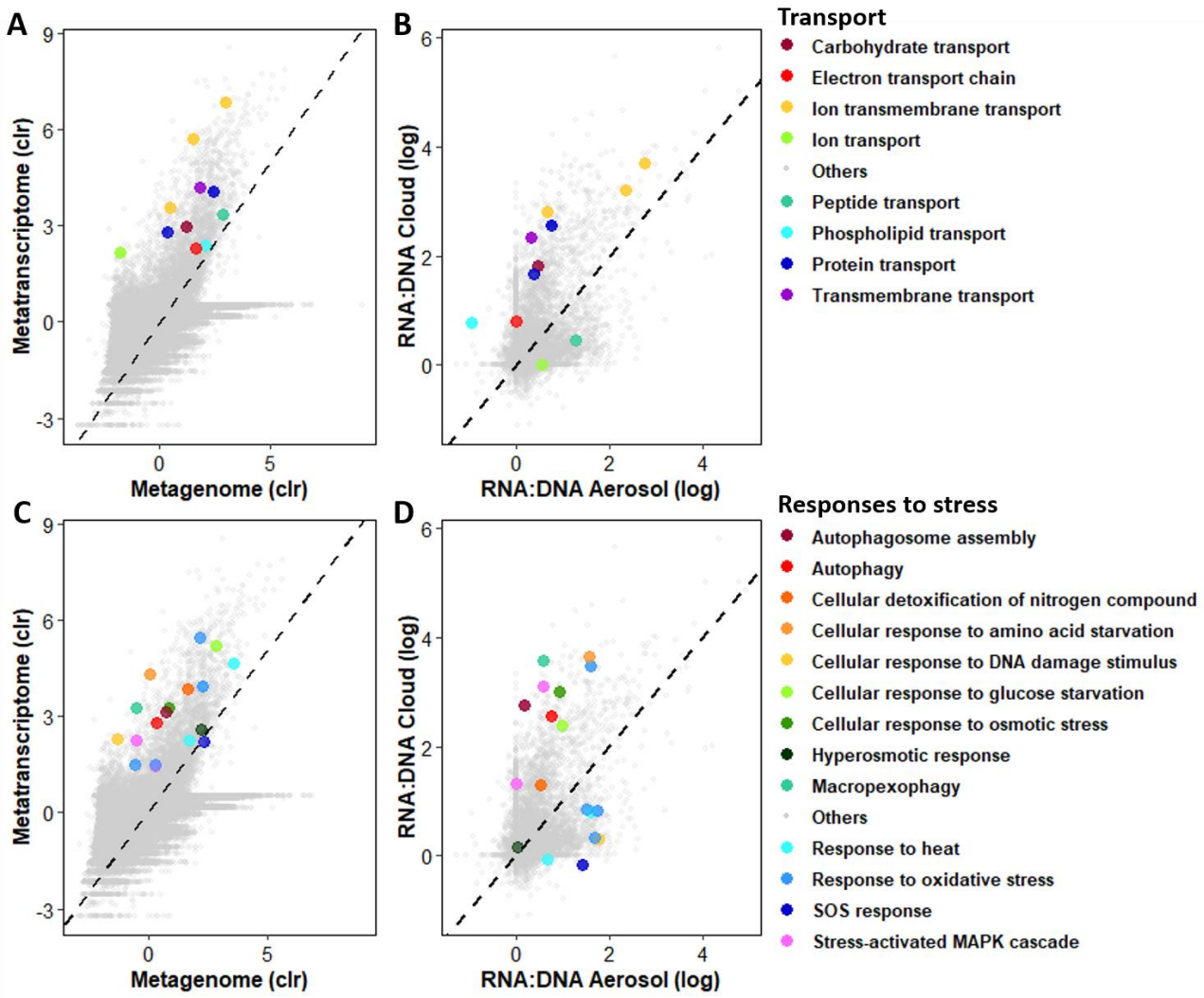


Figure 8: Genes expression regarding (A; C) metatranscriptome versus metagenome data; (B; D) and cloud versus air RNA:DNA content ratios for Biological Processes (Gene Ontologies category) related to (A; B) transport, (C; D) responses to stress. clr: centered log ratio transformation.

Significantly differentially expressed functions between clouds and aerosols

According to the NMDS results including all 21,046 annotated genes (**Figure 9**), The MTs of clouds and aerosols differed, revealing a likely distinct biological functioning. Hierarchical clustering based on RNA:DNA content ratios showed that aerosols were clustered together, with the exception of the air sample A0922, which was clustered with most clouds, and cloud sample C1016, which was grouped with aerosols (**Supplementary Figure 12**). Moreover, the heatmap of RNA:DNA content ratios revealed large clusters of genes expressed in clouds, and less in aerosols. This observation is confirmed by the RNA versus DNA plots (**Supplementary Figure 13**), with higher levels of RNA detected in clouds.

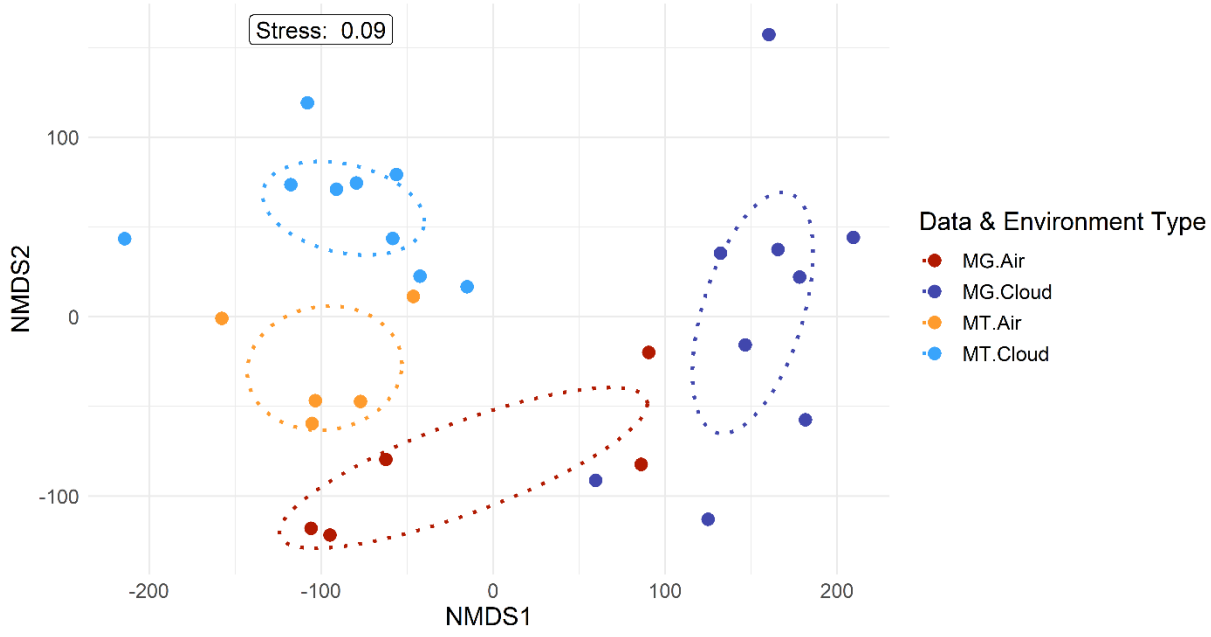


Figure 9: NMDS analysis based on the 21,046 functional gene entries for metagenome (MG) and metatranscriptome (MT) dataset coloured by environment type (Cloud or Air, i.e. aerosols). Ellipses are calculated based on covariance matrix (function *veganCovEllipse*, Vegan R package).

A total of 320 genes were found significantly differentially expressed between clouds and aerosols. Based on these genes, 820 GOs were identified: 147 cellular components, 394 biological processes and 279 molecular functions. Clouds contained more overrepresented GOs than aerosols (**Figure 10**). Most differentially expressed GOs are presented by categories in **Supplementary Table 11** and **Figures 11**, **12** and **13**.

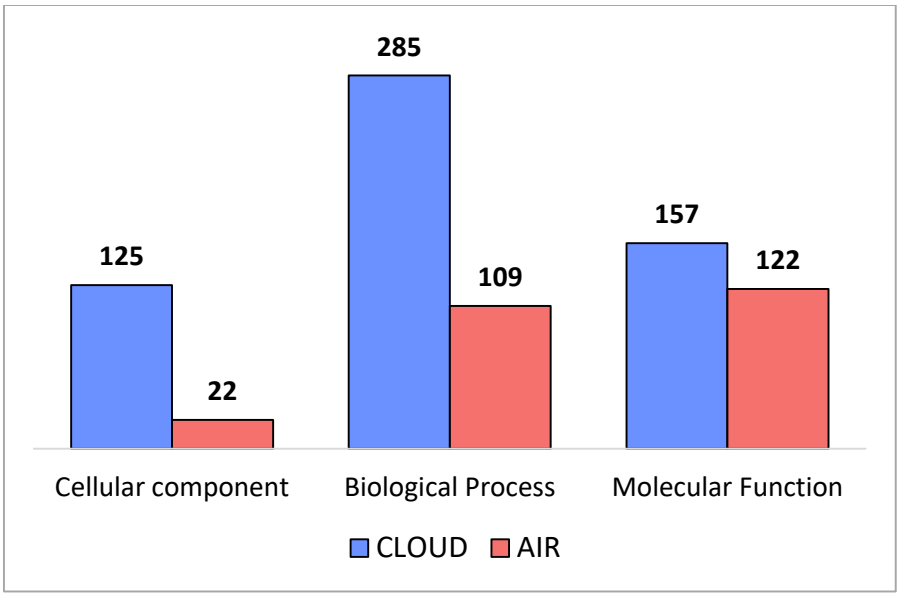


Figure 10: Number of overrepresented gene ontologies (GO) in clouds and in aerosols (i.e., air).

First, concerning aerosols, central metabolism, cell cycle and protein synthesis were expressed with several significantly more represented processes compared to cloudy situations, such as DNA replication, the pentose-phosphate shunt, cell cycle, cell division and transcription (**Figure 12**; **Supplementary Table 11**), but also the tricarboxylic acid cycle (GO:0006099) (DEA coefficient: -0.68), and translation (GO:0006412) (-0.76). Regarding cloudy situations, lipid glycosylation (GO:0030259) (0.86) and cytoplasmic translation (GO:0002181) (0.97) were significantly overrepresented compared with dry air. However, regarding to the potential expression of each gene separately for these GOs associated to central metabolism and cell cycle several genes (sometimes the majority) related to glycolytic process, glyoxylate cycle,

tricarboxylic acid cycle and translation were more active in clouds compared to aerosols (**Figure 6-B,F; Supplementary Figure 15**).

Second, several metabolic, biosynthetic and catabolic pathways were as well overexpressed in dry atmosphere such as citrate (GO:0006101) (-1.36) metabolism, pyridoxine (GO:0008615) (-2.00), pyridoxal phosphate (GO:0042823) (-2.07) biosynthesis and protein (GO:0030163) (-1.98), hydrogen peroxide (GO:0042744) (-0.16) and glycerol (GO:0019563) (-1.62) catabolism (**Figure 7-B,D,F**). In contrast, acetate (GO:0006083) (0.71), glucose (GO:0006006) (0.45), carbohydrate (GO:0005975) (0.53), isocitrate (GO:0006102) (1.50) metabolisms, glutamine (GO:0006542) (1.61) biosynthesis and polysaccharide (GO:0000272) (1.61) catabolism processes were, among others, noticeably significantly more represent in clouds.

Third, regarding energy metabolism (**Figure 6-D**), the main biological processes were significantly overrepresented in clouds (e.g., ATP synthesis, carbon utilization, respiratory electron transport chain), as well as cellular components related to respiratory chain and ATP synthase complex (particularly in mitochondria) (**Supplementary Figure 14**), testifying of a higher energy potential than in aerosols. Only oxidative phosphorylation was significantly more represented in aerosols (**Figure 6**; **Supplementary Figure 16**).

Interestingly, transport and transmembrane transport processes of different compounds (e.g., carbohydrates, proteins) were overall significantly overrepresented in clouds (**Figure 8-B**), perhaps reflecting greater interaction of cells with their environment due to the aqueous environment. Also, cytoplasmic vesicle membrane and peroxisome components were more present and expressed in clouds compared to dry situations (**Figure 11**).

Finally, different responses to the environment were overrepresented such as, cellular responses to starvation (also to glucose and amino acid starvation), pexophagy (i.e. macropexophagy) and other autophagy processes, stress-activated MAPK cascade, cellular

responses to nitrosative stress and to osmotic stress (with also hyperosmotic response), cellular detoxification of nitrogen compound (**Figure 12 and 14**), regulation of intracellular pH (GO:0051453) (1.19) and response to UV (GO:0034644) (1.63) significantly overrepresented in clouds (**Figure 8-D**; **Supplementary Table 11**). In aerosols responses to heat, to DNA damage stimulus, to oxidative stress and SOS response were significantly overrepresented (**Figure 8-D, 12 and 14**; **Supplementary Table 11**).

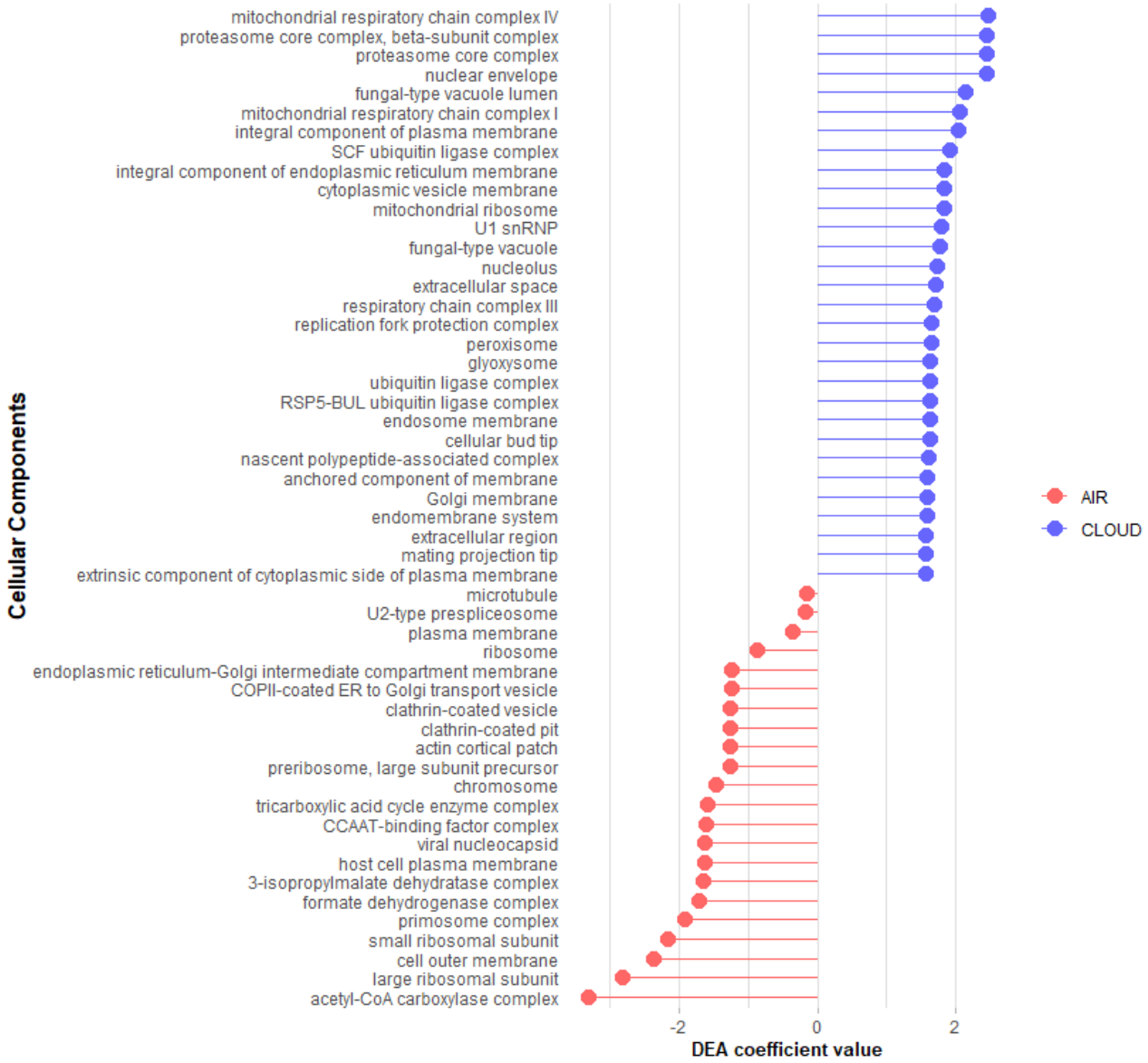
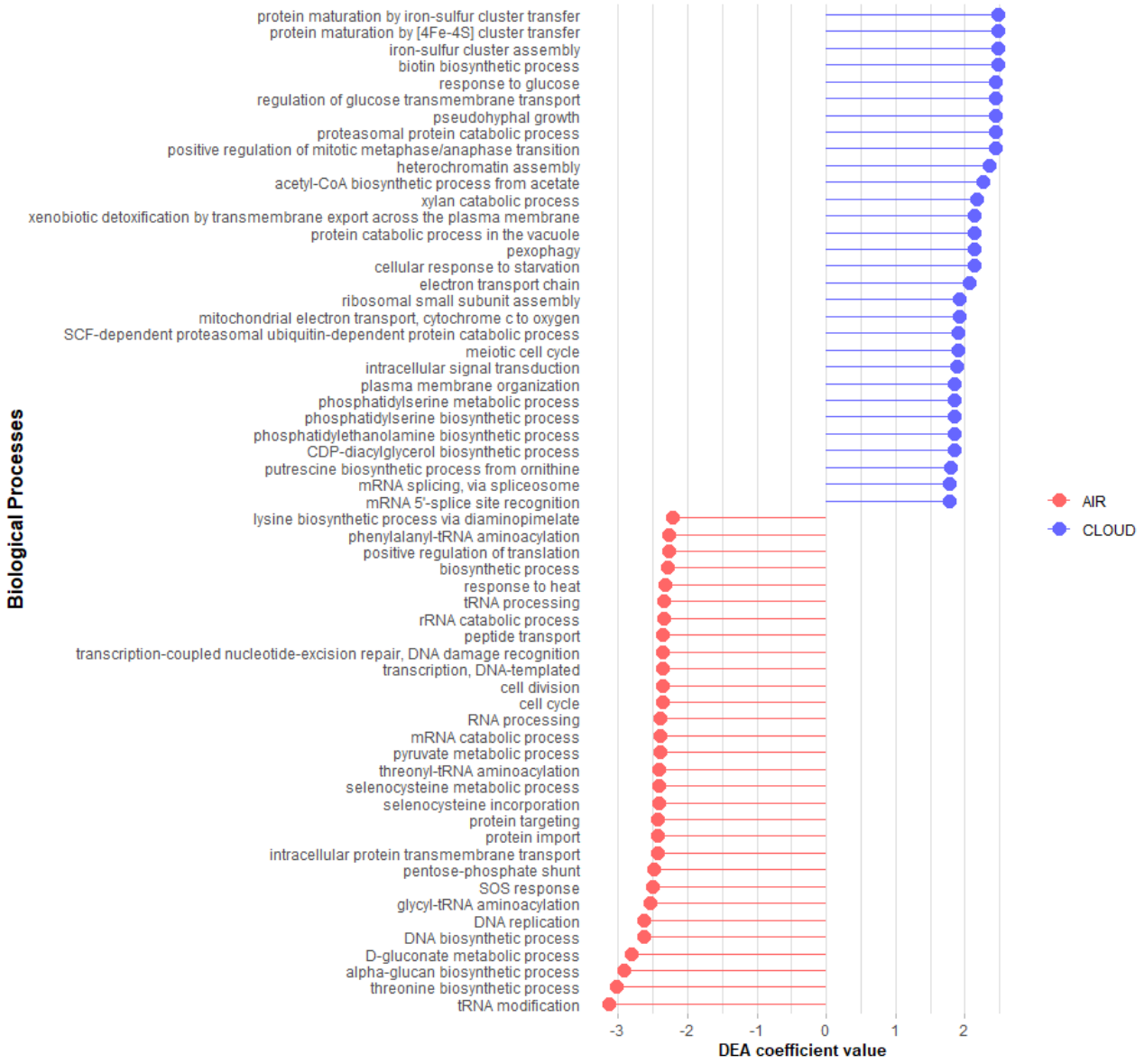


Figure 11: The 30 most overrepresented Cellular Components in clouds (blue) (among 125 GOs) and the total 22 overrepresented in aerosols (red). DEA: differential expression analysis; GO: gene ontology; AIR: aerosol.

572



573

574 **Figure 12: The 30 most overrepresented Biological Processes in clouds (blue) (among 285**
575 **GOs) and in aerosols (red) (among 109 GOs).** DEA: differential expression analysis; GO:
576 gene ontology; AIR: aerosol.

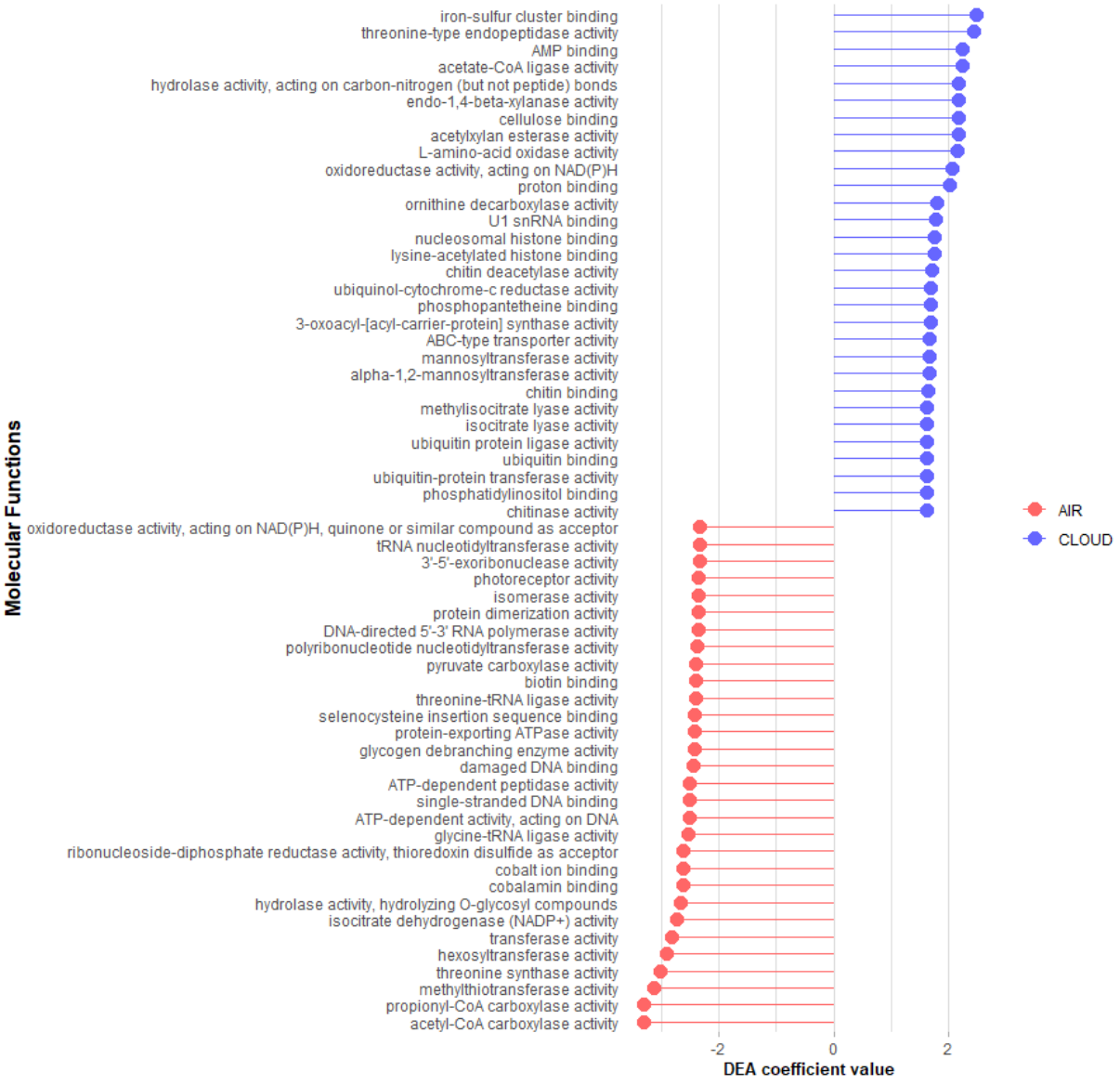
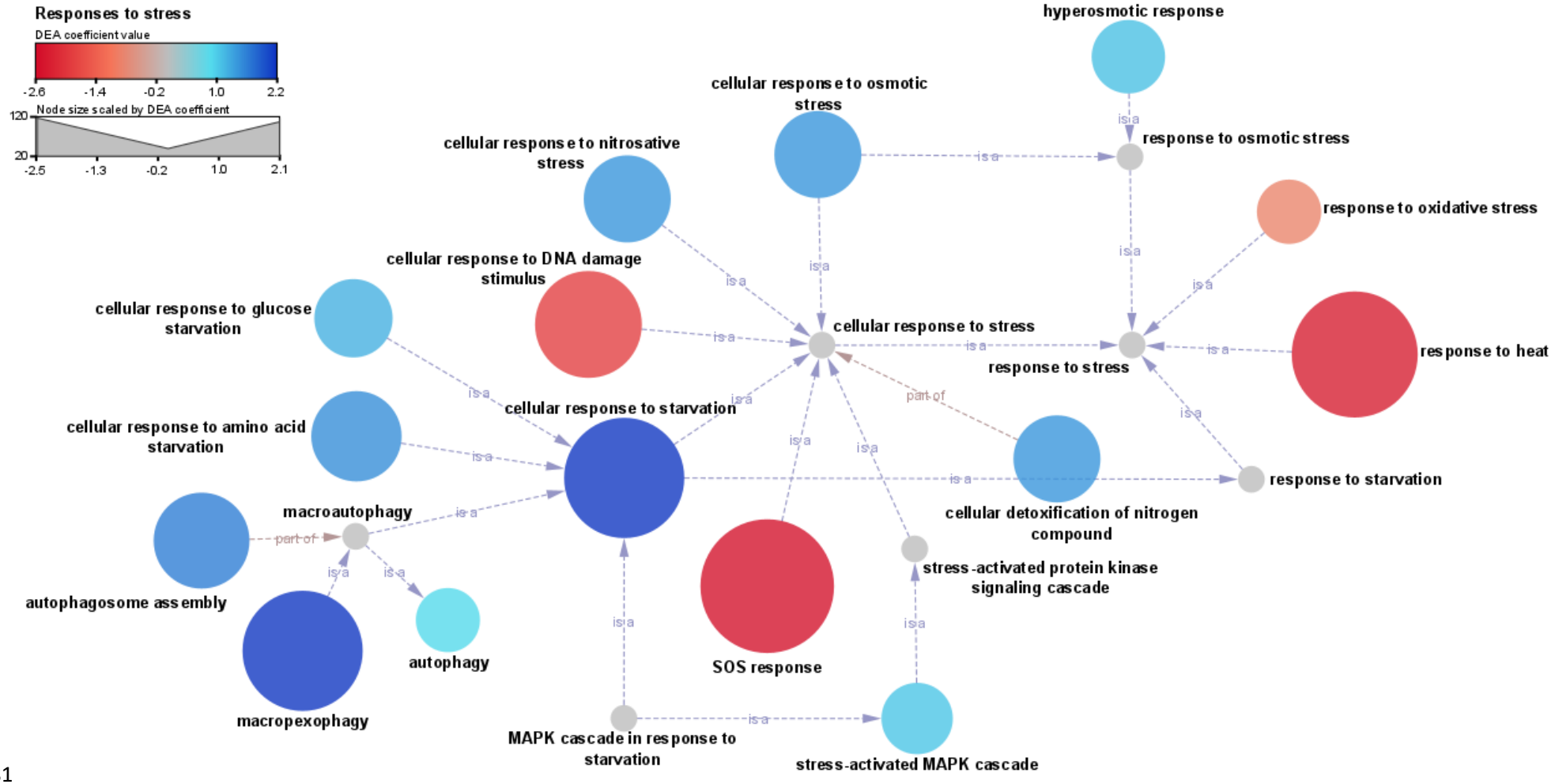


Figure 13: The 30 most overrepresented Molecular Functions in clouds (blue) (among 157 GOs) and in aerosols (red) (among 122 GOs). DEA: differential expression analysis; GO: gene ontology; AIR: aerosol.



581

582 **Figure 14: Gene ontology (GO) relationship tree for Biological Processes related to stress responses in clouds and aerosols.** The red to blue
583 colour scale represents the Differential Expression Analysis (DEA) coefficient value, with negative values (red shades) indicating a significant
584 overrepresentation in aerosols as opposed to positive values (blue shades) signifying an overrepresentation in clouds. The size of the nodes is scaled
585 by the absolute value of the DEA coefficient.

Discussion

Cloud and aerosol microbial communities do not differ in composition

The microbial diversities in clouds and aerosols were in agreement with typical airborne microbial communities (Amato et al., 2019; Fröhlich-Nowoisky et al., 2012; Tignat-perrier et al., 2019) and not clearly distinct between the two atmospheric situations. The distinction between seasons was not apparent unlike what was observed by Tignat-Perrier et al. (2020) and **Chapter 2**, **article 2** at this site. However, this may be explained by the small number of samples collected in summer and winter and the lack of sampling replicates here unlike in **article 2**.

Bacteria represented ~50 % of the affiliated ribosomal and non-ribosomal sequences in cloud samples and the majority of the affiliated sequences in aerosols, which was not expected given that eukaryotes have a higher number of rDNA copy (1 to thousands copies for fungi and even more for some protist, compared to 1-7 copies for main bacterial genomes) (Lavrinienko et al., 2021; Lofgren et al., 2019; Medinger et al., 2010; Stoddard et al., 2015). The low content of eukaryotic sequences in aerosols compared to clouds could possibly be explained by the presence of condensed water (thus higher relative humidity) and more autumnal meteorological conditions (compared to aerosols collected mainly in summer) favouring the survival of fungal cells, or by a seasonal and punctual effect of emission sources. Indeed, atmospheric distribution and abundance of fungal cells and spores were correlated with relative humidity and temperature (Almaguer et al., 2014; Xu et al., 2017; Yan et al., 2016).

Both Clouds and aerosols harbour active microbial organisms

ATP concentrations and average RNA:DNA content ratios reflect the potential activity of communities in both clouds and aerosols. Eukaryotes were more abundant (mainly fungi) than bacteria in terms of affiliated sequences number in MTs, which can imply that eukaryotes were

611 more active than bacteria. However, there were many more potentially active bacterial than
612 eukaryotic taxa (greater richness) and with overall higher expression coefficients, indicating
613 that bacteria were the primary domain of active life in clouds and aerosols. Moreover, this tends
614 to indicate that there were probably only a few highly active eukaryotic taxa in the active
615 community as a whole.

616 Potentially active microbial communities were not significantly distinct between clouds
617 and aerosols, with high variabilities between samples. This supported that the composition of
618 active taxa was mainly driven by emission sources, rather than by the specificities of the
619 atmospheric environment. The active fraction clearly differed from the whole community (MG
620 data) (cf **Supplementary Figure 8**) and represented only ~20 % of the total richness,
621 suggesting specific selection processes.

622 Cyanobacteria, Clostridiales and Spirochaetes were among the most active taxa here and
623 were already reported active in Artic area (Šantl-Temkiv et al., 2018). Cyanobacteria and
624 Alpha-Proteobacteria phyla were also previously found potentially active in clouds (Amato et
625 al., 2019). Interestingly, *Pseudomonas* was found significantly overrepresented in MT data,
626 particularly, if not only, in clouds (**Figure 3**) which correlates with observations done in Amato
627 *et al*. (2017) and Šantl-Temkiv *et al.* (2018) with a high activity potential for *Pseudomonas* in
628 clouds and rain but less or none in air. Regarding eukaryotes, significantly differentially active
629 taxa were mostly unicellular algae (Cryptophyta and Bacillariophyta) known to live in aqueous
630 and marine environments. Moreover, the Cryptomonadaceae family was significantly
631 overexpressed in clouds. Photosynthetic microorganisms adapted to aqueous environments
632 (Cyanobacteria and micro-algae) were therefore highly represented in the active part of airborne
633 communities.

634

Microbial cells expressed central metabolic processes in response to atmospheric conditions in clouds and aerosols

Central and energy metabolisms as well as biological processes related to cell cycle, translation, carbon utilization or photosynthesis were overexpressed in both clouds and aerosols. These observations testify at least to a minimal metabolism for the cells to survive and be potentially active, with the degradation of substrates, the synthesis of proteins, the functioning of cellular respiration and potentially of photosynthesis. Interestingly, the pentose phosphate shunt and the glyoxylate pathway were overrepresented, which may be related with a response to oxidative stress (production of NADPH) (Christodoulou et al., 2018; Slekar et al., 1996) and to the use of acetate instead of the more complex sugars in the case of a lack of nutrients, as well as a potential response to oxidative stress (Ensign, 2006; Park et al., 2019).

Moreover, cell outer membrane cellular components were among the most overexpressed in airborne communities, which highlights the strong interaction of the cells with the extracellular phase. The outer member is indeed a key component in the acclimation of microbial cells to their environments (e.g., regulation of nutrient uptake and solute transport, participation in cell division, adhesion, signalling and sensing) and is directly affected by osmolarity change and balance. Metabolism of compounds such as mannitol, glycine and glycerol were as well overrepresented and are known to be compatible solutes which can be stocked in cells and potentially used to balance osmolarity or can also act as a cryoprotective (Ghobakhlou et al., 2015; Goordial et al., 2016; Mykytczuk et al., 2013; Robinson, 2001; Sajjad et al., 2020).

Finally, the fact that H_2O_2 catabolic processes were overexpressed supports previous work (Vaïtilingom et al., 2013; Wirgot et al., 2017), and a potential impact on cloud chemistry. This can now be extended to dry atmosphere (aerosols).

Clouds exhibited more biological processes than aerosols

NMDS analysis and hierarchical clustering supported that clouds and aerosols harboured different potential activities. In addition, a higher number of significantly differentially expressed GOs were observed in clouds compared to aerosols. Several clear distinctions can be made, as the strong overrepresentation of energy metabolism (e.g., ATP synthesis coupled electron transport, carbon utilization) (**Figure 6D**), of cytoplasmic translation, of glucose and carbohydrate metabolisms, and of polysaccharide catabolic processes in clouds. All these processes highlight clouds as the main places of microbial activities in the atmosphere, as indicated as well by RNA:DNA content ratios significantly higher in clouds.

The cells responded differently in the presence or absence of condensed water

Differences in expression between clouds and aerosols were also distinguishable for cellular responses to environmental conditions. In clouds, regulation of intracellular pH, response to UV, response to osmotic stress, to starvation (toward glucose and amino acids), autophagy and pexophagy (i.e., macropexophagy) were overexpressed. It appears logical that regulation of pH and osmotic stress became a concern for cells in an aqueous environment such as cloud droplets rather than in dry air. Interestingly, transmembrane transport was also overrepresented in clouds, supporting solutes transport to regulate osmolarity, and also a a higher activity in the water medium with potential uptake of nutrients and others essential compounds (e.g. carbohydrate transport overrepresented in clouds). UV exposure is a concern for airborne microorganisms in aerosols or clouds, and the fact that UV response was significantly higher in clouds suggests that light exposure was greater in clouds than in aerosols and supports important Mie scattering within it. Responses to starvation is intriguing here in clouds rather than in aerosols but can be considered as a consequence of the resurgence of microbial activity in clouds droplets and a higher demand in nutrients. This goes with the

overexpression of autophagy, which is a response to stressful environment but also a way to recycle components and retrieve glucose, amino acid, fatty acid and others essential compounds. Finally, pexophagy is the main way of regulating peroxisomes, organelles mainly responsible in eukaryotic cells for the detoxification of hydrogen peroxide (Till et al., 2012). The overrepresentation of pexophagy in clouds can therefore be interpreted as a negative regulation of peroxisomes related to lower concentrations of hydrogen peroxide in cloud droplets compared to aerosols. Cells were overexpressing as well hydrogen peroxide catabolic processes, that could have contributed to the lower concentrations in cloud droplets.

In aerosols, responses to oxidative stress, to heat and SOS response were overrepresented compared to clouds. This suggests that airborne microorganisms were more exposed to reactive oxygen species (e.g., superoxide anions, H_2O_2 and hydroxyl radicals), as supported by pexophagy in clouds, but also to higher temperatures. Indeed, temperatures were higher during aerosol collection than during cloud sampling. H_2O_2 concentrations in the atmosphere are known to be positively correlated with solar radiation and temperature as the formation of H_2O_2 is promoted by enhanced photochemical activity (Lee et al., 2000). Also, direct exposition of aerosols to sun light may contribute to heat exposure of cells compared to conditions encountered within clouds (**Figure 15**).

Studies investigating the effect of drying-rewetting cycles on soil microbial communities have demonstrated that these cycles result in increased cellular respiration after soil rewetting. Growth of microorganisms increases at a slower rate and was likely due to the release of nutrients from the wet soil organic matter and dead organisms which did not survive to the osmotic shock (Fraser et al., 2016; Iovieno and Bååth, 2008; Kieft et al., 1987). Furthermore, it has been shown that when undergoing multiple drying-rewetting cycles in soils, cells can be shape to respond more quickly to future drying-rewetting cycles (Leizeaga et al., 2022), thus perhaps more adapted to recolonize or colonize new environments in the presence of water. A

similar phenomenon can be expected for airborne microorganisms incorporating clouds droplets. The respiratory electron transport chain was significantly overexpressed in clouds, and nutrient concentrations in cloud droplets could increase with dissolution of organic and inorganic aerosol particles (Marinoni et al., 2004; Sellegri et al., 2003) and cell lysis (Ye et al., 2010) with osmotic shock and freeze-thaw cycles, allowing access to new nutrients compared to dry aerosols. In addition, it was estimated that clouds undergo 10-11 evaporation-condensation cycles before precipitating (Pruppacher and Jaenicke, 1995). Cells that survive these drastic conditions may therefore already be formed in clouds to be better prepared for future drying-rewetting cycles when they reach the ground with precipitation. This may confer adaptations on viable cells to better colonize new ecosystems.

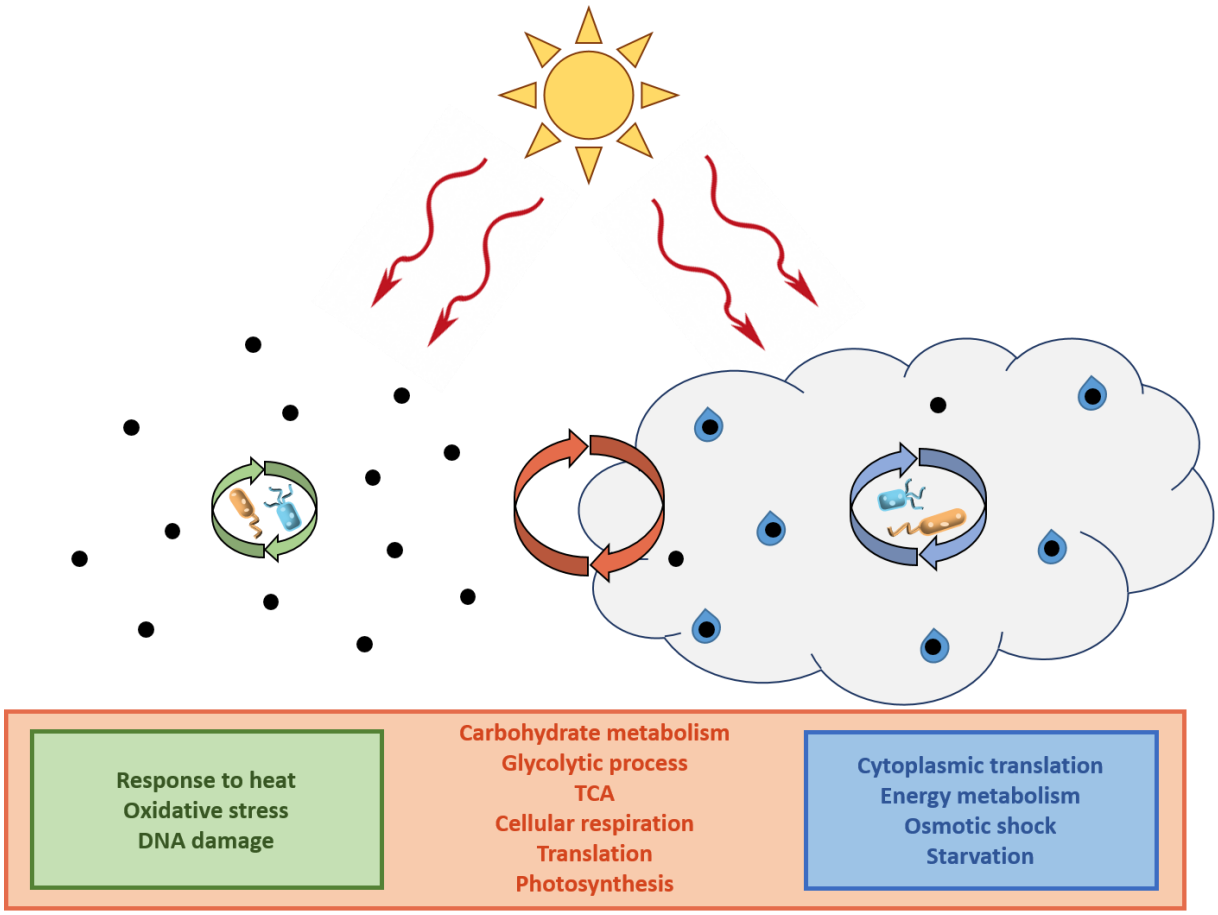


Figure 15: Main metabolic processes expressed by airborne microbial communities in the atmosphere. Green: aerosol specific; blue: cloud specific; orange: common to both atmospheric situations.

Conclusions

Clouds and aerosols microbial communities were for the first time compared in terms of functional profile using metatranscriptomics. They harboured similar communities typical of atmospheric environments where only a fraction of the community was potentially active, testifying of a demanding and selective environment. Central metabolic activities were expressed in both atmospheric situations, placing aerosols as a potential player in atmospheric microbial metabolism alongside clouds. This first point calls into question what has been assumed until now about microbial activity in the atmosphere and its potential impact since atmospheric models currently only consider biologic activity in clouds (Ervens and Amato, 2020; Khaled et al., 2021). Furthermore, environment-specific responses indicated attempts by microorganisms to acclimatize to atmospheric conditions and rewetting upon incorporation into cloud droplets. Finally, and above all, clouds were harbouring much more potential microbial activities than aerosols and a significantly higher energy metabolism, which can testify of the "revival" of microorganisms within clouds. Data support our hypothesis about clouds as microbial habitats in the atmosphere, providing nutrients (substrates), water (cloud droplet) and shelter to microbial communities and therefore allowing airborne microbes to better survive to atmospheric conditions and a resumption of microbial activity. This also brings the question of microorganism's dispersion and ecology as clouds are the source of precipitation.

To summarized here the potential fate of microorganisms in the atmosphere: microbes are aerosolized and mixed over long distances, with only a small proportion remaining viable and active in the dry atmosphere (aerosols) and being exposed to extreme living conditions including high exposure to UV and free radicals. The airborne microorganisms that will reach high altitudes and will be integrated in clouds will have to face osmotic shocks and freeze-thaw cycles but will have in exchange an aqueous microenvironment "reviving" them. Successive drying-rewetting cycles of soils have even been shown to promote faster reactivation of cell

metabolism in the presence of water (Leizeaga et al., 2022). Applying this to the atmosphere, the microorganisms that have survived so far would thus be revived and prepared to be more reactive in water presence, before being redeposited on surface ecosystems with precipitation. Precipitation would thus disseminate subsamples of microbial strains from the clouds (Péguilhan et al., 2021) ready to colonize and impact local ecosystems such as lake (Noirmain et al., 2022), crop and vegetation (Morris et al., 2008).

Further studies should be conducted with larger datasets from different geographical area to have a better view on the potential link between functional profile and biodiversity and emission sources of microorganisms. Moreover, it would be interesting to have a temporal monitoring of the transition from dry air to cloudy state to better estimate the impact of these two atmospheric situations on the microbial communities in close temporal conditions. At last, it would be an advance to have absolute quantifications of several microbial activities of interest in clouds and aerosols (e.g., hydrogen peroxide catabolism), by quantitative PCR for example. This would allow a better view of the impact of microorganisms on atmospheric chemical compounds and potentially integrate these data into new atmospheric models. This would also give clues to the potential impact of those airborne microorganisms (sometimes originating from continental-scale transport) on the metabolism of local ecosystems.

Acknowledgments

References

Ahern, H.E., Walsh, K.A., Hill, T.C.J., and Moffett, B.F. (2007). Fluorescent pseudomonads isolated from Hebridean cloud and rain water produce biosurfactants but do not cause ice nucleation. Biogeosciences *4*, 115–124.

Aho, K., Weber, C.F., Christner, B.C., Vinatzer, B.A., Morris, C.E., Joyce, R., Failor, K., Werth, J.T., Bayless-Edwards, A.L.H., and Schmale III, D.G. (2019). Spatiotemporal patterns of microbial composition and diversity in precipitation. Ecol. Monogr. *0*, 1–26.

Almaguer, M., Aira, M.J., Rodríguez-Rajo, F.J., and Rojas, T.I. (2014). Temporal dynamics of airborne fungi in Havana (Cuba) during dry and rainy seasons: influence of meteorological parameters. Int. J. Biometeorol. *58*, 1459–1470.

Amato, P. (2013). Energy Metabolism at Low-temperature and Frozen Conditions in Cold-adapted Microorganisms. Cold-Adapted Microorg. 71–96.

Amato, P., Ménager, M., Sancelme, M., Laj, P., Mailhot, G., and Delort, A.M. (2005). Microbial population in cloud water at the Puy de Dôme: Implications for the chemistry of clouds. Atmos. Environ. *39*, 4143–4153.

Amato, P., Parazols, M., Sancelme, M., Laj, P., Mailhot, G., and Delort, A.M. (2007). Microorganisms isolated from the water phase of tropospheric clouds at the Puy de Dôme: Major groups and growth abilities at low temperatures. In FEMS Microbiology Ecology, pp. 242–254.

Amato, P., Joly, M., Besaury, L., Oudart, A., Taib, N., Moné, A.I., Deguillaume, L., Delort, A.M., and Debroas, D. (2017). Active microorganisms thrive among extremely diverse communities in cloud water. PLoS One *12*, 1–22.

Amato, P., Besaury, L., Joly, M., Penaud, B., Deguillaume, L., and Delort, A.M. (2019). Metatranscriptomic exploration of microbial functioning in clouds. Sci. Rep. *9*.

American Meteorological Society - AMS (2012). American Meteorological Society - Glossary of Meteorology.

André, F., Jonard, M., and Ponette, Q. (2007). Influence of meteorological factors and polluting environment on rain chemistry and wet deposition in a rural area near Chimay, Belgium. Atmos. Environ. *41*, 1426–1439.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Anesio, A.M., Hodson, A.J., Fritz, A., Psenner, R., and Sattler, B. (2009). High microbial activity on glaciers: Importance to the global carbon cycle. Glob. Chang. Biol. *15*, 955–960.

Arts, I.S., Gennaris, A., and Collet, J.F. (2015). Reducing systems protecting the bacterial cell envelope from oxidative damage. FEBS Lett. *589*, 1559–1568.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. *25*, 25.

Asmi, E., Freney, E., Hervo, M., Picard, D., Rose, C., Colomb, A., and Sellegri, K. (2012). Aerosol cloud activation in summer and winter at puy-de-Dôme high altitude site in France. Atmos. Chem. Phys. *12*, 11589–11607.

Aydogan, E.L., Moser, G., Müller, C., Kämpfer, P., and Glaeser, S.P. (2018). Long-term warming shifts the composition of bacterial communities in the phyllosphere of Galium album in a permanent grassland field-experiment. Front. Microbiol. *9*, 144.

Baldrian, P., Kolaiřík, M., Štursová, M., Kopecký, J., Valášková, V., Větrovský, T., Žifčáková, L.,

Šnajdr, J., Rídl, J., Vlček, Č., et al. (2012). Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. ISME J. *6*, 248–258.

Baray, J.L., Bah, A., Cacault, P., Sellegri, K., Pichon, J.M., Deguillaume, L., Montoux, N., Noel, V., Seze, G., Gabarrot, F., et al. (2019). Cloud occurrence frequency at puy de dome (France) deduced from an automatic camera image analysis: Method, validation, and comparisons with larger scale parameters. Atmosphere (Basel). *10*, 808.

Baray, J.L., Deguillaume, L., Colomb, A., Sellegri, K., Freney, E., Rose, C., Baelen, J. Van, Pichon, J.M., Picard, D., Fréville, P., et al. (2020). Cézeaux-Aulnat-Opme-Puy de Dôme: A multi-site for the long-term survey of the tropospheric composition and climate change. Atmos. Meas. Tech. *13*, 3413–3445.

Barberán, A., Ladau, J., Leff, J.W., Pollard, K.S., Menninger, H.L., Dunn, R.R., and Fierer, N. (2015). Continental-scale distributions of dust-associated bacteria and fungi. PNAS *112*, 5756–5761.

Barnard, R.L., Osborne, C.A., and Firestone, M.K. (2013). Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. ISME J. *7*, 2229–2241.

Bauer, H., Kasper-Giebl, A., Löflund, M., Giebl, H., Hitzenberger, R., Zibuschka, F., and Puxbaum, H. (2002). The contribution of bacteria and fungal spores to the organic carbon content of cloud water, precipitation and aerosols. Atmos. Res. *64*, 109–119.

Bauer, H., Giebl, H., Hitzenberger, R., Kasper-Giebl, A., Reischl, G., Zibuschka, F., and Puxbaum, H. (2003). Airborne bacteria as cloud condensation nuclei. J. Geophys. Res. Atmos. *108*, 4658.

Behzad, H., Gojobori, T., and Mineta, K. (2015). Challenges and opportunities of airborne metagenomics. Genome Biol. Evol. *7*, 1216–1226.

Bertolini, V., Gandolfi, I., Ambrosini, R., Bestetti, G., Innocente, E., Rampazzo, G., and Franzetti, A. (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.

Bertrand, G., Celle-Jeanton, H., Laj, P., Rangognio, J., and Chazot, G. (2008). Rainfall chemistry: Long range transport versus below cloud scavenging. A two-year study at an inland station (Opme, France). J. Atmos. Chem. *60*, 253–271.

Blanco-Alegre, C., Castro, A., Calvo, A.I., Oduber, F., Alonso-Blanco, E., Fernández-González, D., Valencia-Barrera, R.M., Vega-Maray, A.M., and Fraile, R. (2018). Below-cloud scavenging of fine and coarse aerosol particles by rain: The role of raindrop size. Q. J. R. Meteorol. Soc. *144*, 2715–2726.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bourcier, L., Masson, O., Laj, P., Chausse, P., Pichon, J.M., Paulat, P., Bertrand, G., and Sellegri, K. (2012). A new method for assessing the aerosol to rain chemical composition relationships. Atmos. Res. *118*, 295–303.

Bowers, R.M., McLetchie, S., Knight, R., and 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.

Bowers, R.M., Clements, N., Emerson, J.B., Wiedinmyer, C., Hannigan, M.P., and Fierer, N. (2013). Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. Environ. Sci. Technol. *47*, 12097–12106.

Bryan, N.C., Christner, B.C., Guzik, T.G., Granger, D.J., and Stewart, M.F. (2019). Abundance and survival of microbial aerosols in the troposphere and stratosphere. ISME J. *13*, 2789–2799.

Buchfink, B., Xie, C., and Huson, D.H. (2014). Fast and sensitive protein alignment using DIAMOND. Nat. Methods 2014 121 *12*, 59–60.

Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F., Rauf, P., Huettel, B., Reinhardt, R., Schmelzer, E., et al. (2012). Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. Nat. 2012 4887409 *488*, 91–95.

Burrows, S.M., Butler, T., Jöckel, P., Tost, H., Kerkweg, A., Pöschl, U., and Lawrence, M.G. (2009a). Bacteria in the global atmosphere-Part 2: Modeling of emissions and transport between different ecosystems.

Burrows, S.M., Elbert, W., Lawrence, M.G., and Pöschl, U. (2009b). Bacteria in the global atmosphere – Part 1: Review and synthesis of literature data for different ecosystems. Atmos. Chem. Phys. *9*, 9263–9280.

Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A.D. (2019). rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. Gigascience *8*, 1–13.

Button, D.K., and Robertson, B.R. (2001). Determination of DNA Content of Aquatic Bacteria by Flow Cytometry. Appl. Environ. Microbiol. *67*, 1636–1645.

Cáliz, J., Triadó-Margarit, X., Camarero, L., and Casamayor, E.O. (2018). A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. Proc. Natl. Acad. Sci. *115*, 12229–12234.

Camacho-Sanchez, M., Burraco, P., Gomez-Mestre, I., and Leonard, J.A. (2013). Preservation of RNA and DNA from mammal samples under field conditions. Mol. Ecol. Resour. *13*, 663–673.

Campbell, B.J., Yu, L., Heidelberg, J.F., and Kirchman, D.L. (2011). Activity of abundant and rare bacteria in a coastal ocean. Proc. Natl. Acad. Sci. U. S. A. *108*, 12776–12781.

Carbon, S., Douglass, E., Good, B.M., Unni, D.R., Harris, N.L., Mungall, C.J., Basu, S., Chisholm, R.L., Dodson, R.J., Hartline, E., et al. (2021). The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. *49*, D325–D334.

Carvalhais, L.C., Dennis, P.G., Tyson, G.W., and Schenk, P.M. (2012). Application of metatranscriptomics to soil environments. J. Microbiol. Methods *91*, 246–251.

Cha, S., Lee, D., Jang, J.H., Lim, S., Yang, D., and Seo, T. (2016). Alterations in the airborne bacterial community during Asian dust events occurring between February and March 2015 in South Korea. Nat. Publ. Gr. 1–9.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890.

Christodoulou, D., Link, H., Fuhrer, T., Kochanowski, K., Gerosa, L., and Sauer, U. (2018). Reserve Flux Capacity in the Pentose Phosphate Pathway Enables Escherichia coli's Rapid Response to Oxidative Stress. Cell Syst. *6*, 569-578.e7.

Cohen, N.R., Ross, C.A., Jain, S., Shapiro, R.S., Gutierrez, A., Belenky, P., Li, H., and Collins, J.J. (2016). A role for the bacterial GATC methylome in antibiotic stress survival. Nat. Genet. *48*, 581–586.

Collett, J., Oberholzer, B., and Staehelin, J. (1993). Cloud chemistry at Mt Rigi, Switzerland: Dependence on drop size and relationship to precipitation chemistry. Atmos. Environ. Part A, Gen. Top. *27*, 33–42.

Deguillaume, L., Charbouillot, T., Joly, M., Vaïtilingom, M., Parazols, M., Marinoni, A., Amato, P., Delort, A.M., Vinatier, V., Flossmann, A., et al. (2014). Classification of clouds sampled at the puy de Dôme (France) based on 10 yr of monitoring of their physicochemical properties. Atmos. Chem. Phys. *14*, 1485–1506.

DeLeon-Rodriguez, N., Lathem, T.L., Rodriguez-R, L.M., Barazesh, J.M., Anderson, B.E., Beyersdorf, A.J., Ziemba, L.D., Bergin, M., Nenes, A., and Konstantinidis, K.T. (2013). Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric

903 implications. Proc. Natl. Acad. Sci. *110*, 2575–2580.

904 Delort, A.M., Vaïtilingom, M., Amato, P., Sancelme, M., Parazols, M., Mailhot, G., Laj, P., and
905 Deguillaume, L. (2010). A short overview of the microbial population in clouds: Potential roles in
906 atmospheric chemistry and nucleation processes. Atmos. Res. *98*, 249–260.

907 Després, V.R., Alex Huffman, J., Burrows, S.M., Hoose, C., Safatov, A.S., Buryak, G., Fröhlich-
908 Nowoisky, J., Elbert, W., Andreae, M.O., Pöschl, U., et al. (2012). Primary biological aerosol particles
909 in the atmosphere: A review. Tellus, Ser. B Chem. Phys. Meteorol. *64*.

910 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and
911 Gingeras, T.R. (2013). Sequence analysis STAR: ultrafast universal RNA-seq aligner. *29*, 15–21.

912 Drautz-Moses, D.I., Luhung, I., Gusareva, E.S., Kee, C., Gaultier, N.E., Premkrishnan, B.N. V., Lee,
913 C.F., Leong, S.T., Park, C., Yap, Z.H., et al. (2022). Vertical stratification of the air microbiome in the
914 lower troposphere. Proc. Natl. Acad. Sci. *119*, e2117293119.

915 Dueker, M.E., Weathers, K.C., O'Mullan, G.D., Juhl, A.R., and Uriarte, M. (2011). Environmental
916 controls on coastal coarse aerosols: Implications for microbial content and deposition in the near-shore
917 environment. Environ. Sci. Technol. *45*, 3386–3392.

918 Dueker, M.E., O'Mullan, G.D., Weathers, K.C., Juhl, A.R., and Uriarte, M. (2012). Coupling of fog and
919 marine microbial content in the near-shore coastal environment. Biogeosciences *9*, 803–813.

920 Edwards, K.J., Bond, P.L., Gihring, T.M., and Banfield, J.F. (2000). An Archaeal Iron-Oxidizing
921 Extreme Acidophile Important in Acid Mine Drainage. Science (80-.). *287*, 1796–1799.

922 Elbert, W., Taylor, P.E., Andreae, M.O., and Pöschl, U. (2007). Contribution of fungi to primary
923 biogenic aerosols in the atmosphere: Wet and dry discharged spores, carbohydrates, and inorganic ions.
924 Atmos. Chem. Phys. *7*, 4569–4588.

925 Els, N., Baumann-Stanzer, K., Larose, C., Vogel, T.M., and Sattler, B. (2019). Beyond the planetary
926 boundary layer: Bacterial and fungal vertical biogeography at Mount Sonnblick, Austria. Geo Geogr.
927 Environ. *6*.

928 Ensign, S.A. (2006). Revisiting the glyoxylate cycle: Alternate pathways for microbial acetate
929 assimilation. Mol. Microbiol. *61*, 274–276.

930 Ervens, B., and Amato, P. (2020). The global impact of bacterial processes on carbon mass. Atmos.
931 Chem. Phys. Discuss. *20*, 1–25.

932 Evans, S.E., Dueker, M.E., Logan, J.R., and Weathers, K.C. (2019). The biology of fog: results from
933 coastal Maine and Namib Desert reveal common drivers of fog microbial composition. Sci. Total
934 Environ. *647*, 1547–1556.

935 Fankhauser, A.M., Antonio, D.D., Krell, A., Alston, S.J., Banta, S., and McNeill, V.F. (2019).
936 Constraining the Impact of Bacteria on the Aqueous Atmospheric Chemistry of Small Organic
937 Compounds. ACS Earth Sp. Chem. *3*, 1485–1491.

938 Fong, N.J.C., Burgess, M.L., Barrow, K.D., and Glenn, D.R. (2001). Carotenoid accumulation in the
939 psychrotrophic bacterium Arthrobacter agilis in response to thermal and salt stress. Appl. Microbiol.
940 Biotechnol. *56*, 750–756.

941 Forouzan, E., Shariati, P., Mousavi Maleki, M.S., Karkhane, A.A., and Yakhchali, B. (2018). Practical
942 evaluation of 11 de novo assemblers in metagenome assembly. J. Microbiol. Methods *151*, 99–105.

943 Fraser, F.C., Corstanje, R., Deeks, L.K., Harris, J.A., Pawlett, M., Todman, L.C., Whitmore, A.P., and
944 Ritz, K. (2016). On the origin of carbon dioxide released from rewetted soils. Soil Biol. Biochem. *101*,
945 1–5.

946 Freitas, T.A.K., Li, P.E., Scholz, M.B., and Chain, P.S.G. (2015). Accurate read-based metagenome

947 characterization using a hierarchical suite of unique signatures. Nucleic Acids Res. *43*, e69–e69.

948 Fröhlich-Nowoisky, J., Pickersgill, D.A., Després, V.R., and Pöschl, U. (2009). High diversity of fungi
949 in air particulate matter. Proc. Natl. Acad. Sci. *106*, 12814.

950 Fröhlich-Nowoisky, J., Burrows, S.M., Xie, Z., Engling, G., Solomon, P.A., Fraser, M.P., Mayol-
951 Bracero, O.L., Artaxo, P., Begerow, D., Conrad, R., et al. (2012). Biogeography in the air: Fungal
952 diversity over land and oceans. Biogeosciences *9*, 1125–1136.

953 Fröhlich-Nowoisky, J., Kampf, C.J., Weber, B., Huffman, J.A., Pöhlker, C., Andreae, M.O., Lang-Yona,
954 N., Burrows, S.M., Gunthe, S.S., Elbert, W., et al. (2016). Bioaerosols in the Earth system: Climate,
955 health, and ecosystem interactions. Atmos. Res. *182*, 346–376.

956 Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-
957 generation sequencing data. Bioinformatics *28*, 3150–3152.

958 Fuzzi, S., Mandrioli, P., and Perfetto, A. (1996). Fog droplets - An atmospheric source of secondary
959 biological aerosol particles. Atmos. Environ. *31*, 287–290.

960 Fuzzi, S., Andreae, M.O., Huebert, B.J., Kulmala, M., Bond, T.C., Boy, M., Doherty, S.J., Guenther,
961 A., Kanakidou, M., Kawamura, K., et al. (2006). Critical assessment of the current state of scientific
962 knowledge, terminology, and research needs concerning the role of organic aerosols in the atmosphere,
963 climate, and global change. Atmos. Chem. Phys. *6*, 2017–2038.

964 Gandolfi, I., Bertolini, V., Ambrosini, R., Bestetti, G., and Franzetti, A. (2013). Unravelling the bacterial
965 diversity in the atmosphere. Appl. Microbiol. Biotechnol. *97*, 4727–4736.

966 Gandolfi, I., Bertolini, V., Bestetti, G., Ambrosini, R., Innocente, E., Rampazzo, G., Papacchini, M.,
967 and Franzetti, A. (2015). Spatio-temporal variability of airborne bacterial communities and their
968 correlation with particulate matter chemical composition across two urban areas. Appl. Microbiol.
969 Biotechnol. *99*, 4867–4877.

970 Garratt, J.R. (1994). The atmospheric boundary layer. Earth-Science Rev. *37*, 89–134.

971 Ge, Z., Wexler, A.S., and Johnston, M. V (1998). Deliquescence behavior of multicomponent aerosols.
972 J. Phys. Chem. A *102*, 173–180.

973 Ghobakhlou, A.F., Johnston, A., Harris, L., Antoun, H., and Laberge, S. (2015). Microarray
974 transcriptional profiling of Arctic Mesorhizobium strain N33 at low temperature provides insights into
975 cold adaption strategies. BMC Genomics *16*.

976 Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017). Microbiome datasets are
977 compositional: And this is not optional. Front. Microbiol. *8*, 1–6.

978 de Goffau, M.C., Lager, S., Salter, S.J., Wagner, J., Kronbichler, A., Charnock-Jones, D.S., Peacock,
979 S.J., Smith, G.C.S., and Parkhill, J. (2018). Recognizing the reagent microbiome. Nat. Microbiol. *3*,
980 851–853.

981 Gong, F., and Miller, K.M. (2019). Histone methylation and the DNA damage response. Mutat. Res.
982 Mutat. Res. *780*, 37–47.

983 Goordial, J., Raymond-Bouchard, I., Zolotarov, Y., De Bethencourt, L., Ronholm, J., Shapiro, N.,
984 Woyke, T., Stromvik, M., Greer, C.W., Bakermans, C., et al. (2016). Cold adaptive traits revealed by
985 comparative genomic analysis of the eurypsychrophile Rhodococcus sp. JG3 isolated from high
986 elevation McMurdo Dry Valley permafrost, Antarctica. FEMS Microbiol. Ecol. *92*, 154.

987 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L.,
988 Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data
989 without a reference genome. Nat. Biotechnol. 2011 297 *29*, 644–652.

990 Grafstrom, R.H., Hamilton, D.L., and Yuan, R. (1984). DNA Methylation: DNA Replication and Repair.

111–126.

Graham, K.E., Prussin, A.J., Marr, L.C., Sassoubre, L.M., and Boehm, A.B. (2018). Microbial community structure of sea spray aerosols at three California beaches. FEMS Microbiol. Ecol. *94*, 5.

Green, R.L., and Warren, G.J. (1985). Physical and functional repetition in a bacterial ice nucleation gene. Nature *317*, 645–648.

Griffin, D.W. (2007). Atmospheric movement of microorganisms in clouds of desert dust and implications for human health. Clin. Microbiol. Rev. *20*, 459–477.

Hamilton, W.D., and Lenton, T.M. (1998). Spora and gaia : how microbes fly with their clouds. Ethol. Ecol. Evol. *10*, 1–16.

Heald, C.L., and Spracklen, D. V. (2009). Atmospheric budget of primary biological aerosol particles from fungal spores. Geophys. Res. Lett. *36*.

Hervàs, A., Camarero, L., Reche, I., and Casamayor, E.O. (2009). Viability and potential for immigration of airborne bacteria from Africa that reach high mountain lakes in Europe. Environ. Microbiol. *11*, 1612–1623.

Hoffmann, L., Günther, G., Li, D., Stein, O., Wu, X., Griessbach, S., Heng, Y., Konopka, P., Müller, R., Vogel, B., et al. (2019). From ERA-Interim to ERA5: The considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. Atmos. Chem. Phys. *19*, 3097–3214.

Hou, P., Wu, S., and McCarty, J. (2017). Sensitivity of atmospheric aerosol scavenging to precipitation intensity and frequency in the context of global climate change. Atmos. Chem. Phys. Discuss. 1–17.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. *47*, D309–D314.

Huffman, J.A., Prenni, A.J., Demott, P.J., Pöhlker, C., Mason, R.H., Robinson, N.H., Fröhlich-Nowoisky, J., Tobo, Y., Després, V.R., Garcia, E., et al. (2013). High concentrations of biological aerosol particles and ice nuclei during and after rain. Atmos. Chem. Phys. *13*, 6151–6164.

Hunt, D.E., Lin, Y., Church, M.J., Karl, D.M., Tringe, S.G., Izzo, L.K., and Johnson, Z.I. (2013). Relationship between abundance and specific activity of bacterioplankton in open ocean surface waters. Appl. Environ. Microbiol. *79*, 177–184.

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics *11*, 1–11.

Imshenetsky, A.A., Lysenko, S. V., and Kazakov, G.A. (1978). Upper boundary of the biosphere. Appl. Environ. Microbiol. *35*, 1–5.

Iovieno, P., and Bååth, E. (2008). Effect of drying and rewetting on bacterial growth rates in soil. FEMS Microbiol. Ecol. *65*, 400–407.

Jaing, C., Thissen, J., Morrison, M., Dillon, M.B., Waters, S.M., Graham, G.T., Be, N.A., Nicoll, P., Verma, S., Caro, T., et al. (2020). Sierra Nevada sweep: metagenomic measurements of bioaerosols vertically distributed across the troposphere. Sci. Rep. *10*, 12399.

Joly, M., Amato, P., Sancelme, M., Vinatier, V., Abrantes, M., Deguillaume, L., and Delort, A.M. (2015). Survival of microbial isolates from clouds toward simulated atmospheric stress factors. Atmos. Environ. *117*, 92–98.

Jones, B.E., Grant, W.D., Duckworth, A.W., Owenson, G.G., Horikoshi, K., Jones, B.E., Grant, W.D., Duckworth, · A W, and Owenson, · G G (1998). Microbial diversity of soda lakes. Extrem. 1998 23 *2*,

191–200.

Joung, Y.S., Ge, Z., and Buie, C.R. (2017). Bioaerosol generation by raindrops on soil. Nat. Commun. *8*, 1–10.

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *28*, 27–30.

Karin, E.L., Mirdita, M., and Söding, J. (2020). MetaEuk-sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. Microbiome *8*, 1–15.

Khaled, A., Zhang, M., Amato, P., Delort, A.M., and Ervens, B. (2021). Biodegradation by bacteria in clouds: An underestimated sink for some organics in the atmospheric multiphase system. Atmos. Chem. Phys. *21*, 3123–3141.

Kieft, T.L., and Ahmadjian, V. (1989). Biological ice nucleation activity in lichen mycobionts and photobionts. Lichenol. *21*, 355–362.

Kieft, T.L., Soroker, E., and Firestone, M.K. (1987). Microbial biomass response to a rapid increase in water potential when dry soil is wetted. Soil Biol. Biochem. *19*, 119–126.

Kim, D., Song, L., Breitwieser, F.P., and Salzberg, S.L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. *26*, 1721–1729.

Klein, A.M., Bohannan, B.J.M., Jaffe, D.A., Levin, D.A., and Green, J.L. (2016). Molecular evidence for metabolically active bacteria in the atmosphere. Front. Microbiol. *7*.

Klingenberg, H., and Meinicke, P. (2017). How to normalize metatranscriptomic count data for differential expression analysis. PeerJ *2017*.

Kobziar, L.N., Vuono, D., Moore, R., Christner, B.C., Dean, T., Betancourt, D., Watts, A.C., Aurell, J., and Gullett, B. (2022). Wildland fire smoke alters the composition, diversity, and potential atmospheric function of microbial life in the aerobiome. ISME Commun. *2*, 1–9.

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics *28*, 3211–3217.

Krumins, V., Mainelis, G., Kerkhof, L.J., and Fennell, D.E. (2014). Substrate-Dependent rRNA Production in an Airborne Bacterium. Environ. Sci. Technol. Lett. *1*, 376–381.

Ladino, L., Stetzer, O., Hattendorf, B., Günther, D., Croft, B., and Lohmann, U. (2011). Experimental Study of Collection Efficiencies between Submicron Aerosols and Cloud Droplets. J. Atmos. Sci. *68*, 1853–1864.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lavrinienko, A., Jernfors, T., Koskimäki, J.J., Pirttilä, A.M., and Watts, P.C. (2021). Does Intraspecific Variation in rDNA Copy Number Affect Analysis of Microbial Communities? Trends Microbiol. *29*, 19–27.

Lazaridis, M. (2019). Bacteria as Cloud Condensation Nuclei (CCN) in the Atmosphere. Atmosphere (Basel). *10*, 786.

Lee, M., Heikes, B.G., and O'Sullivan, D.W. (2000). Hydrogen peroxide and organic hydroperoxide in the troposphere: A review. Atmos. Environ. *34*, 3475–3494.

Leizeaga, A., Meisner, A., Rousk, J., and Bååth, E. (2022). Repeated drying and rewetting cycles accelerate bacterial growth recovery after rewetting. Biol. Fertil. Soils *58*, 365–374.

Leung, H.C.M., Yiu, S.M., Parkinson, J., and Chin, F.Y.L. (2013). IDBA-MT: De Novo Assembler for Metatranscriptomic Data Generated from Next-Generation Sequencing Technology.

Https://Home.Liebertpub.Com/Cmb *20*, 540–550.

Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2014). IDBA-MTP: A Hybrid MetaTranscriptomic Assembler Based on Protein Information. Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) *8394 LNBI*, 160–172.

Li, H., and Durbin, R. (2009). Fast and accurate short-read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659.

Li, D., Liu, C.M., Luo, R., Sadakane, K., and Lam, T.W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics *31*, 1674–1676.

Li, H., Zhou, X.Y., Yang, X.R., Zhu, Y.G., Hong, Y.W., and Su, J.Q. (2019). Spatial and seasonal variation of the airborne microbiome in a rapidly developing city of China. Sci. Total Environ. *665*, 61–68.

Li, X., Chen, H., and Yao, M. (2020). Microbial emission levels and diversities from different land use types. Environ. Int. *143*, 105988.

Lighthart, B. (1997). The ecology of bacteria in the alfresco atmosphere. FEMS Microbiol. Ecol. *23*, 263–274.

Lighthart, B., and Shaffer, B.T. (1995). Viable bacterial aerosol particle size distributions in the midsummer atmosphere at an isolated location in the high desert chaparral. Aerobiologia (Bologna). *11*, 19–25.

Lighthart, B., and Shaffer, B.T. (1997). Increased airborne bacterial survival as a function of particle content and size. Aerosol Sci. Technol. *27*, 439–446.

Lindemann, J., and Upper, C.D. (1985). Aerial Dispersal of Epiphytic Bacteria. Appl. Environ. Microbiol. *50*, 1229–1232.

Lindemann, J., Constantinidou, H.A., Barchet, W.R., and Upper, C.D. (1982). Plants as sources of airborne bacteria, including ice nucleation-active bacteria. Appl. Environ. Microbiol. *44*, 1059–1063.

Lindow, S.E., and Brandl, M.T. (2003). Microbiology of the phyllosphere. Appl. Environ. Microbiol. *69*, 1875–1883.

Lo, C.C., and Chain, P.S.G. (2014). Rapid evaluation and quality control of next generation sequencing data with FaQCs. BMC Bioinformatics *15*, 1–8.

Lofgren, L.A., Uehling, J.K., Branco, S., Bruns, T.D., Martin, F., and Kennedy, P.G. (2019). Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. Mol. Ecol. *28*, 721–730.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 1–21.

Maki, T., Hara, K., Iwata, A., Lee, K.C., Kawai, K., Kai, K., Kobayashi, F., Pointing, S.B., Archer, S., Hasegawa, H., et al. (2017). Variations in airborne bacterial communities at high altitudes over the Noto Peninsula (Japan) in response to Asian dust events. Atmos. Chem. Phys. *17*, 11877–11897.

Manirajan, B.A., Maisinger, C., Ratering, S., Rusch, V., Schwiertz, A., Cardinale, M., and Schnell, S. (2018). Diversity, specificity, co-occurrence and hub taxa of the bacterial-fungal pollen microbiome.

FEMS Microbiol. Ecol. *94*, 1–11.

Marinoni, A., Laj, P., Sellegri, K., and Mailhot, G. (2004). Cloud chemistry at the Puy de Dôme: variability and relationships with environmental factors.

Maron, P.A., Mougel, C., David, D.P., Carvalho, E., Bizet, K., Marck, G., Cubito, N., Lemanceau, P., and Ranjard, L. (2006). Temporal variability of airborne bacterial community structure in an urban area. Atmos. Environ. *40*, 8074–8080.

Matthias-Maser, S., Gruber, S., and Jaenicke, R. (2000). The size distribution of primary biological aerosol particles in cloud water on the mountain Kleiner Feldberg/Taunus (FRG). Atmos. Res. *54*, 1–13.

Medinger, R., Nolte, V., Pandey, R.V., Jost, S., Ottenwälder, B., Schlötterer, C., and Boenigk, J. (2010). Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Mol. Ecol. *19*, 32–40.

Menke, S., Gillingham, M.A.F., Wilhelm, K., and Sommer, S. (2017). Home-made cost effective preservation buffer is a better alternative to commercial preservation methods for microbiome research. Front. Microbiol. *8*.

Michaud, J.M., Thompson, L.R., Kaul, D., Espinoza, J.L., Richter, R.A., Xu, Z.Z., Lee, C., Pham, K.M., Beall, C.M., Malfatti, F., et al. (2018). Taxon-specific aerosolization of bacteria and viruses in an experimental ocean-atmosphere mesocosm. Nat. Commun. *9*.

Mikhailov, E.F., Pöhlker, M.L., Reinmuth-Selzle, K., Vlasenko, S.S., Krüger, O.O., Fröhlich-Nowoisky, J., Pöhlker, C., Ivanova, O.A., Kiselev, A.A., Kremper, L.A., et al. (2021). Water uptake of subpollen aerosol particles: Hygroscopic growth, cloud condensation nuclei activation, and liquid-liquid phase separation. Atmos. Chem. Phys. *21*, 6999–7022.

Miles, N.L., Verlinde, J., and Clothiaux, E.E. (2002). Cloud Droplet Size Distributions in Low-Level Stratiform Clouds. J. Atmos. Sci. *57*, 295–311.

Mircea, M., Stefan, S., and Fuzzi, S. (2000). Precipitation scavenging coefficient: influence of measured aerosol and raindrop size distributions. Atmos. Environ. *34*, 5169–5174.

Möhler, O., DeMott, P.J., Vali, G., and Levin, Z. (2007). Microbiology and atmospheric processes: The role of biological particles in cloud physics. Biogeosciences *4*, 1059–1071.

Moore, D., Robson, G.D., and Trinci, A.P.J. (2011). 21st Century Guidebook to Fungi (Cambridge University Press).

Moore, R.A., Hanlon, R., Powers, C., Schmale, D.G., and Christner, B.C. (2020a). Scavenging of Sub-Micron to Micron-Sized Microbial Aerosols during Simulated Rainfall. Atmos. Chem. Phys 1–13.

Moore, R.A., Bomar, C., Kobziar, L.N., and Christner, B.C. (2020b). Wildland fire as an atmospheric source of viable microbial aerosols and biological ice nucleating particles. ISME J.

Morita, R.Y. (1975). Psychrophilic Bacteria. Am. Soc. Microbiol. *39*, 144–167.

Morris, C.E., Georgakopoulos, D.G., and Sands, D.C. (2004). Ice nucleation active bacteria and their potential role in precipitation. J. Phys. IV *121*, 87–103.

Morris, C.E., Sands, D.C., Vinatzer, B.A., Glaux, C., Guilbaud, C., Buffière, A., Yan, S., Dominguez, H., and Thompson, B.M. (2008). The life history of the plant pathogen Pseudomonas syringae is linked to the water cycle. ISME J. *2*, 321–334.

Mueller, D.R., Vincent, W.F., Bonilla, S., and Laurion, I. (2005). Extremotrophs, extremophiles and broadband pigmentation strategies in a high arctic ice shelf ecosystem. FEMS Microbiol. Ecol. *53*, 73–87.

Mykytczuk, N.C., Foote, S.J., Omelon, C.R., Southam, G., Greer, C.W., and Whyte, L.G. (2013). Bacterial growth at -15 °C; molecular insights from the permafrost bacterium Planococcus halocryophilus Or1. ISME J. *7*, 1211–1226.

Noguchi, H., Taniguchi, T., and Itoh, T. (2008). MetaGeneAnnotator: Detecting Species-Specific Patterns of Ribosomal Binding Site for Precise Gene Prediction in Anonymous Prokaryotic and Phage Genomes. DNA Res. *15*, 387–396.

Noirmain, F., Baray, J., Tridon, F., Cacault, P., Billard, H., Voyard, G., Baelen, J. Van, and Latour, D. (2022). Interdisciplinary strategy to survey phytoplankton dynamics of a eutrophic lake under rain forcing: description of the instrumental set-up and first results. Biogeosciences.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome Res. *27*, 824–834.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745.

Park, C., Shin, B., and Park, W. (2019). Alternative fate of glyoxylate during acetate and hexadecane metabolism in Acinetobacter oleivorans DR1. Sci. Rep. *9*, 1–12.

Pascault, N., Loux, V., Derozier, S., Martin, V., Debroas, D., Maloufi, S., Humbert, J.F., and Leloup, J. (2015). Technical challenges in metatranscriptomic studies applied to the bacterial communities of freshwater ecosystems. Genetica *143*, 157–167.

Pasteur, L., Chamberland, C., and Joubert, J. (1878). Théorie des germes et ses applications à la médecine et à la chirurgie. *7*, 107–115.

Péguilhan, R., Besaury, L., Rossi, F., Enault, F., Baray, J., Deguillaume, L., and Amato, P. (2021). Rainfalls sprinkle cloud bacterial diversity while scavenging biomass. FEMS Microbiol. Ecol. 1–15.

Pelizzola, M., and Ecker, J.R. (2011). The DNA methylome. FEBS Lett. *585*, 1994–2000.

Peng, Y., Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics *28*, 1420–1428.

Peng, Y., Leung, H.C.M., Yiu, S.M., Lv, M.J., Zhu, X.G., and Chin, F.Y.L. (2013). IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. Bioinformatics *29*, i326–i334.

Petrenchuk, O.P., and Drozdova, V.M. (1966). On the chemical composition of cloud water. Tellus *18*, 280–286.

Petters, M.D., and Kreidenweis, S.M. (2007). A single parameter representation of hygroscopic growth and cloud condensation nucleus activity. Atmos. Chem. Phys. *7*, 1961–1971.

Prass, M., Andreae, M.O., Araùjo, A.C. De, Artaxo, P., and Ditas, F. (2021). Bioaerosols in the Amazon rain forest: Temporal variations and vertical profiles of Eukarya, Bacteria and Archaea. Biogeosciences Discuss. 1–23.

Pruppacher, H.R., and Jaenicke, R. (1995). The processing of water vapor and aerosols by atmospheric clouds, a global estimate. Atmos. Res. *38*, 283–295.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. *41*, 590–596.

Radke, L.F., Hobbs, P. V., and Eltgroth, M.W. (1980). Scavenging of aerosol particles by precipitation. Am. Meteorol. Soc. *19*, 715–722.

Rastogi, G., Coaker, G.L., and Leveau, J.H.J. (2013). New insights into the structure and function of phyllosphere microbiota through high-throughput molecular approaches. FEMS Microbiol. Lett. *348*, 1–10.

Reche, I., D'Orta, G., Mladenov, N., Winget, D.M., and Suttle, C.A. (2018). Deposition rates of viruses and bacteria above the atmospheric boundary layer. ISME J. *12*, 1154–1162.

Renard, P., Bianco, A., Baray, J.L., Bridoux, M., Delort, A.M., and Deguillaume, L. (2020). Classification of clouds sampled at the puy de Dôme station (France) based on chemical measurements and air mass history matrices. Atmosphere (Basel). *11*, 732.

Rho, M., Tang, H., and Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res. *38*, e191–e191.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47–e47.

Robinson, C.H. (2001). Cold adaptation in Arctic and Antarctic fungi. New Phytol. *151*, 341–353.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Romano, S., Di Salvo, M., Rispoli, G., Alifano, P., Perrone, M.R., and Talà, A. (2019). Airborne bacteria in the Central Mediterranean: Structure and role of meteorology and air mass transport. Sci. Total Environ. *697*.

Rothschild, L.J., and Mancinelli, R.L. (2001). Life in extreme environments. Nature *409*, 1092–1101.

Ruiz-Gil, T., Acuña, J.J., Fujiyoshi, S., Tanaka, D., Noda, J., Maruyama, F., and Jorquera, M.A. (2020). Airborne bacterial communities of outdoor environments and their associated influencing factors. Environ. Int. *145*, 106156.

Sajjad, W., Din, G., Rafiq, M., Iqbal, A., Khan, S., Zada, S., Ali, B., and Kang, S. (2020). Pigment production by cold-adapted bacteria and fungi: colorful tale of cryosphere with wide range applications. Extremophiles *24*, 447–473.

Salazar, G., Paoli, L., Alberti, A., Huerta-Cepas, J., Ruscheweyh, H.J., Cuenca, M., Field, C.M., Coelho, L.P., Cruaud, C., Engelen, S., et al. (2019). Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. Cell *179*, 1068-1083.e21.

Samaké, A., Bonin, A., Jaffrezo, J.L., Taberlet, P., Weber, S., Uzu, G., Jacob, V., Conil, S., and Martins, J.M.F. (2020). High levels of primary biogenic organic aerosols are driven by only a few plant-associated microbial taxa. Atmos. Chem. Phys. *20*, 5609–5628.

Sánchez-Romero, M.A., Cota, I., and Casadesús, J. (2015). DNA methylation in bacteria: From the methyl group to the methylome. Curr. Opin. Microbiol. *25*, 9–16.

Sands D.C (1982). The association between bacteria and rain and possible resultant meteorological implications.

Šantl-Temkiv, T., Finster, K., Hansen, B.M., Pašić, L., and Karlson, U.G. (2013). Viable methanotrophic bacteria enriched from air and rain can oxidize methane at cloud-like conditions. Aerobiologia (Bologna). *29*, 373–384.

Šantl-Temkiv, T., Amato, P., Gosewinkel, U., Thyrhaug, R., Charton, A., Chicot, B., Finster, K., Bratbak, G., and Löndahl, J. (2017). High-Flow-Rate Impinger for the Study of Concentration, Viability, Metabolic Activity, and Ice-Nucleation Activity of Airborne Bacteria. Environ. Sci. Technol. *51*, 11224–11234.

Šantl-Temkiv, T., Gosewinkel, U., Starnawski, P., Lever, M., and Finster, K. (2018). Aeolian dispersal

of bacteria in southwest Greenland: Their sources, abundance, diversity and physiological states. FEMS Microbiol. Ecol. *94*, 1–10.

Šantl-Temkiv, T., Sikoparija, B., Maki, T., Carotenuto, F., Amato, P., Yao, M., Morris, C.E., Schnell, R., Jaenicke, R., Pöhlker, C., et al. (2020). Bioaerosol field measurements: Challenges and perspectives in outdoor studies. Aerosol Sci. Technol. *54*, 520–546.

Šantl-Temkiv, T., Amato, P., Casamayor, E.O., Lee, P.K.H., and Pointing, S.B. (2022). Microbial ecology of the atmosphere. FEMS Microbiol. Rev.

Sattler, B., Puxbaum, H., and Psenner, R. (2001). Bacterial growth in supercooled cloud droplets. Geophys. Res. Lett. *28*, 239–242.

Schleper, C., Puehler, G., Holz, I., Gambacorta, A., Janekovic, D., Santarius, U., Klenk, H.P., and Zillig, W. (1995). Picrophilus gen. nov., fam. nov.: a novel aerobic, heterotrophic, thermoacidophilic genus and family comprising archaea capable of growth around pH 0. J. Bacteriol. *177*, 7050–7059.

Sellegri, K., Laj, P., Marinoni, A., Dupuy, R., Legrand, M., and Preunkert, S. (2003). Contribution of gaseous and particulate species to droplet solute composition at the Puy de Dôme, France. Atmos. Chem. Phys. *3*, 1509–1522.

Shakya, M., Lo, C.C., and Chain, P.S.G. (2019). Advances and challenges in metatranscriptomic analysis. Front. Genet. *10*.

Slekar, K.H., Kosman, D.J., and Culotta, V.C. (1996). The Yeast Copper/Zinc Superoxide Dismutase and the Pentose Phosphate Pathway Play Overlapping Roles in Oxidative Stress Protection *. J. Biol. Chem. *271*, 28831–28836.

Smets, W., Moretti, S., Denys, S., and Lebeer, S. (2016). Airborne bacteria in the atmosphere: Presence, purpose, and potential. Atmos. Environ. *139*, 214–221.

Smith, D.J., Timonen, H.J., Jaffe, D.A., Griffin, D.W., Birmele, M.N., Perry, K.D., Ward, P.D., and Roberts, M.S. (2013). Intercontinental dispersal of bacteria and archaea by transpacific winds. Appl. Environ. Microbiol. *79*, 1134–1139.

Smith, D.J., Ravichandar, J.D., Jain, S., Griffin, D.W., Yu, H., Tan, Q., Thissen, J., Lusby, T., Nicoll, P., Shedler, S., et al. (2018). Airborne bacteria in earth's lower stratosphere resemble taxa detected in the troposphere: Results from a new NASA Aircraft Bioaerosol Collector (ABC). Front. Microbiol. *9*, 1–20.

Sonwani, S., and Kulshrestha, U.C. (2019). PM10 carbonaceous aerosols and their real-time wet scavenging during monsoon and non-monsoon seasons at Delhi, India. J. Atmos. Chem. *76*, 171–200.

Stoddard, S.F., Smith, B.J., Hein, R., Roller, B.R.K., and Schmidt, T.M. (2015). rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Res. *43*, D593–D598.

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One *6*, 21800.

The UniProt Consortium (2019). UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res. *47*, D506–D515.

Tignat-perrier, R., Dommergue, A., Thollot, A., Keuschnig, C., Magand, O., Vogel, T.M., and Larose, C. (2019). Global airborne microbial communities controlled by surrounding landscapes and wind conditions. Sci. Rep. 1–11.

Tignat-Perrier, R., Dommergue, A., Thollot, A., Magand, O., Amato, P., Joly, M., Sellegri, K., Vogel, T.M., and Larose, C. (2020). Seasonal shift in airborne microbial communities. Sci. Total Environ. 137129.

Till, A., Lakhani, R., Burnett, S.F., and Subramani, S. (2012). Pexophagy: The Selective Degradation of Peroxisomes. Int. J. Cell Biol. *2012*.

Triadó-Margarit, X., Caliz, J., Reche, I., and Casamayor, E.O. (2019). High similarity in bacterial bioaerosol compositions between the free troposphere and atmospheric depositions collected at high-elevation mountains. Atmos. Environ. 79–86.

Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., and Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat. Methods *12*, 902–903.

Vaïtilingom, M., Amato, P., Sancelme, M., Laj, P., Leriche, M., and Delort, A.M. (2010). Contribution of microbial activity to carbon chemistry in clouds. Appl. Environ. Microbiol. *76*, 23–29.

Vaïtilingom, M., Deguillaume, L., Vinatier, V., Sancelme, M., Amato, P., Chaumerliac, N., and Delort, A.-M. (2013). Potential impact of microbial activity on the oxidant capacity and organic carbon budget in clouds. Proc. Natl. Acad. Sci. *110*, 559–564.

Vinatier, V., Wirgot, N., Joly, M., Sancelme, M., Abrantes, M., Deguillaume, L., and Delort, A.M. (2016). Siderophores in cloud waters and potential impact on atmospheric chemistry: Production by microorganisms isolated at the puy de Dôme station. Environ. Sci. Technol. *50*, 9315–9323.

Wex, H., Stratmann, F., Topping, D., and McFiggans, G. (2008). The Kelvin versus the raoult term in the köhler equation. J. Atmos. Sci. *65*, 4004–4016.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998). Prokaryotes: The unseen majority. Proc. Natl. Acad. Sci. U. S. A. *95*, 6578–6583.

Wilhelm, L., Besemer, K., Fasching, C., Urich, T., Singer, G.A., Quince, C., and Battin, T.J. (2014). Rare but active taxa contribute to community dynamics of benthic biofilms in glacier-fed streams. Environ. Microbiol. *16*, 2514–2524.

Willis, P.T., and Tattelman, P. (1989). Drop-size distribution associated with intense rainfall. J. Appl. Meteorol. *28*, 3–15.

Wirgot, N., Vinatier, V., Deguillaume, L., Sancelme, M., and Delort, A.M. (2017). H2O2 modulates the energetic metabolism of the cloud microbiome. Atmos. Chem. Phys. *17*, 14841–14851.

Womack, A.M., Bohannan, B.J.M., and Green, J.L. (2010). Biodiversity and biogeography of the atmosphere. Philos. Trans. R. Soc. B Biol. Sci. *365*, 3645–3653.

Womack, A.M., Artaxo, P.E., Ishida, F.Y., Mueller, R.C., Saleska, S.R., Wiedemann, K.T., Bohannan, B.J.M., and Green, J.L. (2015). Characterization of active and total fungal communities in the atmosphere over the Amazon rainforest. Biogeosciences *12*, 6337–6349.

Woo, C., and Yamamoto, N. (2020). Falling bacterial communities from the atmosphere. Environ. Microbiomes *15*, 22.

Wood, D.E., and Salzberg, S.L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. Genome Biol. *15*, 1–12.

Xu, C., Wei, M., Chen, J., Zhu, C., Li, J., Lv, G., Xu, X., Zheng, L., Sui, G., Li, W., et al. (2017). Fungi diversity in PM2.5 and PM1 at the summit of Mt. Tai: Abundance, size distribution, and seasonal variation. Atmos. Chem. Phys. *17*, 11247–11260.

Yan, D., Zhang, T., Su, J., Zhao, L.L., Wang, H., Fang, X.M., Zhang, Y.Q., Liu, H.Y., and Yu, L.Y. (2016). Diversity and composition of airborne fungal community associated with particulate matters in Beijing during haze and non-haze days. Front. Microbiol. *7*, 487.

Yang, K., Li, L., Wang, Y., Xue, S., Han, Y., and Liu, J. (2018). Airborne bacteria in a wastewater treatment plant: Emission characterization, source analysis and health risk assessment. Water Res. *149*,

596–606.

Ye, Y., and Tang, H. (2016). Utilizing de Bruijn graph of metagenome assembly for metatranscriptome analysis. Bioinformatics *32*, 1001–1008.

Ye, L., Wu, X., Tan, X., Shi, X., Li, D., Yu, Y., Zhang, M., and Kong, F. (2010). Cell Lysis of Cyanobacteria and Its Implications for Nutrient Dynamics. Int. Rev. Hydrobiol. *95*, 235–245.

Yilmaz, S., Allgaier, M., and Hugenholtz, P. (2010). Multiple displacement amplification compromises quantitative analysis of metagenomes. Nat. Methods *7*, 943–944.

Zhang, M., Khaled, A., Amato, P., Delort, A.M., and Ervens, B. (2021a). Sensitivities to biological aerosol particle properties and ageing processes: Potential implications for aerosol-cloud interactions and optical properties. Atmos. Chem. Phys. *21*, 3699–3724.

Zhang, N., Castlebury, L.A., Miller, A.N., Huhndorf, S.M., Schoch, C.L., Seifert, K.A., Rossman, A.Y., Rogers, J.D., Kohlmeyer, J., Volkmann-Kohlmeyer, B., et al. (2007). An overview of the systematics of the Sordariomycetes based on a four-gene phylogeny. Mycologia *98*, 1076–1087.

Zhang, Y., Zhao, Z., Dai, M., Jiao, N., and Herndl, G.J. (2014). Drivers shaping the diversity and biogeography of total and active bacterial communities in the South China Sea. Mol. Ecol. *23*, 2260–2274.

Zhang, Y., Thompson, K.N., Huttenhower, C., and Franzosa, E.A. (2021b). Statistical approaches for differential expression analysis in metatranscriptomics. Bioinformatics *37*, I34–I41.

1359

SUPPLEMENTARY INFORMATION to:

Clouds as atmospheric habitats for microorganisms

Raphaëlle Péguilhan[1], Florent Rossi[1], Engy Nasr[2], Bérénice Batut[2], François Enault[3], Pierre Amato[1]

[1] Université Clermont Auvergne, CNRS, SIGMA Clermont, ICCF, F-63000 CLERMONT-FERRAND, FRANCE
[2] University of Freiburg, Department of Computer Science, Bioinformatics group, 79110 FREIBURG, Germany.
[3] Université Clermont Auvergne, CNRS, Laboratoire Microorganismes : Génome et Environnement, F-63000 CLERMONT-FERRAND, France.

List of supplements:

Supplementary Table 1: Processing information for (A) clouds and (B) aerosols metagenomes.

Supplementary Table 2: Processing information for (A) clouds and (B) aerosols metatranscriptomes.

Supplementary Table 3: Chemical characteristics of aerosol and cloud samples.

Supplementary Figure 4: (A) Correlation between meteorological and biological data in aerosols, (B) and chemical content in clouds. Spearman's correlation.

Supplementary Table 5: Taxonomy table for Bacteria. A: 30 top phyla, **B**: 30 top orders, and **C**: 30 top genera tables ordered by abundance. Sample name: "A" or "C" for respectively Aerosol or cloud sample, and the date (mm/dd).

Supplementary Table 6: Taxonomy table for Eukaryota. A: all phyla, **B**: all orders, and **C**: 30 top genera tables ordered by abundance. Sample name: "A" or "C" for respectively Aerosol or cloud sample, and the date (mm/dd).

Supplementary Table 7: Taxonomy table for Archaea. A: all phyla, **B**: all orders, and **C**: 30 top genera tables ordered by abundance. Sample name: "A" or "C" for respectively Aerosol or cloud sample, and the date (mm/dd).

Supplementary Table 8: Taxonomy table for Viruses. A: all phyla, and **B**: all orders tables ordered by abundance. Sample name: "A" or "C" for respectively Aerosol or cloud sample, and the date (mm/dd).

Supplementary Table 9: Differential expression analysis (DEA) coefficients for the 30 most overexpressed A) bacterial families, and B) bacterial genera, and the total overexpressed C) eukaryotic families, and D) eukaryotic genera. Positive significant results (i.e., coef) from MTXmodel R package. Metadata: feature name; value: reference category for coefficient values; coef: DEA result; stderr: standard deviation of the model; N: total number of data point; N.not.0: total of non-zero data point; pval: p-value from the calculation; qval: computed with the correlation method of the model (p.adjust) (cf R package Maaslin2 protocol).

Supplementary Figure 10: RNA:DNA content ratios at genus level for (A) bacteria and (B) eukaryotes. The main heatmap for bacteria (left) represents the 176 top abundant genera from the heatmap including all genera (1,250 genera; lower right). Hierarchical clusterings were done with the Ward's method (ward.D2). Intensity scale depicts log abundances of corresponding sequencing reads. EnvType: environment type. Sample name = "A" for aerosol or "C" for cloud and sampling date under the format "mmdd" (month and day).

Supplementary Table 11: The 20 most overrepresented Gene Ontologies (GOs) for each GO categories (Cellular Component, Biological process, and Molecular Function) in clouds (positive coefficients) or in aerosols (negative coefficients). Mean by GO of the coefficient from the differential expression analysis (DEA) for genes in clouds versus aerosols.

Supplementary Figure 1: Seventy-two-hours backward trajectory plots and associated sources areas for each sampling event, extracted from ERA5 data analysis.

Supplementary Figure 2: Stacked numbers of bacterial and eukaryotic sequences affiliated with kraken2 in metagenomic (MG) and metatranscriptomic (MT) datasets. A: in MG data; **B**: in MT data; C: mean proportions in MG and MT data.

Supplementary Figure 3: Distribution of the most abundant bacterial orders (top cluster), and corresponding hierarchical clusterings (Ward's method, "ward.D2"). Intensity scale depicts centered-log ratio (clr) abundances. EnvType: environment type. Sample name = "A" for aerosol or "C" for cloud and sampling date under the format "mmdd" (month and day).

Supplementary Figure 4: Distribution of bacterial genera over samples. A: all 1250 genera; **B**: focus on the first top cluster (red asterisk). Hierarchical clusterings were done using the Ward's method (ward.D2). Intensity scale describes centered-log ratio (clr) abundances. EnvType: environment type. Sample name = "A" for aerosol or "C" for cloud and sampling date under the format "mmdd" (month and day).

Supplementary Figure 5: Distribution of the total eukaryotic orders, and corresponding hierarchical clusterings (Ward's method, "ward.D2"). Intensity scale depicts centered-log ratio (clr) abundances. EnvType: environment type. Sample name = "A" for aerosol or "C" for cloud and sampling date under the format "mmdd" (month and day).

Supplementary Figure 6: Distribution of the total 55 eukaryotic genera over samples. Hierarchical clusterings were done using the Ward's method (ward.D2). Intensity scale describes centered-log ratio (clr) abundances. EnvType: environment type. Sample name = "A" for aerosol or "C" for cloud and sampling date under the format "mmdd" (month and day).

Supplementary Figure 7: Distribution of the 14 archaeal orders over samples. Hierarchical clusterings were done using the Ward's method (ward.D2). Intensity scale describes centered-log ratio (clr) abundances. EnvType: environment type. Sample name = "A" for aerosol or "C" for cloud and sampling date under the format "mmdd" (month and day).

Supplementary Figure 8: Principal component analysis based on the total biodiversity in metagenomic (MG) and metatranscriptomic (MT) data for clouds and aerosols (air). Based on 6,373 taxa. Count data were centered-log ratio (clr) transformed. (R package factoextra, ellipse type: "confidence")

Supplementary Figure 9: Significantly differentially expressed bacterial and eukaryotic families (A) and genera (B) between cloud and aerosol samples. Values represent differential expression analysis coefficients from MTXmodel R package. Positive coefficient means taxon is more expressed in clouds.

Supplementary Figure 10: Significantly differentially overrepresented bacterial and eukaryotic families (A) and genera (B) between cloud and aerosol samples. Values represent differential expression analysis coefficients from MTXmodel R package. Positive coefficient means taxon is more expressed in clouds.

Supplementary Figure 11: Proportions (%) of overexpressed genes affiliated to eukaryota or bacteria and their respective phyla. Based on the 488 significantly expressed genes from differential expression analysis.

Supplementary Figure 12: RNA:DNA log ratios for each gene entries in cloud and aerosol samples. Only 8,627 genes are represented here (ratios were not calculable for the others). Red scale means values of RNA:DNA ratios in log. RNA and DNA data were first normalized as relative counts. EnvType: environment type. Sample name = "A" for aerosol or "C" for cloud and sampling date under the format "mmdd" (month and day).

Supplementary Figure 13: RNA data against DNA data plots for all the samples, or clouds and aerosols only. Based on the 21,046 genes IDs recovered.

Supplementary Figure 14: Gene ontology (GO) relationship tree for Cellular Components related to the proton-transporting ATP synthase complex in clouds and aerosols. The blue shade scale represents the Differential Expression Analysis (DEA) coefficient value, with positive values indicating a significant overrepresentation in clouds as opposed to aerosols. The size of the nodes is scaled by the absolute value of the DEA coefficient.

Supplementary Figure 15: Gene ontology (GO) relationship tree for Biological Processes related to translation in clouds and aerosols. The red to blue color scale represents the Differential Expression Analysis (DEA) coefficient value, with negative values (red shades) indicating a significant overrepresentation in aerosols as opposed to positive values (blue shades) signifying an overrepresentation in clouds. The size of the nodes is scaled by the absolute value of the DEA coefficient.

Supplementary Figure 16: Gene ontology (GO) relationship tree for Biological Processes related to ATP synthesis and ion transport in clouds and aerosols. The red to blue colour scale represents the Differential Expression Analysis (DEA) coefficient value, with negative values (red shades) indicating a significant overrepresentation in aerosols as opposed to positive values (blue shades) signifying an overrepresentation in clouds. The size of the nodes is scaled by the absolute value of the DEA coefficient.

Tables and Figures

Supplementary Table 1: Processing information for (A) clouds and (B) aerosols metagenomes.

A)	CLOUD 20191002	CLOUD 20191022	CLOUD 20200311	CLOUD 20200717	CLOUD 20201016	CLOUD 20201028	CLOUD 20201103	CLOUD 20201110	CLOUD 20201119
Nb of raw reads	65 812 666	259 998 456	58 746 330	43 757 944	97 184 944	66 400 646	54 975 726	60 932 548	77 768 050
After trimming	41 675 340	175 692 186	40 113 524	29 200 252	64 669 674	42 495 176	35 964 782	40 917 502	51 587 742
% removed	37 %	32 %	32 %	33 %	33 %	36 %	35 %	33 %	34 %
Nb of rRNA reads	479 510	2 105 871	681 376	497 119	825 886	553 734	461 444	569 711	608 520
% of rRNA reads	1.15 %	1.20 %	1.70 %	1.70 %	1.28 %	1.30 %	1.28 %	1.39 %	1.18 %
Nb of non-rRNA reads	41 195 830	173 586 315	39 432 148	28 703 133	63 843 788	41 941 442	35 503 338	40 347 791	50 979 222
% of non-rRNA reads	98.85 %	98.80 %	98.30 %	98.30 %	98.72 %	98.70 %	98.72 %	98.61 %	98.82 %
% of human reads in non-rRNA reads	0.11 %	0.71 %	0.02 %	0.04 %	0.08 %	0.14 %	0.07 %	0.08 %	0.09 %
Nb of assembled contig	194 547	495 663	275 249	129 639	316 632	213 954	193 399	241 377	207 227
Nb de reads properly paired	2 060 240	17 067 006	3 109 058	1 853 656	5 065 598	3 131 000	3 707 712	3 154 518	2 198 260
% of properly mapped reads	5 %	9.9 %	7.9 %	6.4 %	7.9 %	7.5 %	10.4 %	7.8 %	4.3 %
Nb of affiliated reads	554 018	2 271 135	480 815	506 200	769 566	469 334	337 543	374 346	597 562
% affiliated	2.66 %	2.59 %	2.40 %	3.47 %	2.38 %	2.21 %	1.18 %	1.83 %	2.32 %
Nb of affiliated human reads	41 379	654 422	19 638	19 244	38 655	35 478	18 311	26 211	39 429
% of human reads	7.5 %	28.8 %	4.1 %	3.8 %	5 %	7.6 %	5.4 %	7 %	6.6 %

B)

	AIR 20200707	AIR 20200708	AIR 20200709	AIR 20200922	AIR 20201118	AIR 20201124
Nb of raw reads	47 409 014	30 435 818	44 295 022	41 909 288	40 730 342	41 086 722
After trimming	31 161 586	19 145 910	30 985 400	31 152 980	25 543 444	28 294 184
% removed	34 %	37 %	30 %	26 %	37 %	31 %
Nb of rRNA reads	524 095	361 927	333 292	437 338	384 484	305 043
% of rRNA reads	1.68 %	1.89 %	1.08 %	1.40 %	1.51 %	1.08 %
Nb of non-rRNA reads	30 637 491	18 783 983	30 652 108	30 715 642	25 158 960	27 989 141
% of non-rRNA reads	98.32 %	98.11 %	98.92 %	98.60 %	98.49 %	98.92 %
% of human reads in non-rRNA reads	0.02 %	0.02 %	0.01 %	0.02 %	0.14 %	-
Nb of assembled contig	140 392	43 627	99 268	179 815	101 745	-
Nb de reads properly paired	4 511 490	2 149 274	3 511 940	5 361 278	1 450 314	-
% of properly mapped reads	14.6 %	11.3 %	11.4 %	17.4 %	5.8 %	-
Nb of affiliated reads	628 977	361 163	614 940	303 720	499 417	-
% affiliated	4.04 %	3.77 %	3.97 %	1.95 %	3.91 %	-
Nb of affiliated human reads	48 092	27 287	14 197	23 093	36 927	-
% of human reads	7.6 %	7.6 %	2.3 %	7.6 %	7.4 %	-

1509

Supplementary Table 2: Processing information for (A) clouds and (B) aerosols metatranscriptomes.

A)	CLOUD 20191002	CLOUD 20191022	CLOUD 20200311	CLOUD 20200717	CLOUD 20201016	CLOUD 20201028	CLOUD 20201103	CLOUD 20201110	CLOUD 20201019
Nb of raw reads	93 499 990	186 010 124	82 131 152	81 247 584	79 489 694	85 009 094	110 129 198	69 916 188	195 503 428
After trimming	64 702 194	94 221 480	61 733 384	60 764 224	54 020 462	64 420 980	71 515 884	50 964 222	132 952 184
% removed	31 %	49 %	25 %	25 %	32 %	24 %	35 %	27 %	32 %
Nb of rRNA reads	59 259 226	85 291 849	58 058 408	56 202 145	43 175 946	57 199 748	28 842 768	45 243 381	115 553 819
% of rRNA reads	91.59 %	90.52 %	94.05 %	92.49 %	79.93 %	88.79 %	40.33 %	88.77 %	86.91 %
Nb of non-rRNA reads	5 442 968	8 929 631	3 674 976	4 562 079	10 844 516	7 221 232	42 673 116	5 720 841	17 398 365
% of non-rRNA reads	8.41 %	9.48 %	5.95 %	7.51 %	20.07 %	11.21 %	59.67 %	11.23 %	13.09 %
% of human reads in non-rRNA reads	0.86 %	0.05 %	0.32 %	0.42 %	0.82 %	0.46 %	**89.47** %	0.29 %	0.31 %
Nb de reads properly paired	275 906	655 764	262 850	275 216	833 402	697 632	464 440	423 202	1 007 468
% of properly mapped reads	4.6 %	7 %	7 %	5.7 %	7.5 %	8.9 %	3.9 %	7.1 %	5.6 %
Nb of affiliated reads	24 817 282	37 049 964	26 829 087	22 589 666	18 755 564	24 842 334	32 332 229	17 680 329	52 249 268
% affiliated	76.71 %	78.64 %	86.92 %	74.35 %	69.44 %	77.12 %	90.42 %	69.38 %	78.60 %
Nb of affiliated human reads	31 612	9 229	15 483	26 700	64 652	25 003	**≈ 19 000 000**	22 925	46 348
% of human reads	0.1 %	0.02 %	0.06 %	0.1 %	0.3 %	0.1 %	**≈ 59** %	0.1 %	0.1 %

B)

	AIR 20200707	AIR 20200708	AIR 20200709	AIR 20200922	AIR 20201118	AIR 20201124
Nb of raw reads	71 487 464	116 985 074	96 661 644	65 554 022	76 813 684	68 355 146
After trimming	49 784 022	63 268 244	56 706 232	48 733 804	54 728 858	54 728 858
% removed	30 %	46 %	41 %	26 %	29 %	20 %
Nb of rRNA reads	41 199 078	51 568 599	49 530 808	42 808 640	45 507 747	5 986 089
% of rRNA reads	82.76 %	81.51 %	87.35 %	87.84 %	83.15 %	**12.06** %
Nb of non-rRNA reads	8 584 944	11 699 645	7 175 424	5 925 164	9 221 111	43 642 337
% of non-rRNA reads	17.24 %	18.49 %	12.65 %	12.16 %	16.85 %	**87.94** %
% of human reads in non-rRNA reads	1.95 %	1.37 %	12.02 %	0.2 %	0.7 %	-
Nb de reads properly paired	648 418	383 000	190 608	667 098	467 706	-
% of properly mapped reads	7.6 %	3.3 %	2.7 %	10.8 %	4.9 %	-
Nb of affiliated reads	18 903 493	22 757 792	22 429 335	19 290 161	17 432 686	-
% affiliated	75.94 %	71.94 %	79.11 %	79.17 %	63.71 %	-
Nb of affiliated human reads	103 756	119 150	164 072	18 204	96 729	-
% of human reads	0.5 %	0.5 %	0.7 %	0.1 %	0.6 %	-

Supplementary Table 3: Chemical characteristics of cloud samples.

Sample ID	Cloud category (Renard et al., 2020)	pH	Anions (µM)			Cations (µM)				
			Cl^-	NO_3^-	SO_4^{2-}	Na^+	NH_4^+	K^+	Mg^{2+}	Ca^{2+}
CLOUDS										
20191002CLOUD	Marine	5.2	109.0	42.6	13.2	120.4	4.2	39.0	7.5	19.4
20191022CLOUD	Marine	5.8	52.0	10.0	11.7	68.7	0.0	14.2	3.9	7.4
20200311CLOUD	Marine	5.1	37.8	12.8	21.5	53.0	61.8	16.5	4.1	6.0
20200717CLOUD	Continental	5.4	39.6	162.1	40.0	54.7	339.8	12.6	13.3	36.2
20201016CLOUD	Marine / Continental	NA*	18.0	36.1	16.5	29.2	91.8	14.5	6.8	16.0
20201028CLOUD	Continental	5.5	75.1	13.8	7.5	78.8	52.3	14.7	14.2	24.5
20201103CLOUD	NA*	5.4	782.0	56.8	67.8	128.8	189.4	24.4	83.6	52.6
20201110CLOUD	Marine	5.1	26.9	51.3	15.6	4.5	160.3	5.1	3.6	13.0
20201119CLOUD	Continental	5.0	152.3	11.4	15.6	131.1	51.6	10.9	19.7	19.9
Minimum	**-**	**5.0**	**18.0**	**10.0**	**7.5**	**4.5**	**0.0**	**5.1**	**3.6**	**6.0**
Maximum	**-**	**5.8**	**782.0**	**162.1**	**67.8**	**131.1**	**339.8**	**39.0**	**83.6**	**52.6**
Median	**-**	**5.3**	**52.0**	**36.1**	**15.6**	**68.7**	**61.8**	**14.5**	**7.5**	**19.4**
Mean	**-**	**5.3**	**143.6**	**44.1**	**23.3**	**74.3**	**105.7**	**16.9**	**17.4**	**21.7**
Standard error	**-**	**0.3**	**243.3**	**47.8**	**19.1**	**44.9**	**108.5**	**9.7**	**25.4**	**14.8**

NA* No data available

Supplementary Figure 4: (A) Correlation between meteorological and biological data in aerosols, (B) and chemical content in clouds. Spearman's correlation.

A)	**Average temperature (°C)**	**Relative humidity (%)**	**Wind speed (average [max]) (m s^{-1})**	**Sampling duration (h)**	**Microbial cell number concentration (cells m^{-3})**	**ATP concentration (pmol m^{-3})**	**RNA:DNA ratio (ng m^{-3})**	**Number of gene in MG data**	**Number of gene in MT data**	**MT/MG ratio for annotated gene**
p-value / Curve coeff	**A**	**B**	**C**	**B**	**E**	**F**	**G**	**H**	**I**	**J**
A		0.49722	0.95972	0.83889	0.80278	0.35556	1	0.86667	0.29167	1
B	-0.37143		0.175	0.44444	0.175	0.65833	0.49722	0.6	1	0.73333
C	0.028571	-0.65714		0.93889	0.75833	0.95972	0.95972	0.68333	0.68333	0.68333
D	-0.11595	0.40584	-0.057977		0.16111	0.50556	0.0055556	0.05	0.29167	0.35
E	-0.14286	-0.65714	0.14286	-0.66674		0.24167	0.11944	0.13333	0.68333	0.45
F	0.48571	0.25714	0.028571	0.34786	-0.6		0.35556	0.6	0.18333	0.6
G	-0.028571	-0.37143	0.028571	-0.98561	0.71429	-0.48571		0.083333	0.35	0.29167
H	-0.1	-0.3	0.3	-0.9	0.8	-0.3	0.9		0.23333	0.18333
I	-0.6	0	0.3	-0.6	0.3	-0.7	0.6	0.7		0.73333
J	0	-0.2	0.3	0.6	-0.5	-0.3	-0.6	-0.7	-0.2	

B)	**Average temperature (°C)**	**Relative humidity (%)**	**Wind speed (average [max]) (m s^{-1})**	**Sampling duration (h)**	**Microbial cell number concentration (cells m^{-3})**	**ATP concentration (pmol m^{-3})**	**RNA:DNA ratio (ng m^{-3})**	**Number of gene in MG data**	**Number of gene in MT data**	**MT/MG ratio for annotated gene**
p-value / Curve coeff	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
A		0.11508	0.41008	0.64364	0.93	0.80998	0.52063	0.7928	0.1938	0.23003
B	-0.60553		0.9127	0.8082	0.44841	0.9127	0.8082	0.8082	0.74868	0.76852
C	-0.31667	0.055048		0.22982	0.1938	0.98157	0.14015	0.84318	0.52063	0.51014
D	-0.18333	0.1101	0.45		0.033995	0.37195	0.74354	0.43662	0.28039	0.19414
E	0.033333	-0.30277	0.48333	0.71667		0.20316	0.55171	0.47816	0.24992	0.21213
F	0.1	-0.055048	-0.016667	0.33333	0.46667		0.13278	0.43662	0.69291	0.55534
G	-0.25	0.1101	0.53333	-0.13333	-0.23333	-0.55		0.88009	0.55171	0.61678
H	-0.1	-0.1101	-0.083333	-0.3	0.26667	0.3	-0.066667		0.28039	0.36146
I	-0.48333	-0.13762	0.25	0.4	0.43333	0.15	0.23333	0.4		2.20E-05
J	-0.44539	-0.1249	0.25211	0.47901	0.4622	0.2269	0.19328	0.34455	0.99163	
K	-0.16667	-0.22019	0.58333	-0.28333	-0.066667	-0.066667	0.4	-0.05	-0.1	-0.14286
L	0.45	-0.082572	-0.56667	-0.4	-0.33333	0.5	-0.38333	0.033333	-0.46667	-0.42859
M	-0.15063	0.35931	-0.49373	-0.20921	-0.53557	0.0083683	-0.27615	-0.43515	-0.45189	-0.44304
N	-0.28333	-0.22019	0.55	-0.26667	-0.066667	-0.1	0.33333	-0.016667	0.016667	-0.033615
O	0.1	0.35781	-0.53333	-0.11667	-0.46667	0.38333	-0.28333	-0.28333	-0.36667	-0.30253
P	-0.016667	-0.027524	0.26667	0.066667	0.45	0.33333	-0.45	0.16667	-0.35	-0.36976
Q	-0.36667	0.082572	0.28333	-0.2	-0.21667	0.3	0.016667	-0.083333	-0.066667	-0.067229
R	0	-0.13762	0.05	-0.23333	-0.21667	0.53333	-0.083333	-0.05	-0.1	-0.067229
S	0.30123	-0.50408	0.49401	0.69885	0.77114	0.6145	-0.28918	-0.21688	0.26508	0.35976

1519

1520

	Cl-	NO3-	SO42-	Na+	NH4+	K+	Mg2+	Ca2+	pH
p-value / Curve coeff	K	L	M	N	O	P	Q	R	S
A	0.66069	0.22982	0.69971	0.46299	0.80998	0.98157	0.33626	1	0.46885
B	0.5873	0.85185	0.36971	0.5873	0.3664	0.98148	0.85185	0.74868	0.21905
C	0.108	0.12057	0.18031	0.13278	0.14753	0.47816	0.44981	0.91162	0.21528
D	0.46299	0.29119	0.58873	0.49333	0.77563	0.88009	0.6134	0.55171	0.063492
E	0.88009	0.38532	0.14145	0.88009	0.21252	0.22982	0.58094	0.58094	0.032937
F	0.88009	0.17766	0.99182	0.7928	0.3125	0.37195	0.43662	0.14015	0.11151
G	0.28039	0.3125	0.47048	0.37195	0.46299	0.22982	0.96498	0.84318	0.48512
H	0.91162	0.93	0.24167	0.98157	0.46299	0.67774	0.84318	0.91162	0.60615
I	0.7928	0.21252	0.22364	0.96498	0.33626	0.34742	0.88009	0.7928	0.52817
J	0.71506	0.25033	0.23088	0.94017	0.4267	0.32698	0.86914	0.86914	0.37798
K		0.94839	0.90351	0.00010747	0.52063	0.33626	0.02139	0.096798	0.82718
L	-0.033333		0.11009	0.66069	0.010769	0.82658	0.56632	0.108	0.96667
M	-0.05021	0.57741		0.80033	0.015289	0.85395	0.49798	0.44202	0.40992
N	0.96667	-0.16667	-0.10042		0.34742	0.37195	0.013828	0.12057	0.98849
O	-0.25	0.81667	0.78662	-0.35		0.66069	0.64364	0.22982	0.78333
P	0.36667	0.083333	0.075314	0.33333	-0.16667		0.43662	0.74354	0.50238
Q	0.76667	0.21667	0.25942	0.8	0.18333	0.3		0.0045084	0.98849
R	0.6	0.58333	0.29289	0.56667	0.45	0.13333	0.86667		0.58671
S	0.096393	0.024098	-0.33334	-0.012049	-0.12049	0.27713	0.012049	0.22893	

Supplementary Table 5: Taxonomy table for Bacteria. **A**: 30 top phyla, **B**: 30 top orders, and **C**: 30 top genera tables ordered by abundance. Sample name: "A" or "C" for respectively Aerosol or cloud sample, and the date (mm/dd).

A)

Phylum	A0707	A0708	A0709	A0922	A1118	C0311	C0717	C1002	C1016	C1022	C1028	C1103	C1110	C1119
Proteobacteria	241883	126822	217697	70274	121534	138320	95497	64808	169482	206092	107057	53473	38582	63675
Actinobacteria	157964	103574	216101	55133	116627	68614	82122	71848	124682	192282	69726	50782	41189	61282
multi-affiliation	94614	56382	94962	41621	44191	63561	34877	22630	38418	56424	12214	10452	12746	17298
Bacteroidetes	14436	6759	11645	3073	23685	6811	5926	3440	16486	106260	3281	1528	3164	6056
Firmicutes	6305	4989	8693	3929	7055	7698	5365	21967	8785	121837	3442	2308	2763	6812
Cyanobacteria	4048	1834	3397	2124	468	3711	4316	576	4129	2838	217	208	747	1922
Deinococcus-Thermus	1027	459	708	153	1434	1438	619	733	778	2527	235	140	284	407
Planctomycetes	1125	689	1650	114	782	421	376	283	435	482	105	143	236	282
Acidobacteria	779	434	798	111	714	362	243	274	207	333	168	75	99	301
Gemmatimonadetes	291	164	307	23	241	129	122	77	81	277	199	15	14	24
Verrucomicrobia	135	89	134	10	72	46	54	14	148	31	8	3	28	23
Chloroflexi	81	49	73	13	57	25	43	38	43	28	14	4	9	13
Chlamydiae	34	6	21	1	41	3	7	20	25	38	4	3	2	41
Nitrospirae	41	24	35	0	22	19	23	17	10	9	3	3	1	4
Fusobacteria	3	4	0	1	3	1	7	8	10	116	0	46	2	1
Tenericutes	1	0	7	7	29	44	4	11	9	13	2	1	8	4
Spirochaetes	12	0	5	0	12	18	3	8	9	59	3	0	7	3
Candidatus Saccharibacteria	15	12	18	0	17	7	5	1	3	3	3	7	4	1
Armatimonadetes	2	0	6	0	6	0	1	3	3	5	3	3	0	4
Chlorobi	0	1	6	0	1	1	1	1	4	4	0	1	4	3
Deferribacteres	0	0	0	0	2	0	0	0	0	6	2	2	0	3
Aquificae	0	0	0	0	1	1	0	1	1	0	0	0	2	6
Fibrobacteres	4	0	1	0	3	0	0	1	0	0	0	0	0	0
Candidatus Cloacimonetes	0	0	4	0	2	0	0	0	0	0	0	0	0	0
Elusimicrobia	0	0	0	0	1	0	0	0	5	0	0	0	0	0
Candidatus Omnitrophica	3	0	0	0	0	0	0	0	0	0	2	0	0	0
Thermotogae	0	0	0	1	0	0	0	0	0	0	0	0	0	3
Balneolaeota	0	0	0	0	0	0	3	0	0	0	0	0	0	0
Calditrichaeota	0	0	0	0	0	0	0	2	0	0	0	0	1	0
Synergistetes	0	0	0	0	0	0	0	0	0	1	0	2	0	0

B)

Order	A0707	A0708	A0709	A0922	A1118	C0311	C0717	C1002	C1016	C1022	C1028	C1103	C1110	C1119
multi-affiliation	235280	142223	267932	74777	142075	140275	100236	68345	117683	156287	40715	39592	38968	55327
Micrococcales	32060	21198	39994	6868	23015	11954	24361	16657	60619	38666	32386	16244	9688	16073
Hyphomicrobiales	55143	24845	57641	3889	30135	16900	15732	13182	24635	25712	8245	4269	8397	12786
Propionibacteriales	18245	13502	28330	2191	15317	9922	10108	9779	10162	75021	8051	6384	6957	8671
Corynebacteriales	24045	13381	31949	36685	13648	9247	6565	10628	13120	22396	6966	3094	5530	10738
Burkholderiales	20757	11582	22235	4293	15643	9982	7092	9652	37591	23868	6841	3853	5371	15396
Pseudomonadales	19932	9698	14369	6759	6073	39708	5727	8152	25025	31057	8091	4192	6192	5216
Cytophagales	10116	5109	8773	2509	19730	5024	4723	1913	13827	101887	2193	1134	1865	4630
Pasteurellales	43434	18409	18082	24580	161	17153	6533	357	737	6702	180	23	95	138
Sphingomonadales	17862	10618	20019	2654	16551	6706	7230	3823	9474	10625	4642	3028	3860	5259
Lactobacillales	848	428	917	938	1773	3276	1229	10757	2569	65477	1613	1403	1038	2771
Rhodobacterales	1800	1272	2818	1462	1631	1182	11714	4182	8956	12456	13035	29880	877	3240
Bacillales	3662	2382	4173	1799	3824	2451	2993	10407	5356	35881	1313	763	1315	3340
Rhodospirillales	1130	582	1227	357	1790	833	1404	781	8764	5942	52145	519	560	465
Enterobacterales	8968	4922	3292	815	2092	1598	1037	4471	2232	4596	1097	473	644	2349
Streptomycetales	4748	3088	5219	537	2998	1738	5286	2121	2406	3304	732	875	1442	1671
Eubacteriales	1461	1926	3125	1053	971	1678	885	454	445	18846	384	78	331	409
Xanthomonadales	3439	1620	4960	477	2276	2017	4626	1809	2208	3041	893	300	1257	1367
Geodermatophilales	1133	829	1023	262	570	312	601	1112	421	12256	466	269	301	223
Micromonosporales	2240	1424	3418	180	1505	1299	966	980	1699	1395	2875	204	462	547
Pseudonocardiales	1760	1261	2205	249	2048	778	689	1097	1437	1503	330	495	625	612
Caulobacterales	1136	527	1517	291	2421	527	508	795	1378	2747	681	248	568	537
Flavobacteriales	1661	435	1142	225	1921	1048	525	724	1233	1558	442	164	722	618
Myxococcales	1571	959	1950	96	599	300	1151	151	4052	349	214	87	147	226
Deinococcales	963	442	694	151	1398	1421	600	711	747	2316	223	134	273	392
Nostocales	1275	434	552	728	113	429	1469	178	1992	1874	96	52	215	100
Legionellales	30	43	11	7	46	2	8	6	19	8372	19	13	7	14
Oscillatoriales	979	914	1394	71	53	56	1304	55	102	523	5	9	199	1152
Streptosporangiales	760	485	823	114	578	400	793	419	393	588	82	114	509	654
Neisseriales	16	2	21	1	8	18	34	117	42	5883	17	7	9	26

C)

Genus	A0707	A0708	A0709	A0922	A1118	C0311	C0717	C1002	C1016	C1022	C1028	C1103	C1110	C1119
multi-affiliation	316951	187178	366276	89630	199255	175592	133853	93936	168533	208033	60805	54252	54220	83099
Hymenobacter	8465	4355	7415	2217	18135	4042	4187	1492	12777	99704	1792	555	1397	4258
Pasteurella	43428	18406	18077	24579	155	17137	6514	335	708	139	167	17	89	125
Bradyrhizobium	28837	11427	24727	1439	10746	6784	7225	6413	7754	6108	1643	1291	2961	3245
Pseudomonas	12502	4642	6349	1219	4244	34531	2606	2900	7302	13170	3307	2873	4393	1886
Cutibacterium	377	129	236	496	425	308	584	3227	2255	68516	2320	1029	3088	1247
Kocuria	559	481	791	751	2153	466	5273	2994	30076	4896	18589	7222	1314	4697
Paracoccus	706	508	1039	1066	786	510	9818	3378	6912	9038	12241	28976	574	1648
Nocardioides	10283	7946	16524	885	7439	5190	5341	2994	4416	2577	3220	1620	1737	3980
Streptococcus	107	110	198	254	934	233	781	5872	537	56807	1104	1021	581	1418
Sphingomonas	10652	6562	11401	1431	9915	2783	3472	1929	4968	5296	2188	1016	2030	2894
Acinetobacter	7228	4868	7853	5107	965	4606	2126	3166	8298	13108	3148	633	1240	2453
Roseomonas	81	60	112	123	78	76	629	94	5543	903	51847	287	89	58
Micrococcus	385	230	116	1128	1880	900	2965	3821	10312	5201	3272	442	843	1608
Massilia	2080	1183	1967	1064	1357	359	780	2685	16094	2434	497	390	273	1836
Streptomyces	4252	2789	4645	487	2711	1577	5063	1928	2114	3070	638	760	1352	1515
Mycobacteroides	12	1	16	29563	39	9	6	3	4	92	118	8	25	23
Corynebacterium	467	240	1029	699	3267	2523	943	4053	3054	4742	4132	484	1273	1645
Mycolicibacterium	4443	2665	6247	288	2511	1517	1210	1442	2638	1315	790	1012	1044	1266
Methylobacterium	1666	1453	5937	400	3406	1518	783	431	3180	4984	986	352	1069	2046
Clostridium	1086	1763	2648	933	205	1396	586	132	210	18243	282	42	198	84
Rhodococcus	4913	2526	6012	316	1094	1322	968	461	877	994	327	221	767	5066
Staphylococcus	451	183	252	604	1723	465	1099	6000	2721	8039	804	496	759	1226
Microbacterium	3318	1751	3747	410	2008	1074	1286	979	2124	1450	1116	2693	1070	1079
Mycobacterium	2279	1156	3259	515	1013	547	576	970	944	11429	189	201	387	473
Janthinobacterium	186	74	129	359	985	832	633	784	2455	8302	1803	1653	1692	2196
Moraxella	128	144	124	398	687	289	931	1957	9122	3505	1501	657	493	792
Listeria	1	1	1	0	0	1	3	2	2	20440	0	0	0	2
Rathayibacter	3553	3053	4665	233	1287	315	1061	655	993	359	208	187	586	470
Pantoea	5176	3682	2229	216	1451	310	347	400	570	408	126	24	182	425

Supplementary Table 6: Taxonomy table for Eukaryota. A: all phyla, **B**: all orders, and **C:** 30 top genera tables ordered by abundance. Sample name: "A" or "C" for respectively Aerosol or cloud sample, and the date (mm/dd).

A)

Phylum	A0707	A0708	A0709	A0922	A1118	C0311	C0717	C1002	C1016	C1022	C1028	C1103	C1110	C1119
Ascomycota	19604	12031	18281	30321	42546	31146	124330	188049	113338	249571	56014	48123	70019	215881
multi-affiliation	4103	2574	3444	24354	47034	59297	58861	61160	111559	303916	84131	71126	78988	83656
Basidiomycota	1995	1519	3821	12793	16589	19135	30873	28802	61586	154015	39965	34461	43847	39108
Apicomplexa	224	162	114	269	457	757	369	331	640	2106	415	180	524	302
Bacillariophyta	29	18	34	175	1086	1616	306	540	131	188	116	100	70	57
Euglenozoa	13	20	5	15	70	90	24	68	83	515	71	32	383	15
Evosea	2	2	2	9	32	17	4	27	30	33	13	7	10	17
Cercozoa	1	0	0	4	2	7	2	1	9	11	8	0	2	2
Microsporidia	0	0	0	0	3	0	0	0	0	0	0	0	1	0

B)

Order	A0707	A0708	A0709	A0922	A1118	C0311	C0717	C1002	C1016	C1022	C1028	C1103	C1110	C1119
multi-affiliation	7862	4741	6630	42040	75351	90665	102962	112469	198492	501984	130701	112425	134649	163100
Helotiales	12048	7059	10949	11892	14141	4707	79213	114759	17036	42040	13670	5616	5594	31766
Tremellales	297	193	303	2424	5762	5746	10281	6719	16714	53102	13962	13304	13809	12750
Hypocreales	1017	809	1202	842	1065	1021	1365	15177	4202	5550	1487	3815	2185	96493
Ustilaginales	1350	1111	3142	4762	1021	909	1702	8160	17311	17841	3301	2104	8822	5777
Mycosphaerellales	2575	1889	2781	2336	2050	1011	8838	6901	10339	16757	1480	1502	2913	5999
Saccharomycetales	158	82	98	790	2525	2191	2314	3396	5826	22823	5049	4387	5789	4576
Schizosaccharomycetales	28	21	38	978	2199	2180	1544	2588	6263	20280	5335	5292	5263	4861
Malasseziales	38	31	44	668	1361	1318	3632	2091	4251	16215	3719	3153	3410	2974
Sordariales	150	60	112	452	818	436	839	3155	2592	4062	722	1140	8404	6486
Eurotiales	193	140	192	372	768	756	1517	2578	2846	5946	588	789	1281	2437
Glomerellales	29	54	113	66	100	64	121	267	425	551	101	150	535	802
Eucoccidiorida	28	14	12	62	117	205	91	110	281	1499	175	66	119	83
Magnaporthales	11	9	4	44	58	40	83	259	335	484	80	118	327	687
Haemosporida	72	76	54	70	78	47	94	100	235	388	69	44	248	84
Trypanosomatida	13	20	5	15	70	90	24	68	83	515	71	32	383	15
Naviculales	9	2	4	33	201	260	81	94	58	124	50	37	30	39
Piroplasmida	58	5	10	24	31	218	34	18	28	75	123	35	47	85
Cryptomonadales	3	3	3	21	12	107	15	16	18	54	15	4	11	1

Thalassiosirales	3	0	0	13	56	61	9	25	8	16	11	5	8	2
Dictyosteliales	2	2	2	9	32	17	4	27	30	33	13	7	10	17
Pyrenomonadales	27	5	3	27	3	16	6	1	3	16	11	4	7	4

C)

Genus	**A0707**	**A0708**	**A0709**	**A0922**	**A1118**	**C0311**	**C0717**	**C1002**	**C1016**	**C1022**	**C1028**	**C1103**	**C1110**	**C1119**
multi-affiliation	8471	5156	7758	43533	78468	93514	107143	117603	207562	526422	136112	117946	145480	173072
Botrytis	12048	7059	10949	11892	14141	4707	79213	114759	17036	42040	13670	5616	5594	31766
Cryptococcus	297	193	303	2424	5762	5746	10281	6719	16714	53102	13962	13304	13809	12750
Fusarium	956	759	1123	696	831	761	1178	14747	3441	4558	1255	3368	1670	94938
Schizosaccharomyces	28	21	38	978	2199	2180	1544	2588	6263	20280	5335	5292	5263	4861
Ustilago	504	96	300	4289	361	225	547	7162	14751	13030	1886	849	7240	4506
Malassezia	38	31	44	668	1361	1318	3632	2091	4251	16215	3719	3153	3410	2974
Zymoseptoria	1932	1514	2446	1451	1139	589	4960	3772	5967	6867	890	862	1854	3525
Cercospora	439	247	172	563	694	323	2769	2191	3179	6975	402	438	665	1896
Aspergillus	139	103	137	268	419	458	1071	1973	1564	3468	267	352	773	1332
Sporisorium	678	835	2062	227	347	342	828	551	1304	2330	597	399	635	618
Neurospora	69	36	34	180	240	120	365	1639	1102	1799	266	292	1707	1520
Lachancea	9	3	3	32	311	62	75	281	649	4333	845	728	744	482
Colletotrichum	29	54	113	66	100	64	121	267	425	551	101	150	535	802
Thermothielavioides	23	6	18	59	90	74	72	205	365	452	109	135	643	801
Pyricularia	11	9	4	44	58	40	83	259	335	484	80	118	327	687
Toxoplasma	9	9	5	33	65	132	50	92	249	1439	158	51	101	67
Thermothelomyces	12	7	16	40	66	47	86	173	219	424	58	109	653	543
Pochonia	21	16	24	42	72	68	69	126	435	496	129	163	245	438
Talaromyces	20	12	22	44	72	37	96	234	294	603	57	104	131	395
Plasmodium	72	76	54	70	78	47	94	100	235	388	69	44	248	84
Sugiyamaella	10	4	6	34	107	246	30	55	140	398	95	67	109	252
Pichia	6	12	4	24	102	35	66	149	73	499	48	72	76	110
Naumovozyma	2	1	4	7	28	6	11	19	11	46	13	6	1054	24
Scheffersomyces	0	2	1	5	32	9	7	36	87	589	122	118	119	68
Debaryomyces	0	2	10	21	44	23	11	522	38	398	38	22	16	41
Brettanomyces	0	0	7	36	96	81	28	42	75	459	48	51	85	89
Phaeodactylum	9	2	4	33	201	260	81	94	58	124	50	37	30	39
Leishmania	10	4	1	6	31	21	10	14	26	400	30	10	368	9
Kluyveromyces	15	16	7	34	29	10	112	103	129	226	25	32	48	58

Supplementary Table 7: Taxonomy table for Archaea. A: all phyla, **B**: all orders, and **C**: 30 top genera tables ordered by abundance. Sample name: "A" or "C" for respectively Aerosol or cloud sample, and the date (mm/dd).

A)

Phylum	A0707	A0708	A0709	A0922	A1118	C0311	C0717	C1002	C1016	C1022	C1028	C1103	C1110	C1119
Euryarchaeota	124	98	175	20	530	164	106	697	147	92	79	34	27	71
Thaumarchaeota	65	50	31	9	38	16	23	26	6	15	10	2	3	9
multi-affiliation	1	0	0	1	1	0	0	2	2	1	0	0	2	3
Crenarchaeota	0	0	0	1	0	0	1	0	0	0	1	0	1	0

B)

Order	A0707	A0708	A0709	A0922	A1118	C0311	C0717	C1002	C1016	C1022	C1028	C1103	C1110	C1119
Methanosarcinales	56	76	116	4	363	113	64	19	49	14	52	8	8	7
Methanomicrobiales	4	0	17	3	56	17	3	624	10	1	2	1	0	6
Methanobacteriales	35	15	32	4	96	31	20	9	64	18	14	6	6	43
Nitrososphaerales	59	43	27	4	36	14	19	22	6	7	10	2	3	6
multi-affiliation	13	8	5	4	7	0	8	41	10	44	5	7	14	8
Natrialbales	12	5	3	4	2	2	4	8	7	4	2	5	0	6
Halobacteriales	1	0	3	1	4	1	3	0	5	10	1	3	0	0
Nitrosopumilales	6	1	3	4	1	2	4	0	0	0	0	0	0	2
Haloferacales	2	0	0	0	3	0	4	2	1	5	3	2	0	0
Thermococcales	0	0	0	2	0	0	0	0	3	4	0	2	0	1
Methanocellales	2	0	0	0	1	0	0	0	0	1	0	0	1	2
Sulfolobales	0	0	0	1	0	0	0	0	0	0	1	0	1	0
Methanococcales	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Thermoproteales	0	0	0	0	0	0	1	0	0	0	0	0	0	0

C)

Genus	A0707	A0708	A0709	A0922	A1118	C0311	C0717	C1002	C1016	C1022	C1028	C1103	C1110	C1119
Methanosarcina	46	68	116	3	336	111	64	18	42	13	52	8	8	7
Methanoculleus	2	0	16	2	49	12	3	576	10	0	2	0	0	3
multi-affiliation	32	17	12	4	29	7	18	93	15	36	7	8	17	10
Methanobrevibacter	23	10	28	2	64	28	16	4	42	16	10	6	5	38
Candidatus Nitrosocosmicus	36	16	17	3	24	8	11	11	4	2	10	2	2	3
Methanobacterium	3	0	1	1	22	2	1	5	21	1	0	0	0	3
Methanothrix	9	5	0	1	27	2	0	1	6	1	0	0	0	0
Natrialba	12	5	3	4	1	2	3	6	2	2	2	5	0	5
Nitrososphaera	6	18	5	1	0	3	1	6	1	5	0	0	0	1
Methanosphaera	9	5	3	0	7	1	2	0	0	1	4	0	0	2
Nitrosopumilus	6	1	3	4	0	2	3	0	0	0	0	0	0	2
Methanocorpusculum	1	0	0	1	4	2	0	0	0	0	0	0	0	2
Halalkalicoccus	0	0	0	0	1	0	0	0	0	8	0	0	0	0
Thermococcus	0	0	0	2	0	0	0	0	2	4	0	1	0	0
Candidatus Nitrosotenuis	0	0	0	0	0	0	0	0	0	8	0	0	0	0
Methanocella	2	0	0	0	1	0	0	0	0	1	0	0	1	2
Natrinema	0	0	0	0	0	0	1	0	5	0	0	0	0	0
Halorhabdus	0	0	0	0	0	0	3	0	1	0	0	1	0	0
Haloquadratum	0	0	0	0	0	0	0	0	0	3	0	0	0	0
Halorubrum	0	0	0	0	0	0	2	0	0	1	0	0	0	0
Methanofollis	0	0	0	0	0	0	0	2	0	0	0	0	0	1
Methanomethylovorans	0	3	0	0	0	0	0	0	0	0	0	0	0	0
Methanothermobacter	0	0	0	1	1	0	1	0	0	0	0	0	0	0
Natronomonas	0	0	0	0	1	0	0	0	1	1	0	0	0	0
Pyrococcus	0	0	0	0	0	0	0	0	1	0	0	1	0	1
Halarchaeum	0	0	0	0	0	0	0	0	0	0	0	2	0	0
Halobacterium	0	0	1	0	1	0	0	0	0	0	0	0	0	0
Halobaculum	0	0	0	0	0	0	0	1	0	0	0	1	0	0
Haloferax	0	0	0	0	0	0	0	0	0	0	2	0	0	0
Halolamina	0	0	0	0	1	0	0	1	0	0	0	0	0	0

Supplementary Table 8: Taxonomy table for Viruses. A: all phyla, and B: all orders tables ordered by abundance. Sample name: "A" or "C" for respectively Aerosol or cloud sample, and the date (mm/dd).

A)

Phylum	A0707	A0708	A0709	A0922	A1118	C0311	C0717	C1002	C1016	C1022	C1028	C1103	C1110	C1119
Uroviricota	78	18	6	12	10	11	5	53	35	336	10	3	3	7
multi-affiliation	35	9	10	13	12	8	79	2	7	3	0	2	23	2
Negarnaviricota	37	3	10	51	1	16	23	3	4	3	3	1	5	0
Nucleocytoviricota	31	1	0	3	5	1	3	11	4	5	4	2	4	12
Pisuviricota	0	0	0	1	0	30	0	0	0	0	0	0	0	0
Artverviricota	1	0	0	3	0	0	1	2	4	12	0	0	1	0
Cossaviricota	2	0	0	0	0	0	0	0	0	19	0	0	0	0
Peploviricota	0	0	0	2	0	0	0	0	0	0	2	1	1	2
Cressdnaviricota	0	1	0	0	0	0	0	1	0	0	1	0	0	0
Kitrinoviricota	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Preplasmiviricota	0	0	0	0	0	0	0	0	0	0	0	1	0	0

B)

Order	A0707	A0708	A0709	A0922	A1118	C0311	C0717	C1002	C1016	C1022	C1028	C1103	C1110	C1119
Caudovirales	78	18	6	12	10	11	5	53	35	336	10	3	3	7
Lefavirales	29	6	7	13	10	3	78	1	7	3	0	2	23	2
Bunyavirales	37	3	10	51	1	16	23	3	4	3	3	1	5	0
Pimascovirales	29	0	0	1	0	0	0	2	1	3	1	0	1	0
Algavirales	2	1	0	2	5	0	0	5	3	0	2	2	1	12
Sobelivirales	0	0	0	1	0	30	0	0	0	0	0	0	0	0
Ortervirales	1	0	0	3	0	0	1	2	4	12	0	0	1	0
multi-affiliation	6	3	3	0	2	5	1	1	0	0	0	0	0	0
Sepolyvirales	2	0	0	0	0	0	0	0	0	16	0	0	0	0
Imitervirales	0	0	0	0	0	1	3	4	0	2	1	0	1	0
Herpesvirales	0	0	0	2	0	0	0	0	0	0	2	1	1	2
Geplafuvirales	0	1	0	0	0	0	0	1	0	0	1	0	0	0
Piccovirales	0	0	0	0	0	0	0	0	0	3	0	0	0	0
Chitovirales	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Martellivirales	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Rowavirales	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Supplementary Table 9: Differential expression analysis (DEA) coefficients for the 30 most overexpressed A) bacterial families, and B) bacterial genera, and the total overexpressed C) eukaryotic families, and D) eukaryotic genera. Positive significant results (i.e., coef) from MTXmodel R package. Metadata: feature name; value: reference category for coefficient values; coef: DEA result; stderr: standard deviation of the model; N: total number of data point; N.not.0: total of non-zero data point; pval: p-value from the calculation; qval: computed with the correlation method of the model (p.adjust) (cf R package Maaslin2 protocol).

A)

Bacterial families	metadata	value	coef	stderr	N	N.not.0	pval	qval
Halomonadaceae	MTvsMG	MT	**3.230676**	0.275906	28	28	7.20E-12	1.31E-09
Amoebophilaceae	MTvsMG	MT	**3.133931**	0.414888	28	15	5.10E-08	9.52E-07
Aphanizomenonaceae	MTvsMG	MT	**3.068177**	0.472505	28	11	6.98E-07	8.14E-06
Mycoplasmataceae	MTvsMG	MT	**3.02639**	0.404392	28	15	6.04E-08	1.06E-06
Clostridiaceae	MTvsMG	MT	**3.023235**	0.622843	28	28	4.94E-05	2.02E-04
Fimbriimonadaceae	MTvsMG	MT	**2.961249**	0.255887	28	13	9.33E-12	1.31E-09
Peptoniphilaceae	MTvsMG	MT	**2.578585**	0.462713	28	19	7.46E-06	4.54E-05
Parachlamydiaceae	MTvsMG	MT	**2.534424**	0.3346	28	16	4.85E-08	9.52E-07
Chroococcaceae	MTvsMG	MT	**2.466467**	0.402713	28	15	1.79E-06	1.65E-05
Rickettsiaceae	MTvsMG	MT	**2.429085**	0.311973	28	24	2.93E-08	6.96E-07
Treponemataceae	MTvsMG	MT	**2.13919**	0.464782	28	8	9.60E-05	0.00034
Peptostreptococcaceae	MTvsMG	MT	**2.041419**	0.324678	28	28	1.18E-06	1.27E-05
Anaerolineaceae	MTvsMG	MT	**2.001948**	0.236507	28	6	6.04E-09	2.72E-07
Candidatus.Paracaedibacteraceae	MTvsMG	MT	**1.96274**	0.247476	28	2	2.08E-08	6.47E-07
Ilumatobacteraceae	MTvsMG	MT	**1.917951**	0.232337	28	23	9.77E-09	3.42E-07
Chthonomonadaceae	MTvsMG	MT	**1.898614**	0.316637	28	2	2.49E-06	1.93E-05
Chitinophagaceae	MTvsMG	MT	**1.883672**	0.207437	28	28	1.52E-09	1.06E-07
Blattabacteriaceae	MTvsMG	MT	**1.84606**	0.530111	28	8	0.001774	0.004556
Kofleriaceae	MTvsMG	MT	**1.841809**	0.269426	28	6	2.95E-07	3.76E-06
Pirellulaceae	MTvsMG	MT	**1.825652**	0.248029	28	5	8.13E-08	1.20E-06
Isosphaeraceae	MTvsMG	MT	**1.789886**	0.228027	28	28	2.52E-08	6.96E-07
Bdellovibrionaceae	MTvsMG	MT	**1.746573**	0.229332	28	7	4.39E-08	9.46E-07
Caldilineaceae	MTvsMG	MT	**1.72953**	0.297863	28	3	4.06E-06	2.71E-05
Fulvivirgaceae	MTvsMG	MT	**1.721637**	0.275715	28	11	1.32E-06	1.36E-05
Pasteurellaceae	MTvsMG	MT	**1.710233**	0.867663	28	28	0.059447	0.09734
Trueperaceae	MTvsMG	MT	**1.686917**	0.310577	28	9	1.08E-05	6.05E-05
Verrucomicrobiaceae	MTvsMG	MT	**1.682447**	0.281815	28	16	2.66E-06	1.96E-05

Steroidobacteraceae	MTvsMG	MT	**1.648999**	0.315074	28	11	1.82E-05	9.41E-05
Bryobacteraceae	MTvsMG	MT	**1.539126**	0.316294	28	7	4.78E-05	2.00E-04
Turicibacteraceae	MTvsMG	MT	**1.533587**	0.335103	28	21	1.03E-04	0.000355

1555 **B)**

Bacterial genera	**metadata**	**value**	**coef**	**stderr**	**N**	**N.not.0**	**pval**	**qval**
Dolichospermum	MTvsMG	MT	**4.30682105**	0.51877724	28	14	8.77E-09	2.53E-07
Moorea	MTvsMG	MT	**4.0459583**	0.48703863	28	16	8.66E-09	2.53E-07
Anaerococcus	MTvsMG	MT	**3.68256226**	0.41419254	28	16	2.31E-09	9.17E-08
Halomonas	MTvsMG	MT	**3.62753226**	0.27767805	28	28	6.24E-13	2.03E-10
Gloeocapsa	MTvsMG	MT	**3.5706384**	0.40993148	28	16	3.46E-09	1.19E-07
Mycoplasma	MTvsMG	MT	**3.44397578**	0.40421629	28	15	5.32E-09	1.74E-07
Clostridium	MTvsMG	MT	**3.39896344**	0.62999297	28	28	1.19E-05	8.14E-05
Fimbriimonas	MTvsMG	MT	**3.32538924**	0.25680598	28	14	7.62E-13	2.03E-10
Tuwongella	MTvsMG	MT	**3.31291704**	0.26587802	28	15	1.81E-12	3.26E-10
Delftia	MTvsMG	MT	**3.26032098**	0.47131396	28	27	2.41E-07	3.16E-06
Acidibrevibacterium	MTvsMG	MT	**3.14109508**	0.39123899	28	22	1.65E-08	3.80E-07
Candidatus Amoebophilus	MTvsMG	MT	**3.07132913**	0.39700267	28	13	3.30E-08	6.25E-07
Singulisphaera	MTvsMG	MT	**2.98894508**	0.23191731	28	25	8.48E-13	2.03E-10
Neokomagataea	MTvsMG	MT	**2.82211676**	0.27600085	28	11	1.33E-10	1.37E-08
Candidatus Xiphinematobacter	MTvsMG	MT	**2.71565**	0.28282087	28	13	4.90E-10	3.21E-08
Candidatus Sulcia	MTvsMG	MT	**2.67879985**	0.6502084	28	12	0.00034175	0.00130192
Rickettsia	MTvsMG	MT	**2.59450113**	0.38983351	28	24	4.64E-07	5.30E-06
Candidatus Protochlamydia	MTvsMG	MT	**2.53968815**	0.28706198	28	16	2.55E-09	9.17E-08
Dyadobacter	MTvsMG	MT	**2.53813715**	0.26673229	28	21	5.90E-10	3.54E-08
Treponema	MTvsMG	MT	**2.5033301**	0.46677955	28	13	1.29E-05	8.78E-05
Isosphaera	MTvsMG	MT	**2.46905451**	0.29689115	28	12	8.48E-09	2.53E-07
Romboutsia	MTvsMG	MT	**2.39288673**	0.36538258	28	27	6.07E-07	6.72E-06
Pasteurella	MTvsMG	MT	**2.34385252**	0.91337445	28	28	0.01639437	0.03451446
Flavisolibacter	MTvsMG	MT	**2.33417503**	0.25153747	28	24	9.82E-10	4.95E-08
Candidatus Paracaedibacter	MTvsMG	MT	**2.32687985**	0.25100376	28	8	1.00E-09	4.95E-08
Candidatus Hodgkinia	MTvsMG	MT	**2.29901283**	0.44925587	28	8	2.47E-05	0.00014674
Neochlamydia	MTvsMG	MT	**2.29764761**	0.25627345	28	9	1.96E-09	8.29E-08
Ilumatobacter	MTvsMG	MT	**2.28209118**	0.23326798	28	26	3.34E-10	2.67E-08
Chthonomonas	MTvsMG	MT	**2.26275456**	0.31895464	28	9	1.56E-07	2.32E-06
Verrucomicrobium	MTvsMG	MT	**2.25708487**	0.28851087	28	13	2.68E-08	5.37E-07

C)

Eukaryotic families	metadata	value	coef	stderr	N	N.not.0	pval	qval
Hemiselmidaceae	MTvsMG	MT	**1.68979084**	0.402199192	28	0	0.000276045	0.001301355
Geminigeraceae	MTvsMG	MT	**1.59734011**	0.636994627	28	5	0.018731473	0.051511552
Phaeodactylaceae	MTvsMG	MT	**1.30680215**	0.284335264	28	9	9.77E-05	0.000537079
Cryptomonadaceae	MTvsMG	MT	**1.293962126**	0.335868138	28	1	0.000685879	0.002829249
Plasmodiidae	MTvsMG	MT	**1.06151257**	0.392590496	28	13	0.011922463	0.036777513
Cryptosporidiidae	MTvsMG	MT	**0.666151612**	0.313493456	28	0	0.043259531	0.083974384
Saccharomycetaceae	MTvsMG	MT	**0.662918094**	0.246260453	28	27	0.012259171	0.036777513
Malasseziaceae	MTvsMG	MT	**0.66051082**	0.399032776	28	28	0.109890526	0.161721858
Trypanosomatidae	MTvsMG	MT	**0.658810928**	0.271343097	28	5	0.022409143	0.056884748
Cryptococcaceae	MTvsMG	MT	**0.606113098**	0.344357018	28	28	0.090147457	0.14166029
Trichocomaceae	MTvsMG	MT	**0.570264142**	0.247248549	28	8	0.029312384	0.069093476
Mycosphaerellaceae	MTvsMG	MT	**0.532738382**	0.259717797	28	28	0.050447758	0.085143385
Clavicipitaceae	MTvsMG	MT	**0.501731245**	0.234945595	28	11	0.042305569	0.083974384
Trichomonascaceae	MTvsMG	MT	**0.453910183**	0.302785142	28	8	0.145888589	0.185166285
Sarcocystidae	MTvsMG	MT	**0.38734229**	0.26952294	28	5	0.162604409	0.198738722
Debaryomycetaceae	MTvsMG	MT	**0.373328036**	0.247651297	28	12	0.143744641	0.185166285
multi.affiliation	MTvsMG	MT	**0.350659941**	0.213608389	28	28	0.112715234	0.161721858
Dictyosteliaceae	MTvsMG	MT	**0.281970461**	0.176269358	28	0	0.121757528	0.167416601

D)

Eukaryotic genera	metadata	value	coef	stderr	N	N.not.0	pval	qval
Kluyveromyces	MTvsMG	MT	**2.67901299**	0.23995166	28	20	2.04E-11	5.50E-10
Hemiselmis	MTvsMG	MT	**2.43047699**	0.40269462	28	7	2.25E-06	7.14E-06
Guillardia	MTvsMG	MT	**2.33802626**	0.62882404	28	7	0.00097088	0.00194176
Tetrapisispora	MTvsMG	MT	**2.20465389**	0.23107838	28	13	5.59E-10	3.54E-09
Phaeodactylum	MTvsMG	MT	**2.0474883**	0.31286997	28	19	6.14E-07	2.37E-06
Cryptomonas	MTvsMG	MT	**2.03464827**	0.35460617	28	9	4.85E-06	1.38E-05
Candida	MTvsMG	MT	**1.91867063**	0.26780039	28	13	1.31E-07	5.45E-07
Plasmodium	MTvsMG	MT	**1.80219872**	0.36958576	28	28	4.66E-05	0.00010937
Zymoseptoria	MTvsMG	MT	**1.6058694**	0.26368334	28	28	1.95E-06	6.59E-06
Cryptosporidium	MTvsMG	MT	**1.40683776**	0.32998073	28	6	0.00023456	0.00050666
Malassezia	MTvsMG	MT	**1.40119697**	0.42371934	28	28	0.00276053	0.00451723
Cryptococcus	MTvsMG	MT	**1.34679925**	0.37015643	28	28	0.00119129	0.00227899
Talaromyces	MTvsMG	MT	**1.31095029**	0.24796003	28	27	1.58E-05	4.06E-05
Pochonia	MTvsMG	MT	**1.24241739**	0.23384971	28	28	1.47E-05	3.98E-05
Thermothielavioides	MTvsMG	MT	**1.22470645**	0.24812009	28	27	3.98E-05	9.77E-05
Bigelowiella	MTvsMG	MT	**1.2055932**	0.31420623	28	0	0.00071425	0.00148344
Sugiyamaella	MTvsMG	MT	**1.19459633**	0.32927863	28	20	0.0012239	0.00227899
Thalassiosira	MTvsMG	MT	**1.17175453**	0.39363882	28	5	0.00622713	0.00862218
Babesia	MTvsMG	MT	**1.12266154**	0.36902136	28	8	0.00531018	0.00754605
Brettanomyces	MTvsMG	MT	**1.09336707**	0.33501536	28	18	0.00307561	0.00488478
multi.affiliation	MTvsMG	MT	**1.0827933**	0.23267724	28	28	8.39E-05	0.0001887
Dictyostelium	MTvsMG	MT	**1.02265661**	0.17252629	28	4	2.97E-06	8.91E-06
Scheffersomyces	MTvsMG	MT	**0.953537**	0.43015627	28	12	0.03559612	0.0447021
Colletotrichum	MTvsMG	MT	**0.94084593**	0.29864293	28	28	0.00407284	0.0062838
Theileria	MTvsMG	MT	**0.93819934**	0.41369431	28	6	0.03187975	0.04098824
Toxoplasma	MTvsMG	MT	**0.8686916**	0.30124765	28	24	0.00779158	0.01051863
Lachancea	MTvsMG	MT	**0.7741268**	0.42094225	28	25	0.07735841	0.09402603
Eremothecium	MTvsMG	MT	**0.49714426**	0.36309261	28	3	0.18265105	0.20128891
Aspergillus	MTvsMG	MT	**0.44472678**	0.1436014	28	28	0.00464523	0.00690726
Sporisorium	MTvsMG	MT	**0.42645313**	0.33339161	28	28	0.21214798	0.22911982

Supplementary Table 10: The 40 most overrepresented Gene Ontologies (GOs) for each GO categories (Cellular Component, Biological process, and Molecular Function) in both clouds and aerosols. Mean by GO of the coefficients from the differential expression analysis (DEA) for the overexpressed genes.

TermID	Cellular Component	Mean genes coef
GO:0030446	hyphal cell wall	4.09234329
GO:0030445	yeast-form cell wall	4.09234329
GO:0031676	plasma membrane-derived thylakoid membrane	3.12354017
GO:0009523	photosystem II	3.12354017
GO:0001411	hyphal tip	2.46164852
GO:0032126	eisosome	2.40143444
GO:0036286	eisosome filament	2.40143444
GO:0071595	Nem1-Spo7 phosphatase complex	2.38413503
GO:0140220	pathogen-containing vacuole	2.37594961
GO:0000328	fungal-type vacuole lumen	2.14847845
GO:0005834	heterotrimeric G-protein complex	2.14684
GO:0019005	SCF ubiquitin ligase complex	2.02516667
GO:0009279	cell outer membrane	1.95551869
GO:0005856	cytoskeleton	1.9279454
GO:1990077	primosome complex	1.70706328
GO:1990317	Gin4 complex	1.7036205
GO:0032174	cellular bud neck septin collar	1.7036205
GO:0000786	nucleosome	1.68976432
GO:0009986	cell surface	1.58127902
GO:0022626	cytosolic ribosome	1.57362941
GO:0072324	ascus epiplasm	1.56368356
GO:1990063	Bam protein complex	1.55007415
GO:0009507	chloroplast	1.52822961
GO:0031298	replication fork protection complex	1.4157704
GO:0000421	autophagosome membrane	1.40046374
GO:0005753	mitochondrial proton-transporting ATP synthase complex	1.39707298
GO:0005754	mitochondrial proton-transporting ATP synthase, catalytic core	1.38639557
GO:0009522	photosystem I	1.38231899
GO:0062040	fungal biofilm matrix	1.35802281
GO:0005811	lipid droplet	1.34621614
GO:0022627	cytosolic small ribosomal subunit	1.33936676
GO:0033202	DNA helicase complex	1.32510901
GO:0042788	polysomal ribosome	1.32180567
GO:0009277	fungal-type cell wall	1.32143145
GO:0045261	proton-transporting ATP synthase complex, catalytic core F(1)	1.29338131
GO:0005615	extracellular space	1.28906619
GO:0022625	cytosolic large ribosomal subunit	1.26847096
GO:0000138	Golgi trans cisterna	1.25476492
GO:0005775	vacuolar lumen	1.24364637
GO:0009535	chloroplast thylakoid membrane	1.22723367

TermID	Biological process	Mean genes coef
GO:0051701	biological process involved in interaction with host	4.09234329
GO:0009635	response to herbicide	3.12354017
GO:0009772	photosynthetic electron transport in photosystem II	3.12354017
GO:0010827	regulation of glucose transmembrane transport	2.89048939
GO:0009749	response to glucose	2.89048939
GO:0009416	response to light stimulus	2.82320827
GO:0007602	phototransduction	2.82320827
GO:0006520	cellular amino acid metabolic process	2.49690164
GO:0051017	actin filament bundle assembly	2.43534853
GO:0097446	protein localization to eisosome filament	2.43237531
GO:0070941	eisosome assembly	2.40143444
GO:0071071	regulation of phospholipid biosynthetic process	2.38413503
GO:1903740	positive regulation of phosphatidate phosphatase activity	2.38413503
GO:0046889	positive regulation of lipid biosynthetic process	2.38413503
GO:0071072	negative regulation of phospholipid biosynthetic process	2.38413503
GO:0006998	nuclear envelope organization	2.38413503
GO:0140042	lipid droplet formation	2.38413503
GO:0034613	cellular protein localization	2.38413503
GO:0051489	regulation of filopodium assembly	2.37594961
GO:0006616	SRP-dependent cotranslational protein targeting to membrane, translocation	2.32136114
GO:0045903	positive regulation of translational fidelity	2.29851619
GO:0007039	protein catabolic process in the vacuole	2.14847845
GO:0000425	pexophagy	2.14847845
GO:0006796	phosphate-containing compound metabolic process	2.0336122
GO:0034220	ion transmembrane transport	2.03311582
GO:0051453	regulation of intracellular pH	2.03311582
GO:0046898	response to cycloheximide	2.0223878
GO:0006469	negative regulation of protein kinase activity	1.9975986
GO:1990961	xenobiotic detoxification by transmembrane export across the plasma membrane	1.9701832
GO:0032220	plasma membrane fusion involved in cytogamy	1.91951324
GO:0006425	glutaminyl-tRNA aminoacylation	1.90260069
GO:0006424	glutamyl-tRNA aminoacylation	1.90260069
GO:0120029	proton export across plasma membrane	1.879608
GO:0023052	signaling	1.87629647
GO:0044773	mitotic DNA damage checkpoint signaling	1.87629647
GO:0044836	D-xylose fermentation	1.8734787
GO:0006542	glutamine biosynthetic process	1.81685427
GO:0042621	poly(3-hydroxyalkanoate) biosynthetic process	1.80794899
GO:0001933	negative regulation of protein phosphorylation	1.77232125
GO:0015918	sterol transport	1.7519088

TermID	Molecular function	Mean genes coef
GO:0045156	electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity	3.123540165
GO:0010242	oxygen evolving activity	3.123540165
GO:0016682	oxidoreductase activity, acting on diphenols and related substances as donors, oxygen as acceptor	3.123540165
GO:0005503	all-trans retinal binding	2.823208266
GO:0005216	ion channel activity	2.823208266
GO:0016843	amine-lyase activity	2.496901644
GO:0019901	protein kinase binding	2.435348525
GO:0019003	GDP binding	2.435348525
GO:0090729	toxin activity	2.425676381
GO:0001716	L-amino-acid oxidase activity	2.425676381
GO:0106307	protein threonine phosphatase activity	2.384135027
GO:0106306	protein serine phosphatase activity	2.384135027
GO:0070181	small ribosomal subunit rRNA binding	2.298516186
GO:0004568	chitinase activity	2.162140777
GO:0031683	G-protein beta/gamma-subunit complex binding	2.146840002
GO:0001664	G protein-coupled receptor binding	2.146840002
GO:0016791	phosphatase activity	2.069200511
GO:0004427	inorganic diphosphatase activity	2.033612204
GO:0015662	P-type ion transporter activity	2.033115815
GO:0016810	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	2.02345271
GO:0046555	acetylxylan esterase activity	2.02345271
GO:0004022	alcohol dehydrogenase (NAD+) activity	1.991665655
GO:0004819	glutamine-tRNA ligase activity	1.902600693
GO:0008115	sarcosine oxidase activity	1.884203868
GO:0008553	P-type proton-exporting transporter activity	1.879607999
GO:0004683	calmodulin-dependent protein kinase activity	1.876296469
GO:0016881	acid-amino acid ligase activity	1.821249194
GO:0016168	chlorophyll binding	1.817624282
GO:0004356	glutamate-ammonia ligase activity	1.816854269
GO:0016018	cyclosporin A binding	1.791846052
GO:0003746	translation elongation factor activity	1.782249607
GO:0032934	sterol binding	1.751908795
GO:0047456	2-methylisocitrate dehydratase activity	1.741848507
GO:0036381	pyridoxal 5'-phosphate synthase (glutamine hydrolysing) activity	1.699607131
GO:0004365	glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity	1.655056472
GO:0046982	protein heterodimerization activity	1.647572653
GO:0003987	acetate-CoA ligase activity	1.640308663
GO:0004347	glucose-6-phosphate isomerase activity	1.634498939
GO:0004826	phenylalanine-tRNA ligase activity	1.632439873
GO:0019863	IgE binding	1.628578411

1569
1570

Supplementary Table 11: The 20 most overrepresented Gene Ontologies (GOs) for each GO categories (Cellular Component, Biological process, and Molecular Function) in clouds (positive coefficients) or in aerosols (negative coefficients). Mean by GO of the coefficient from the differential expression analysis (DEA) for genes in clouds versus aerosols.

TermID	Cellular Component	Mean genes coef
GO:0005751	mitochondrial respiratory chain complex IV	2.45942143
GO:0005839	proteasome core complex	2.44731466
GO:0019774	proteasome core complex, beta-subunit complex	2.44731466
GO:0005635	nuclear envelope	2.44731466
GO:0000328	fungal-type vacuole lumen	2.13799272
GO:0005747	mitochondrial respiratory chain complex I	2.06611833
GO:0005887	integral component of plasma membrane	2.02944529
GO:0019005	SCF ubiquitin ligase complex	1.90529973
GO:0030659	cytoplasmic vesicle membrane	1.84299433
GO:0030176	integral component of endoplasmic reticulum membrane	1.84299433
GO:0005761	mitochondrial ribosome	1.8349094
GO:0005685	U1 snRNP	1.78556641
GO:0000324	fungal-type vacuole	1.78199138
GO:0005730	nucleolus	1.73975157
GO:0005615	extracellular space	1.70021788
GO:0045275	respiratory chain complex III	1.6919776
GO:0031298	replication fork protection complex	1.6516643
GO:0005777	peroxisome	1.64527919
GO:0009514	glyoxysome	1.63315694
GO:0005934	cellular bud tip	1.63230674
GO:0005886	plasma membrane	-0.35767675
GO:0005840	ribosome	-0.87637983
GO:0033116	endoplasmic reticulum-Golgi intermediate compartment membrane	-1.23951263
GO:0030134	COPII-coated ER to Golgi transport vesicle	-1.23951263
GO:0005905	clathrin-coated pit	-1.26341198
GO:0030136	clathrin-coated vesicle	-1.26341198
GO:0030479	actin cortical patch	-1.26341198
GO:0030687	preribosome, large subunit precursor	-1.26768186
GO:0005694	chromosome	-1.46309657
GO:0045239	tricarboxylic acid cycle enzyme complex	-1.58487927
GO:0016602	CCAAT-binding factor complex	-1.61458721
GO:0019013	viral nucleocapsid	-1.63651833
GO:0020002	host cell plasma membrane	-1.63651833
GO:0009316	3-isopropylmalate dehydratase complex	-1.64879015
GO:0009326	formate dehydrogenase complex	-1.70415855
GO:1990077	primosome complex	-1.92041159
GO:0015935	small ribosomal subunit	-2.16974652
GO:0009279	cell outer membrane	-2.36655281
GO:0015934	large ribosomal subunit	-2.81651614
GO:0009317	acetyl-CoA carboxylase complex	-3.30779414

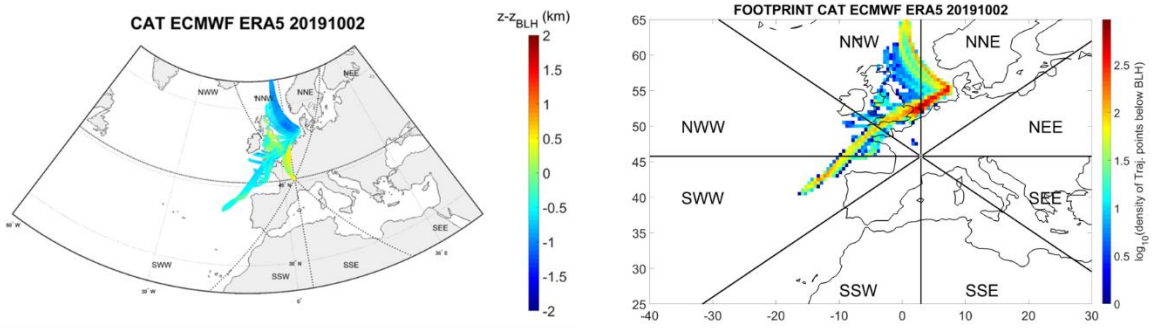
TermID	Biological process	Mean genes coef
GO:0016226	iron-sulfur cluster assembly	2.47847402
GO:0097428	protein maturation by iron-sulfur cluster transfer	2.47847402
GO:0009102	biotin biosynthetic process	2.47847402
GO:0106035	protein maturation by [4Fe-4S] cluster transfer	2.47847402
GO:0010827	regulation of glucose transmembrane transport	2.44878371
GO:0007124	pseudohyphal growth	2.44878371
GO:0009749	response to glucose	2.44878371
GO:0010498	proteasomal protein catabolic process	2.44731466
GO:0045842	positive regulation of mitotic metaphase/anaphase transition	2.44731466
GO:0031507	heterochromatin assembly	2.34459385
GO:0019427	acetyl-CoA biosynthetic process from acetate	2.25438831
GO:0045493	xylan catabolic process	2.16794316
GO:1990961	xenobiotic detoxification by transmembrane export across the plasma membrane	2.1445394
GO:0009267	cellular response to starvation	2.13799272
GO:0000425	pexophagy	2.13799272
GO:0007039	protein catabolic process in the vacuole	2.13799272
GO:0022900	electron transport chain	2.06611833
GO:0000028	ribosomal small subunit assembly	1.92063455
GO:0006123	mitochondrial electron transport, cytochrome c to oxygen	1.91669405
GO:0051321	meiotic cell cycle	1.90529973
GO:0051301	cell division	-2.36655281
GO:0007049	cell cycle	-2.36655281
GO:0006402	mRNA catabolic process	-2.3883259
GO:0006396	RNA processing	-2.3883259
GO:0006090	pyruvate metabolic process	-2.40092613
GO:0006435	threonyl-tRNA aminoacylation	-2.40979398
GO:0001514	selenocysteine incorporation	-2.42098397
GO:0016259	selenocysteine metabolic process	-2.42098397
GO:0065002	intracellular protein transmembrane transport	-2.42548751
GO:0006605	protein targeting	-2.42548751
GO:0017038	protein import	-2.42548751
GO:0006098	pentose-phosphate shunt	-2.47670667
GO:0009432	SOS response	-2.51078768
GO:0006426	glycyl-tRNA aminoacylation	-2.52926533
GO:0071897	DNA biosynthetic process	-2.63178511
GO:0006260	DNA replication	-2.63178511
GO:0019521	D-gluconate metabolic process	-2.81648927
GO:0030979	alpha-glucan biosynthetic process	-2.91083233
GO:0009088	threonine biosynthetic process	-3.02702115
GO:0006400	tRNA modification	-3.13783851

TermID	Molecular function	Mean genes coef
GO:0051536	iron-sulfur cluster binding	2.47847402
GO:0004298	threonine-type endopeptidase activity	2.44731466
GO:0016208	AMP binding	2.25438831
GO:0003987	acetate-CoA ligase activity	2.25438831
GO:0030248	cellulose binding	2.16794316
GO:0046555	acetylxylan esterase activity	2.16794316
GO:0016810	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	2.16794316
GO:0031176	endo-1,4-beta-xylanase activity	2.16794316
GO:0001716	L-amino-acid oxidase activity	2.1512539
GO:0016651	oxidoreductase activity, acting on NAD(P)H	2.06611833
GO:1901691	proton binding	2.02944529
GO:0004586	ornithine decarboxylase activity	1.79714674
GO:0030619	U1 snRNA binding	1.78556641
GO:0070577	lysine-acetylated histone binding	1.7675845
GO:0031493	nucleosomal histone binding	1.7675845
GO:0004099	chitin deacetylase activity	1.70769189
GO:0008121	ubiquinol-cytochrome-c reductase activity	1.6919776
GO:0004315	3-oxoacyl-[acyl-carrier-protein] synthase activity	1.68840942
GO:0031177	phosphopantetheine binding	1.68840942
GO:0140359	ABC-type transporter activity	1.67860068
GO:0004829	threonine-tRNA ligase activity	-2.40979398
GO:0035368	selenocysteine insertion sequence binding	-2.42098397
GO:0008564	protein-exporting ATPase activity	-2.42548751
GO:0004133	glycogen debranching enzyme activity	-2.43579599
GO:0003684	damaged DNA binding	-2.43665884
GO:0004176	ATP-dependent peptidase activity	-2.50808636
GO:0003697	single-stranded DNA binding	-2.51078768
GO:0008094	ATP-dependent activity, acting on DNA	-2.51078768
GO:0004820	glycine-tRNA ligase activity	-2.52926533
GO:0050897	cobalt ion binding	-2.63178511
GO:0031419	cobalamin binding	-2.63178511
GO:0004748	ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor	-2.63178511
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	-2.67331416
GO:0004450	isocitrate dehydrogenase (NADP+) activity	-2.73088306
GO:0016740	transferase activity	-2.81651614
GO:0016758	hexosyltransferase activity	-2.91083233
GO:0004795	threonine synthase activity	-3.02702115
GO:0035596	methylthiotransferase activity	-3.13783851
GO:0004658	propionyl-CoA carboxylase activity	-3.30779414
GO:0003989	acetyl-CoA carboxylase activity	-3.30779414

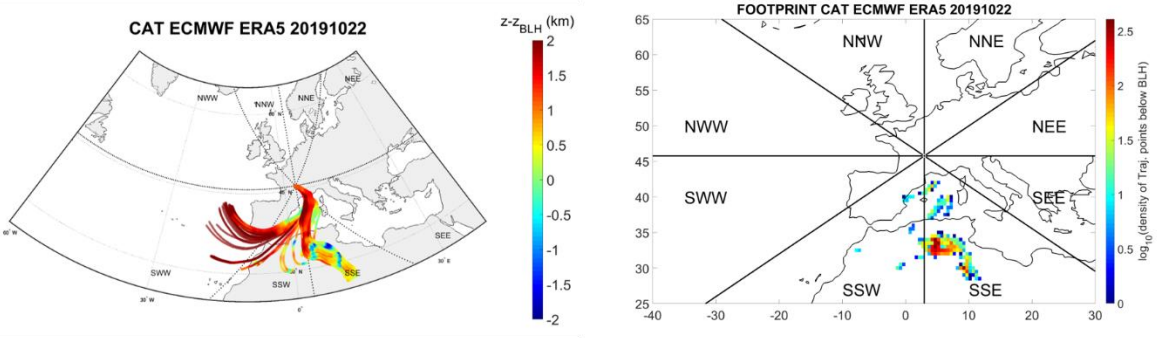
1579

Supplementary Figure 1: Seventy-two-hours backward trajectory plots and associated sources areas for each sampling event, extracted from ERA5 data analysis.

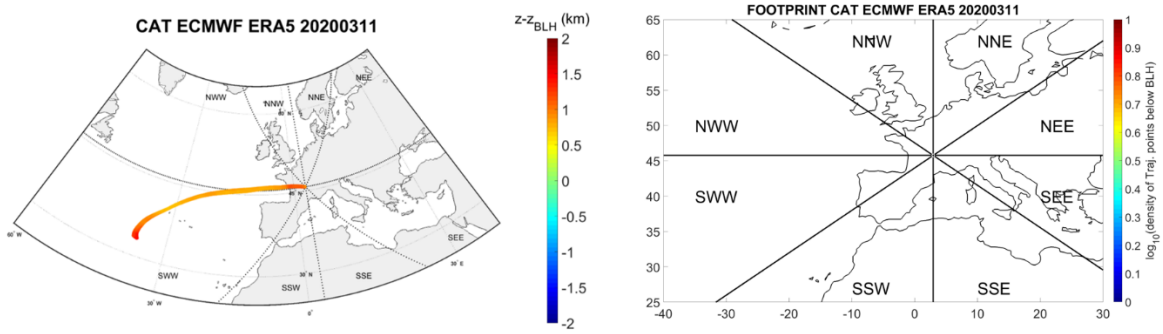
CLOUD20191002



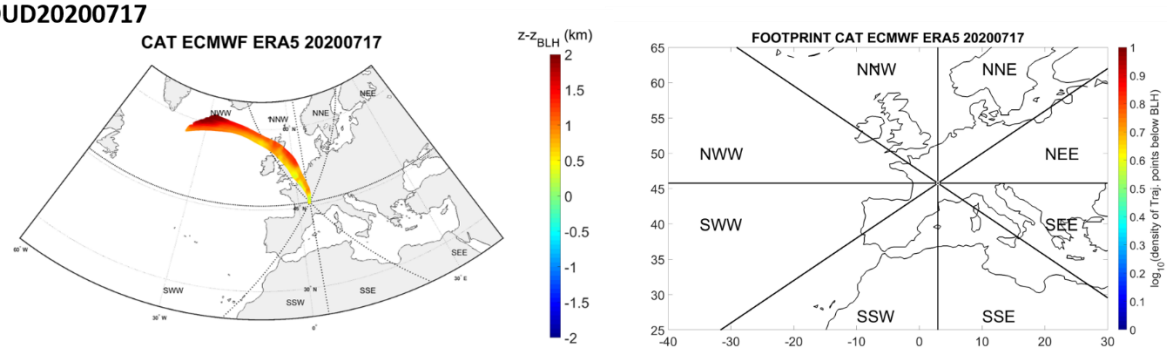
CLOUD20191022



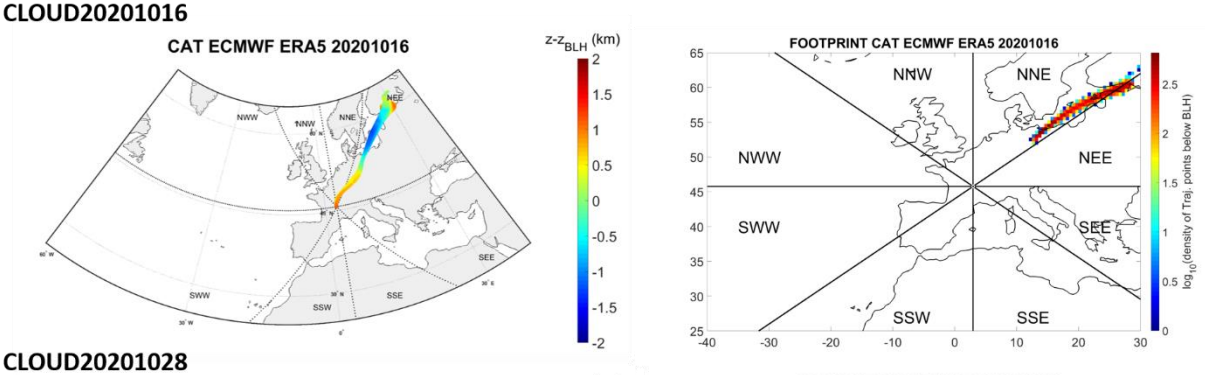
CLOUD20200311



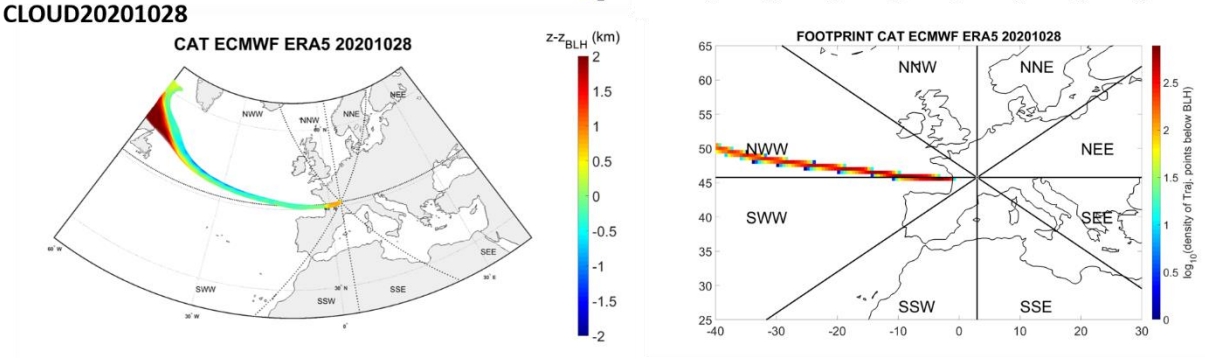
CLOUD20200717



CLOUD20201016

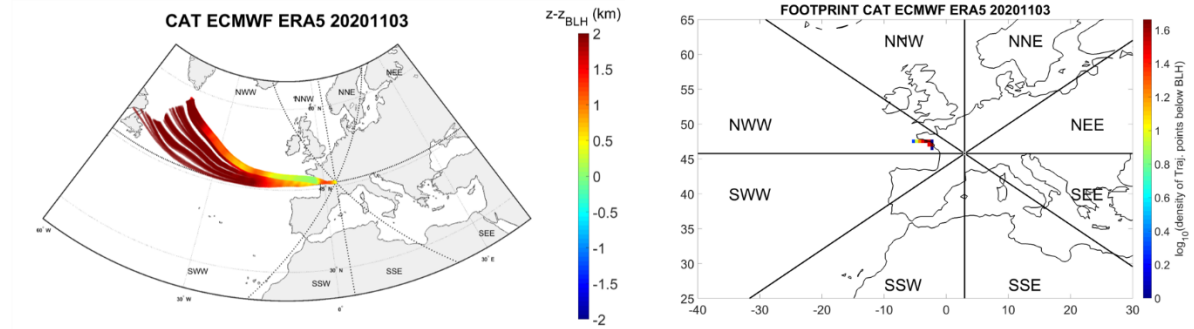


CLOUD20201028

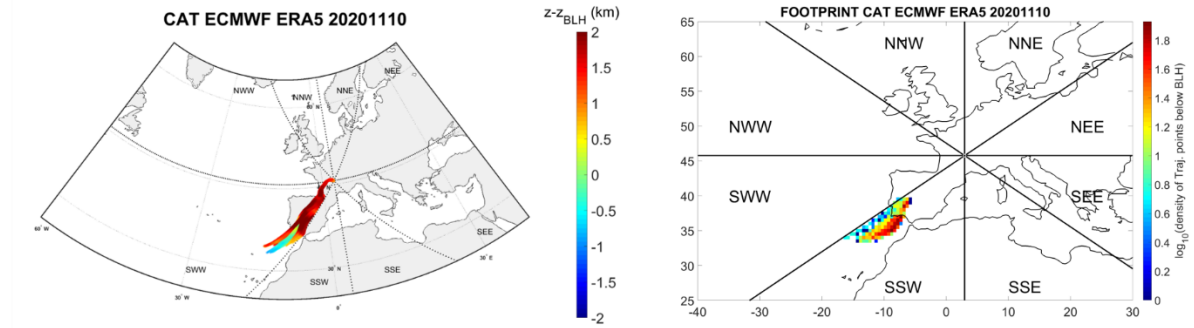


1584

CLOUD20201103

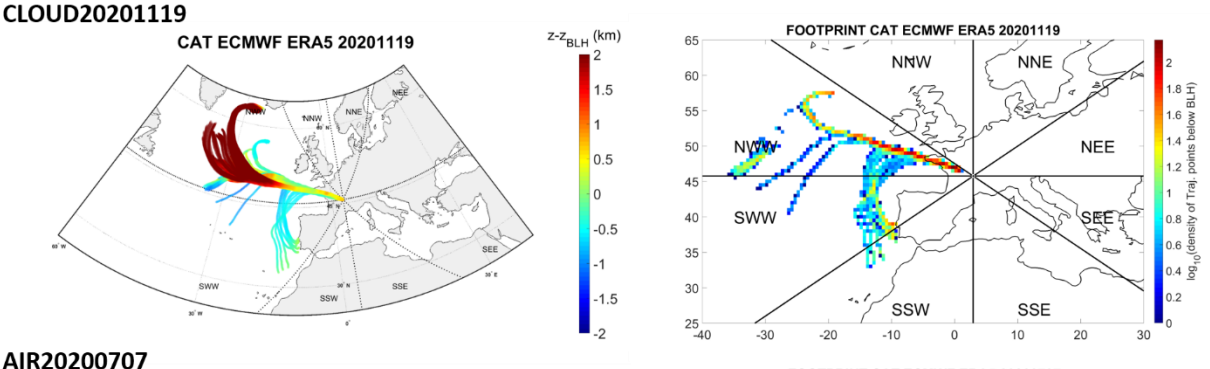


CLOUD20201110

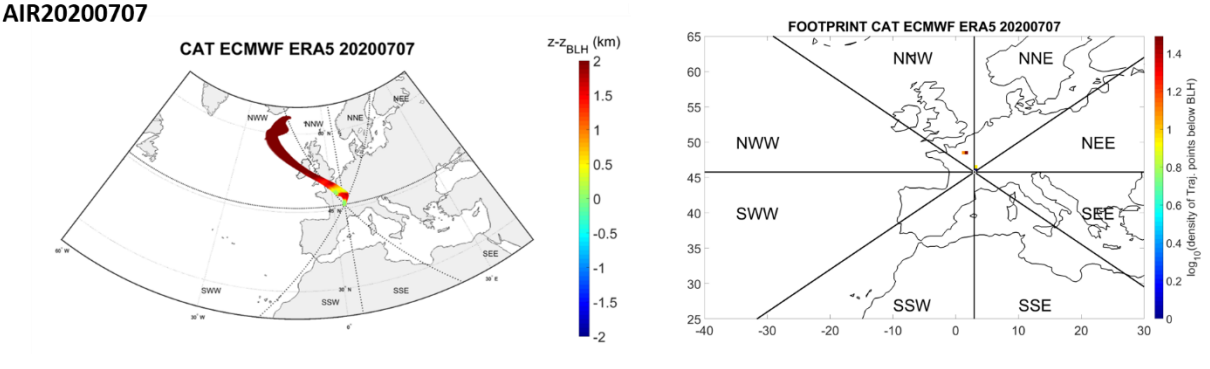


1585

CLOUD20201119

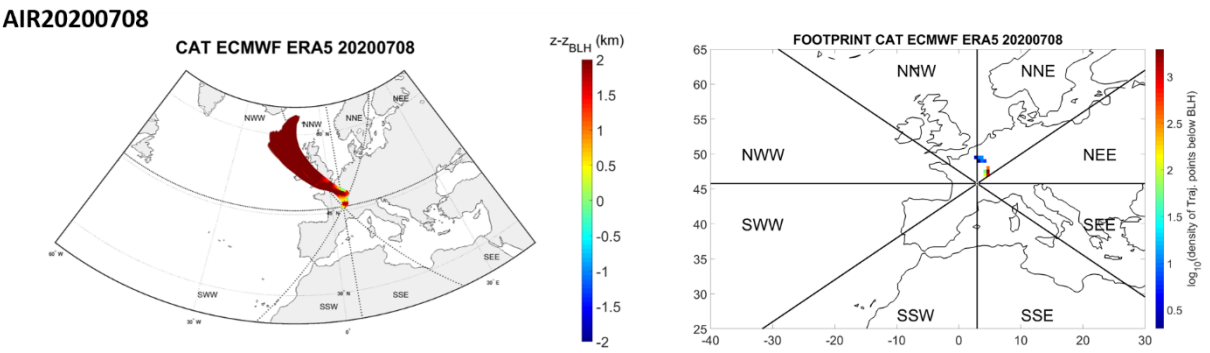


AIR20200707

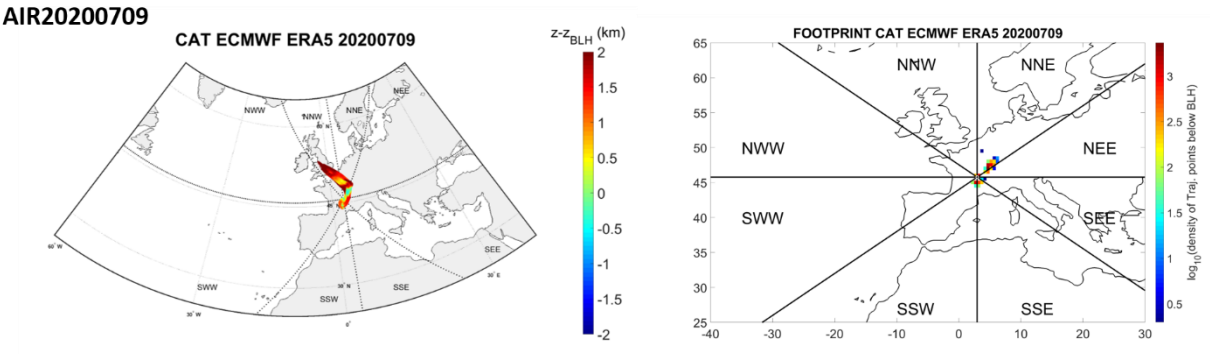


1586

AIR20200708

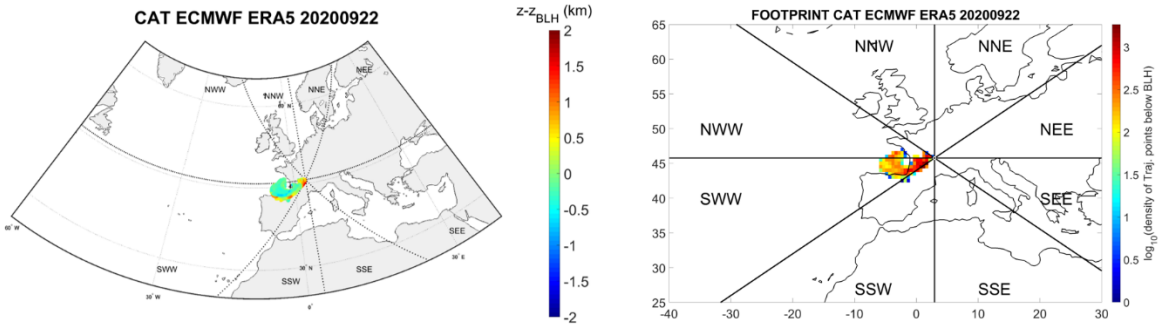


AIR20200709

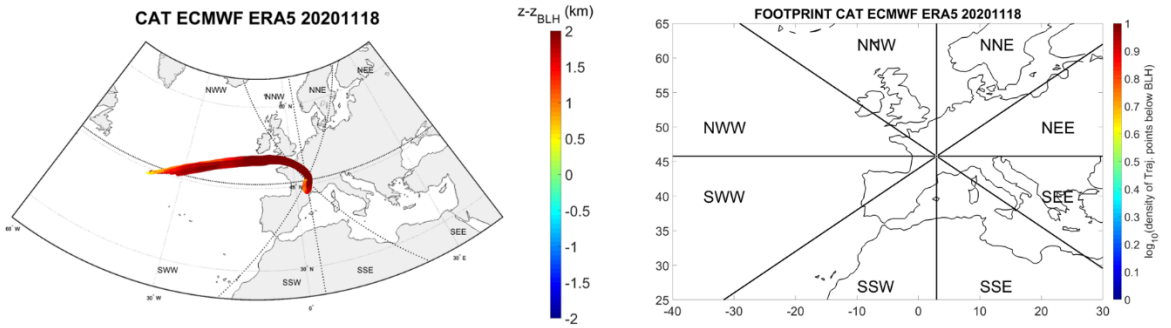


1587

AIR20200922

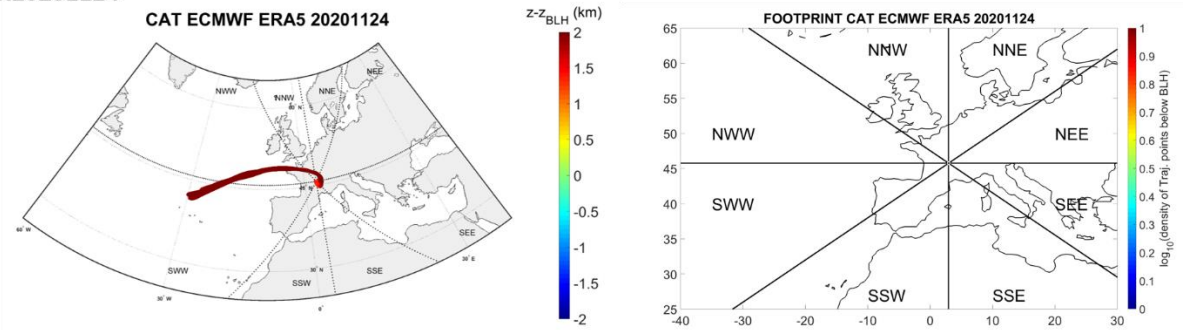


AIR20201118

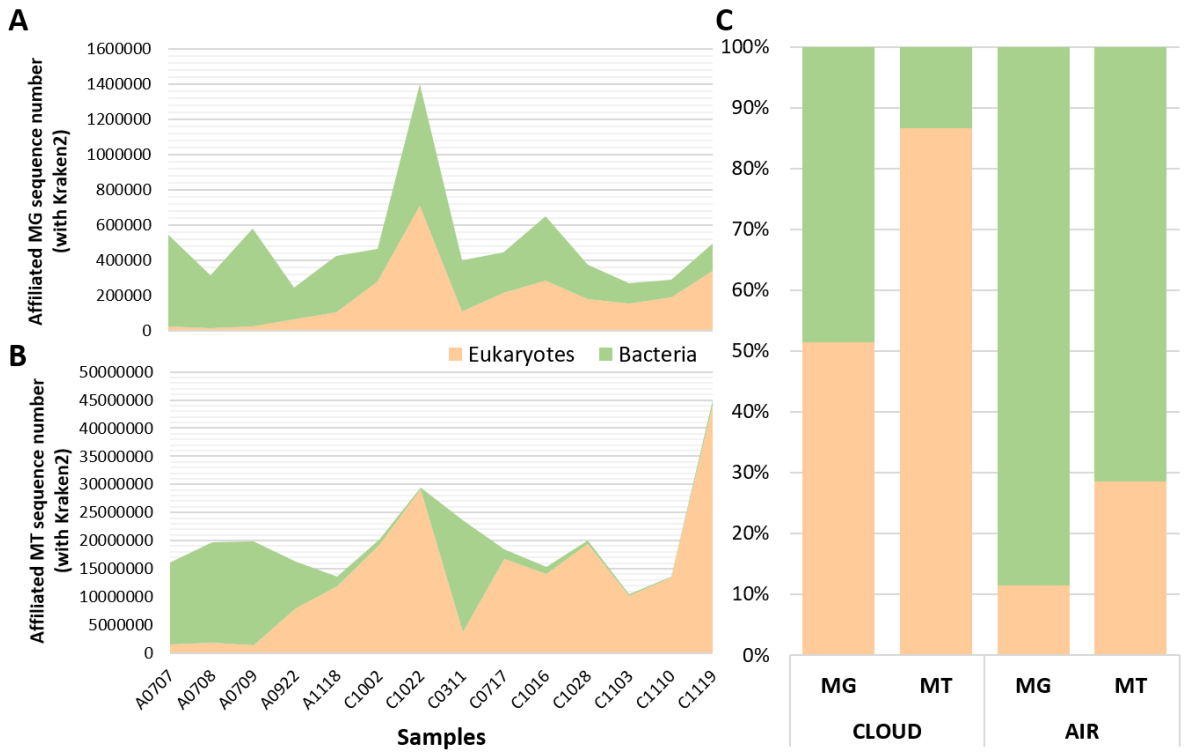


1588

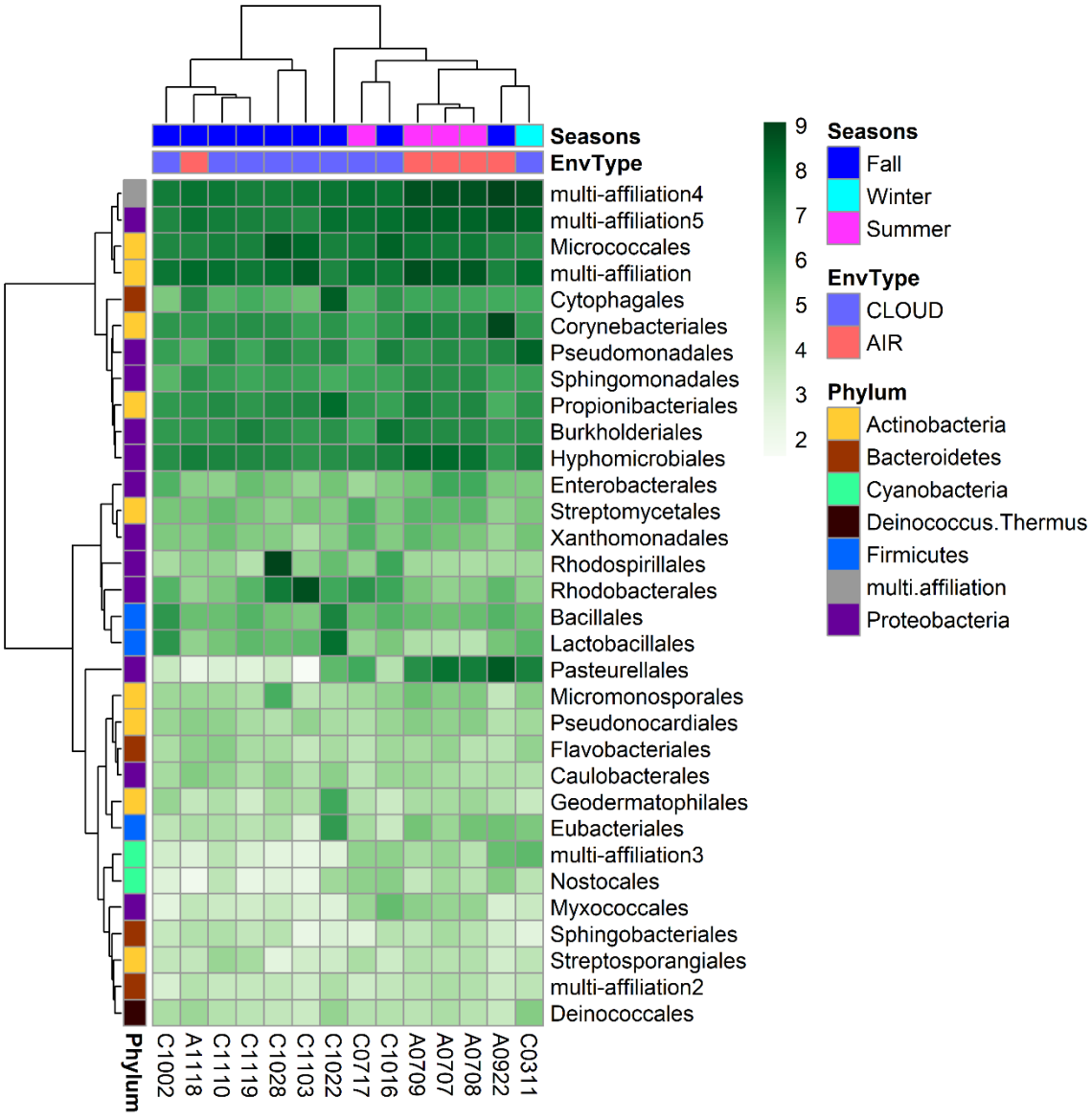
AIR20201124



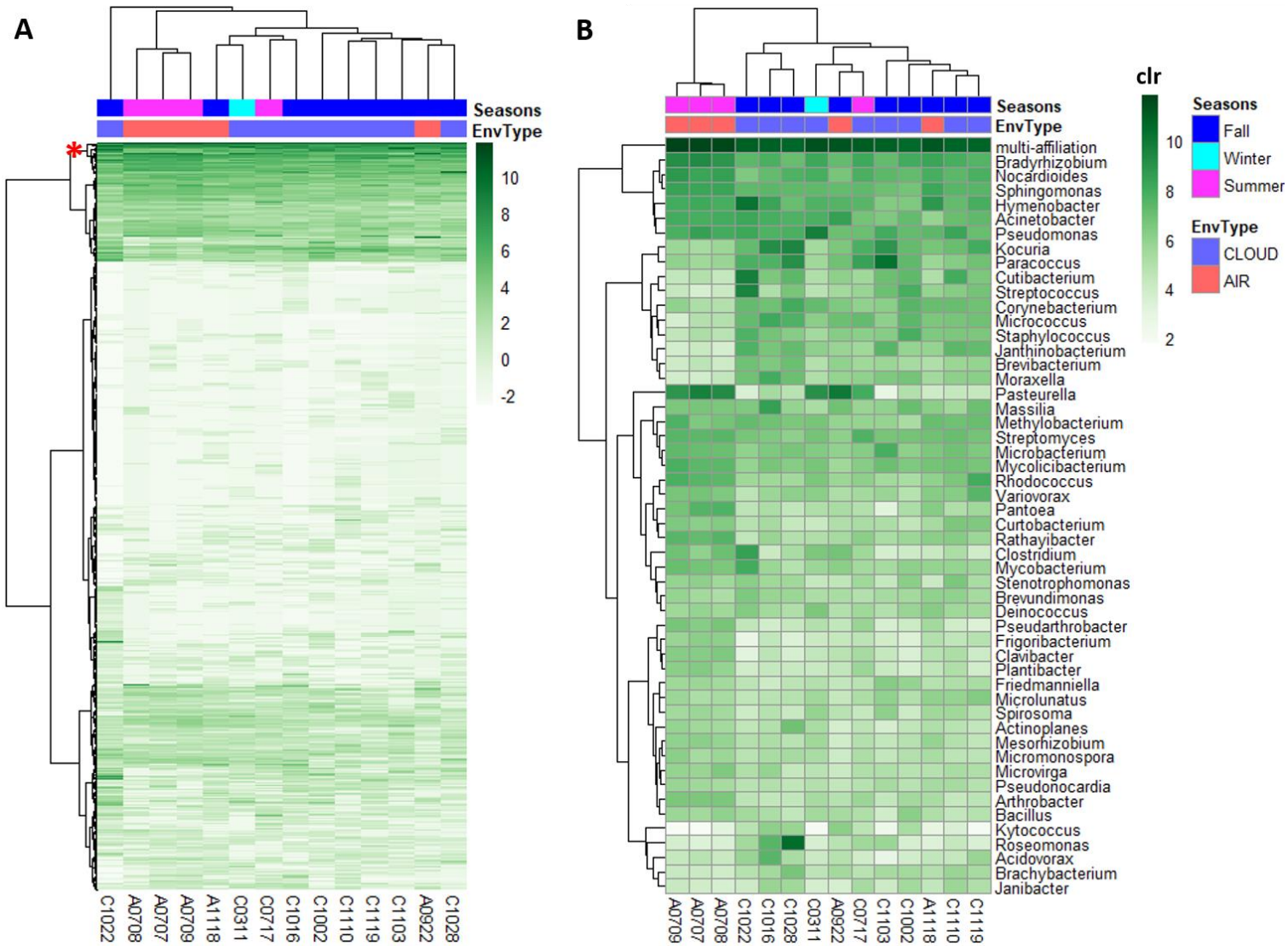
1589



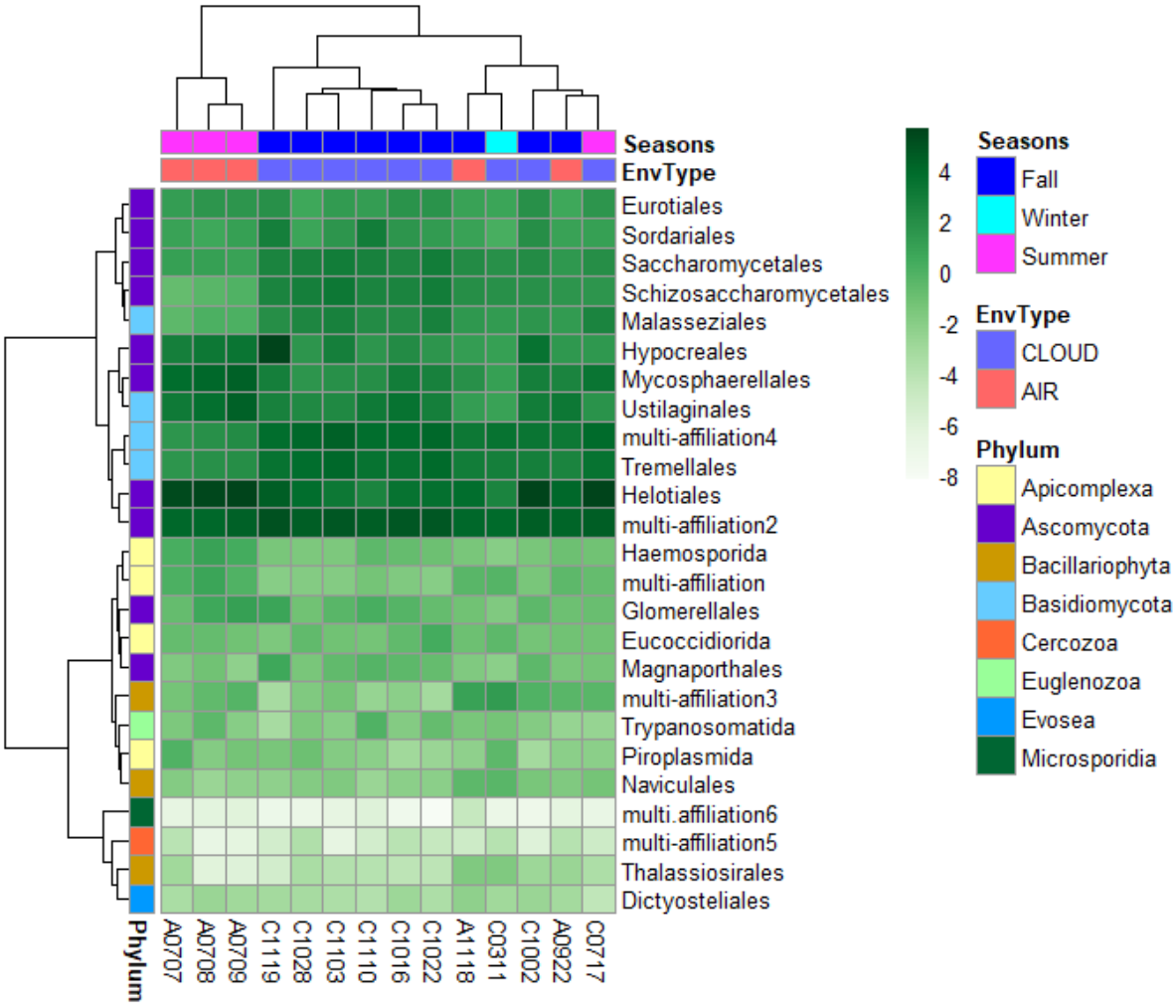
Supplementary Figure 2: Stacked numbers of bacterial and eukaryotic sequences affiliated with kraken2 in metagenomic (MG) and metatranscriptomic (MT) datasets. A: in MG data; **B**: in MT data; **C**: mean proportions in MG and MT data.



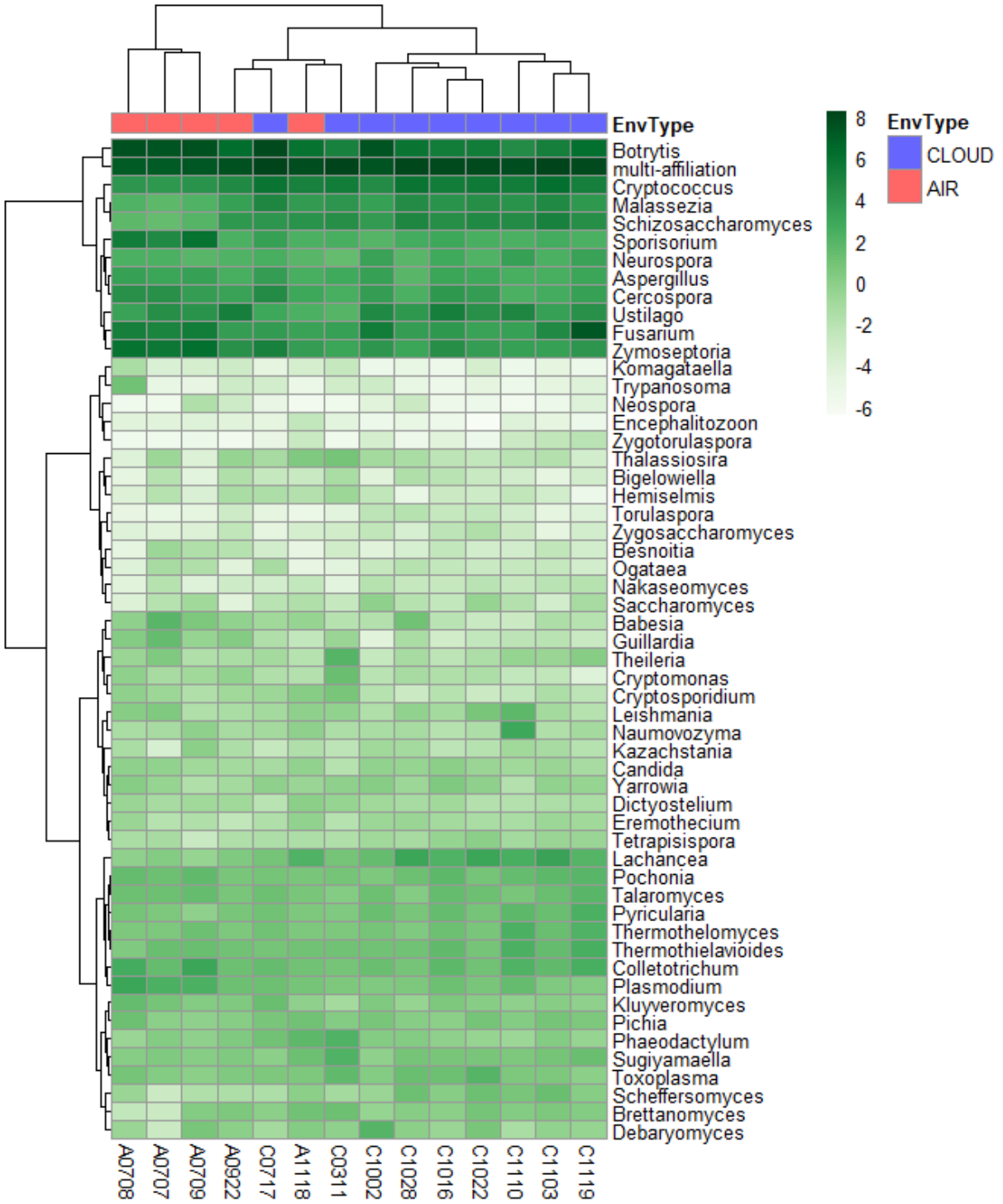
Supplementary Figure 3: Distribution of the most abundant bacterial orders (top cluster), and corresponding hierarchical clusterings (Ward's method, "ward.D2"). Intensity scale depicts centered-log ratio (clr) abundances. EnvType: environment type. Sample name = "A" for aerosol or "C" for cloud and sampling date under the format "mmdd" (month and day).



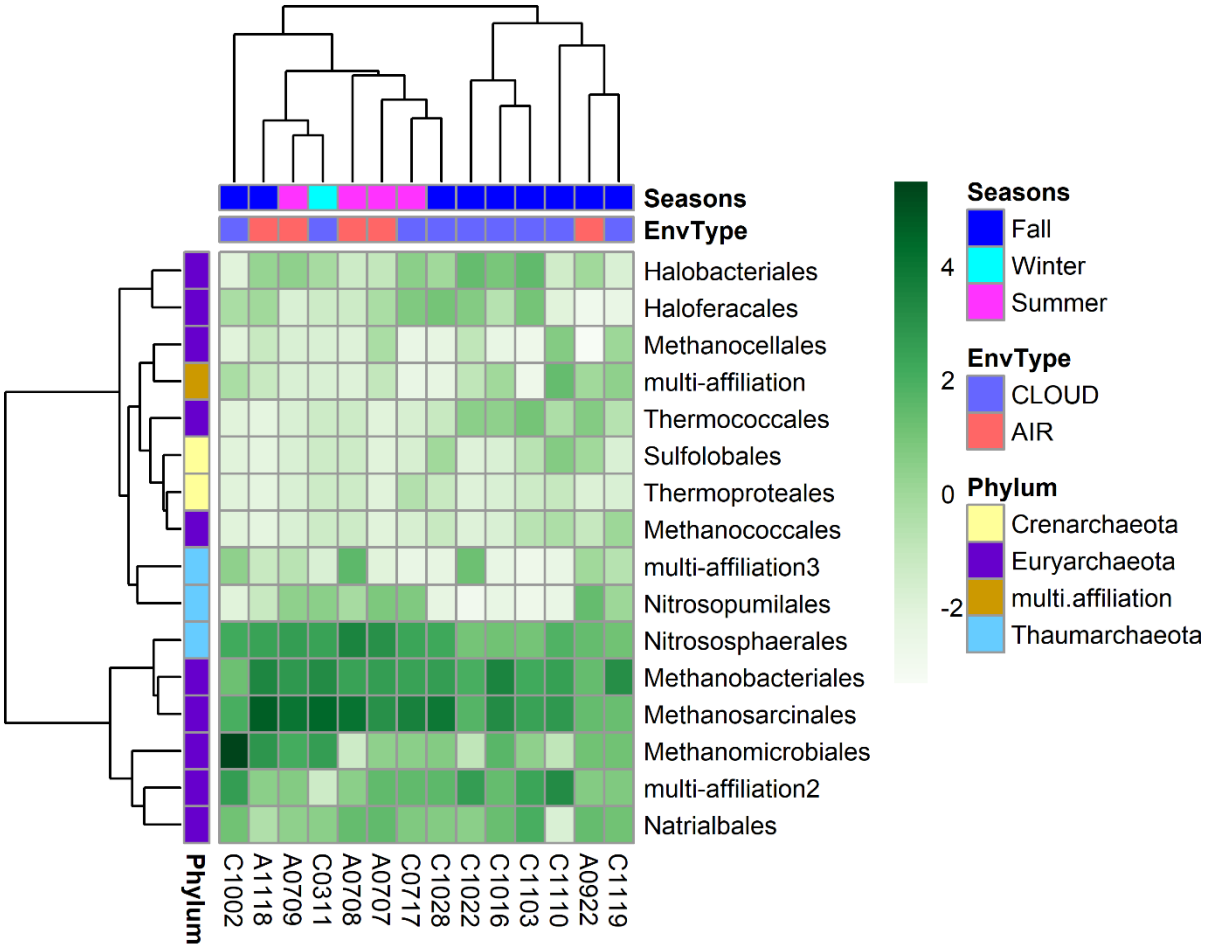
Supplementary Figure 4: Distribution of bacterial genera over samples. A: all 1250 genera; **B**: focus on the first top cluster (red asterisk). Hierarchical clusterings were done using the Ward's method (ward.D2). Intensity scale describes centered-log ratio (clr) abundances. EnvType: environment type. Sample name = "A" for aerosol or "C" for cloud and sampling date under the format "mmdd" (month and day).



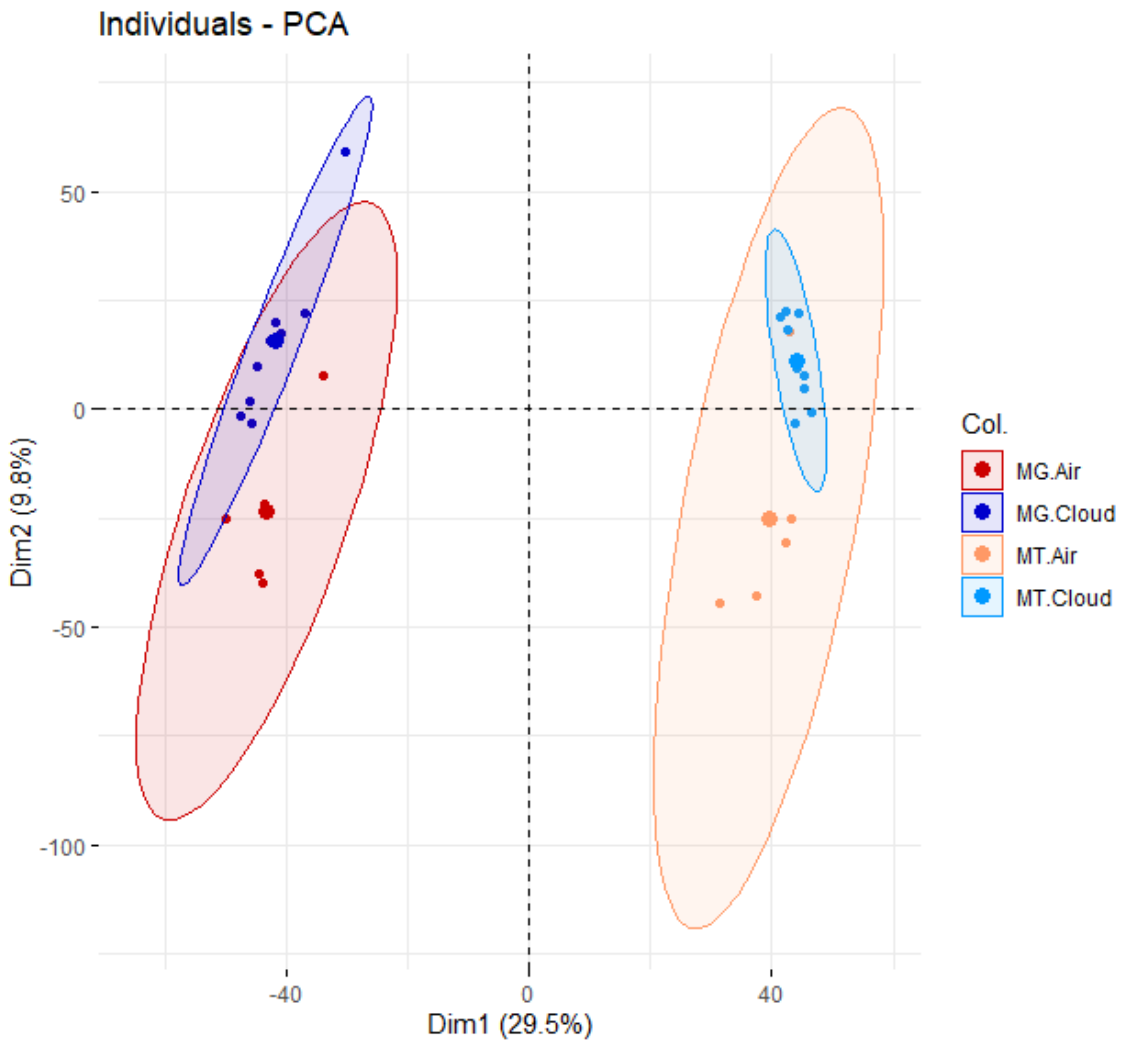
Supplementary Figure 5: Distribution of the total eukaryotic orders, and corresponding hierarchical clusterings (Ward's method, "ward.D2"). Intensity scale depicts centered-log ratio (clr) abundances. EnvType: environment type. Sample name = "A" for aerosol or "C" for cloud and sampling date under the format "mmdd" (month and day).



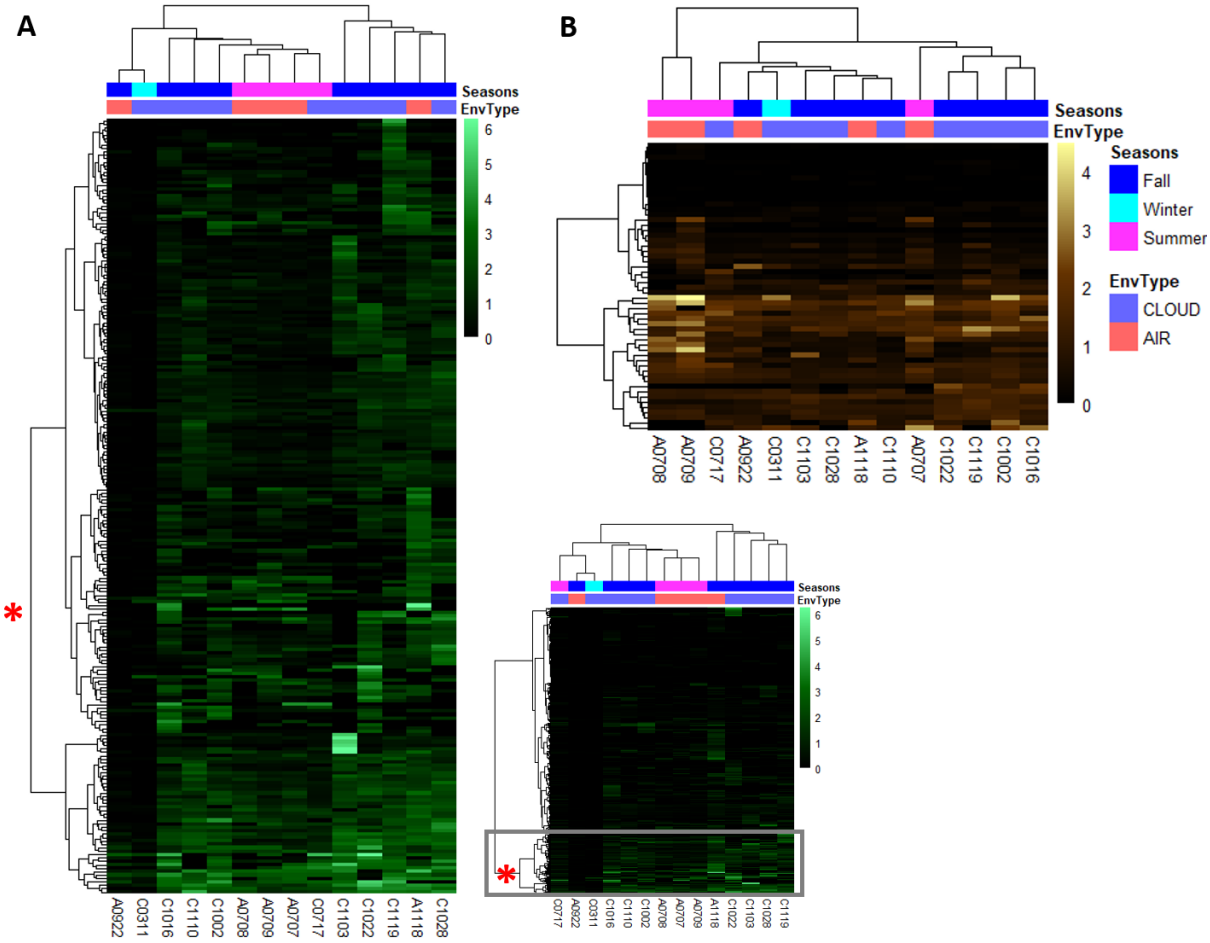
Supplementary Figure 6: Distribution of the total 55 eukaryotic genera over samples. Hierarchical clusterings were done using the Ward's method (ward.D2). Intensity scale describes centered-log ratio (clr) abundances. EnvType: environment type. Sample name = "A" for aerosol or "C" for cloud and sampling date under the format "mmdd" (month and day).



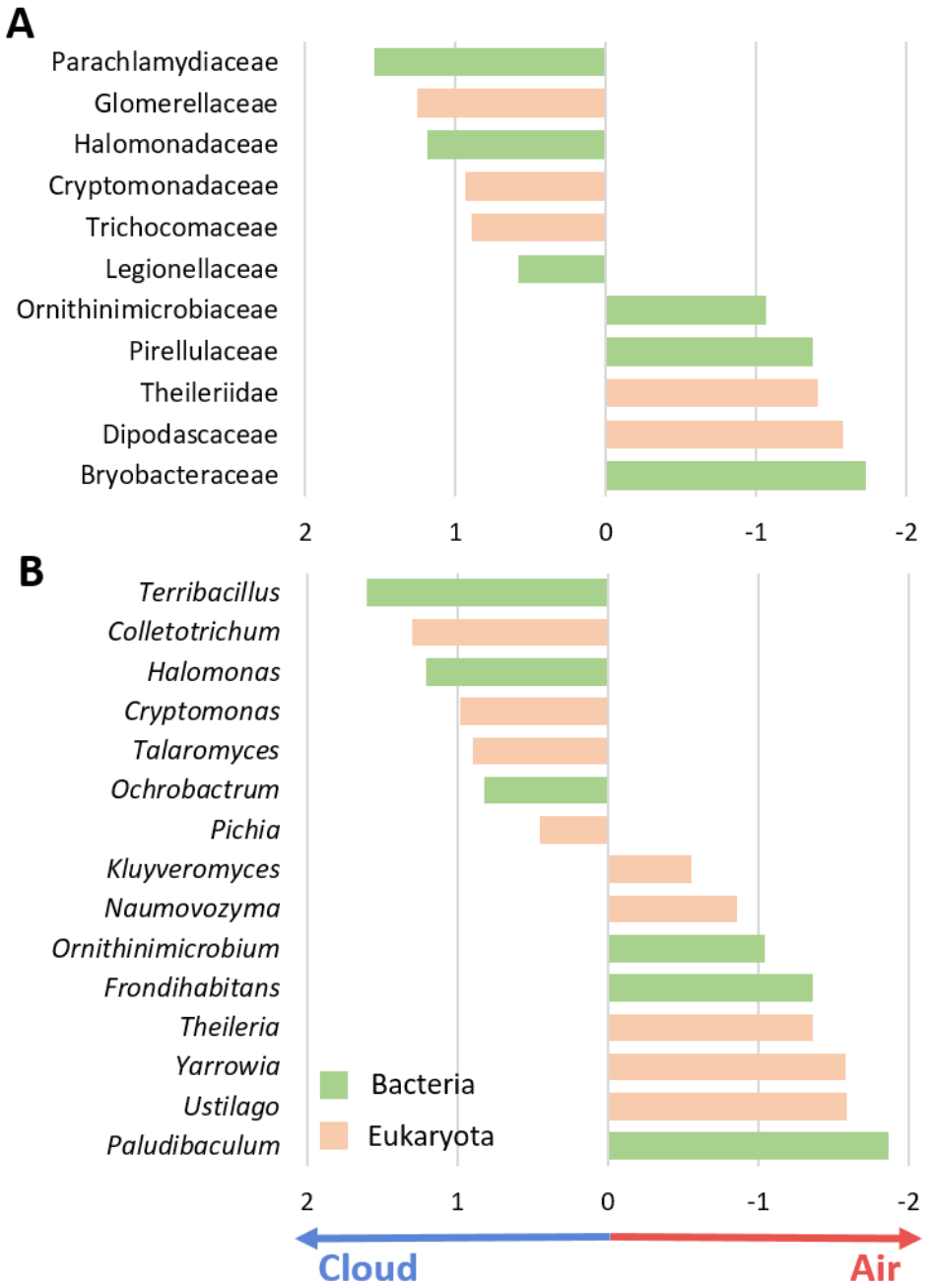
Supplementary Figure 7: Distribution of the 14 archaeal orders over samples. Hierarchical clusterings were done using the Ward's method (ward.D2). Intensity scale describes centered-log ratio (clr) abundances. EnvType: environment type. Sample name = "A" for aerosol or "C" for cloud and sampling date under the format "mmdd" (month and day).



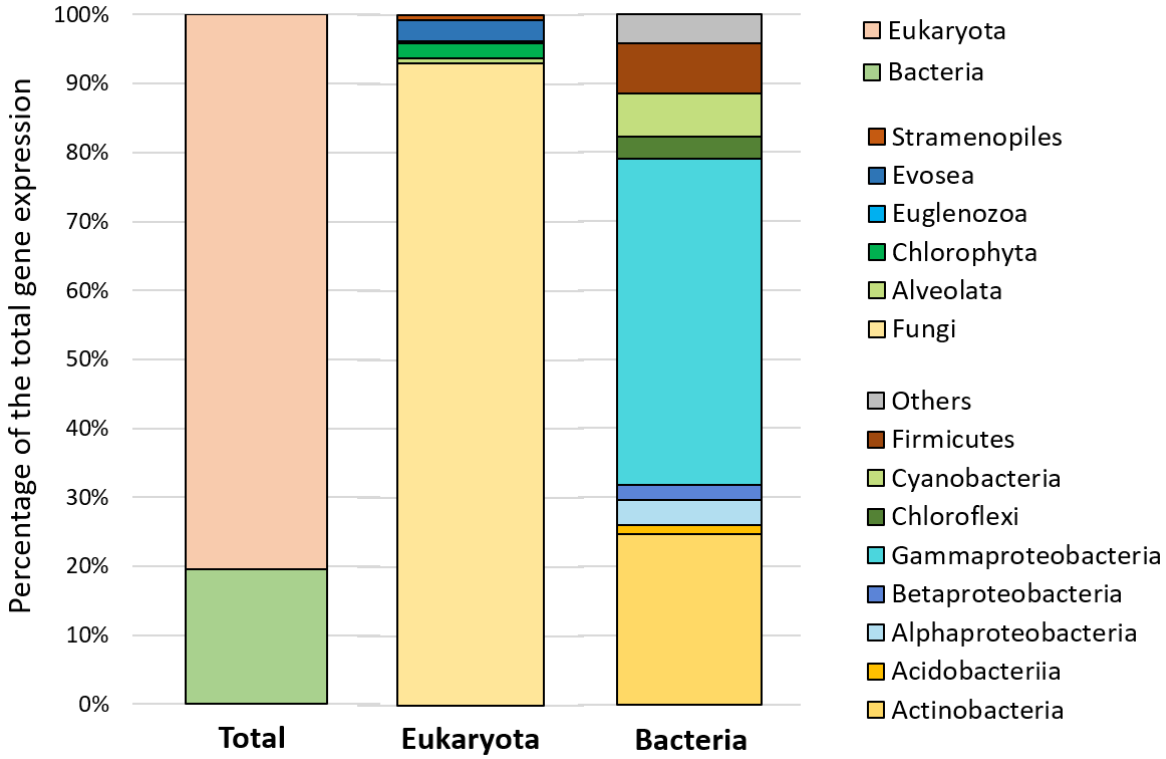
Supplementary Figure 8: Principal component analysis based on the total biodiversity in metagenomic (MG) and metatranscriptomic (MT) data for clouds and aerosols (air). Based on 6,373 taxa. Count data were centered-log ratio (clr) transformed. (R package factoExtra; ellipse type: "confidence")



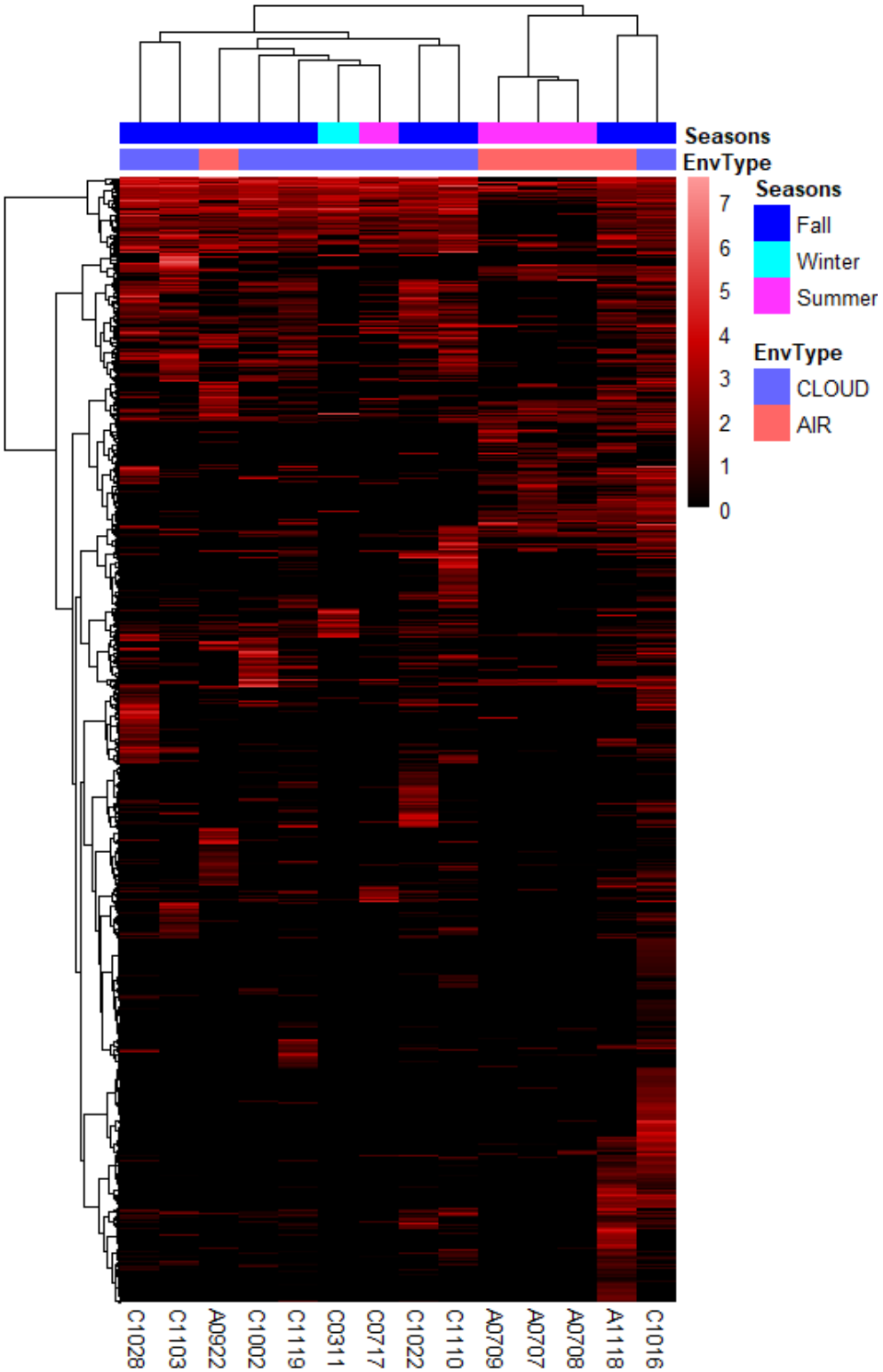
Supplementary Figure 9: RNA:DNA ratios at genus level for bacteria (A) and eukaryotes (B). The main heatmap for bacteria (left) represents the 176 top abundant genera from the heatmap including all genera (1,250 genera; lower right). Hierarchical clusterings were done with the Ward's method (ward.D2). Intensity scale depicts log abundances of corresponding sequencing reads. EnvType: environment type. Sample name = "A" for aerosol or "C" for cloud and sampling date under the format "mmdd" (month and day).



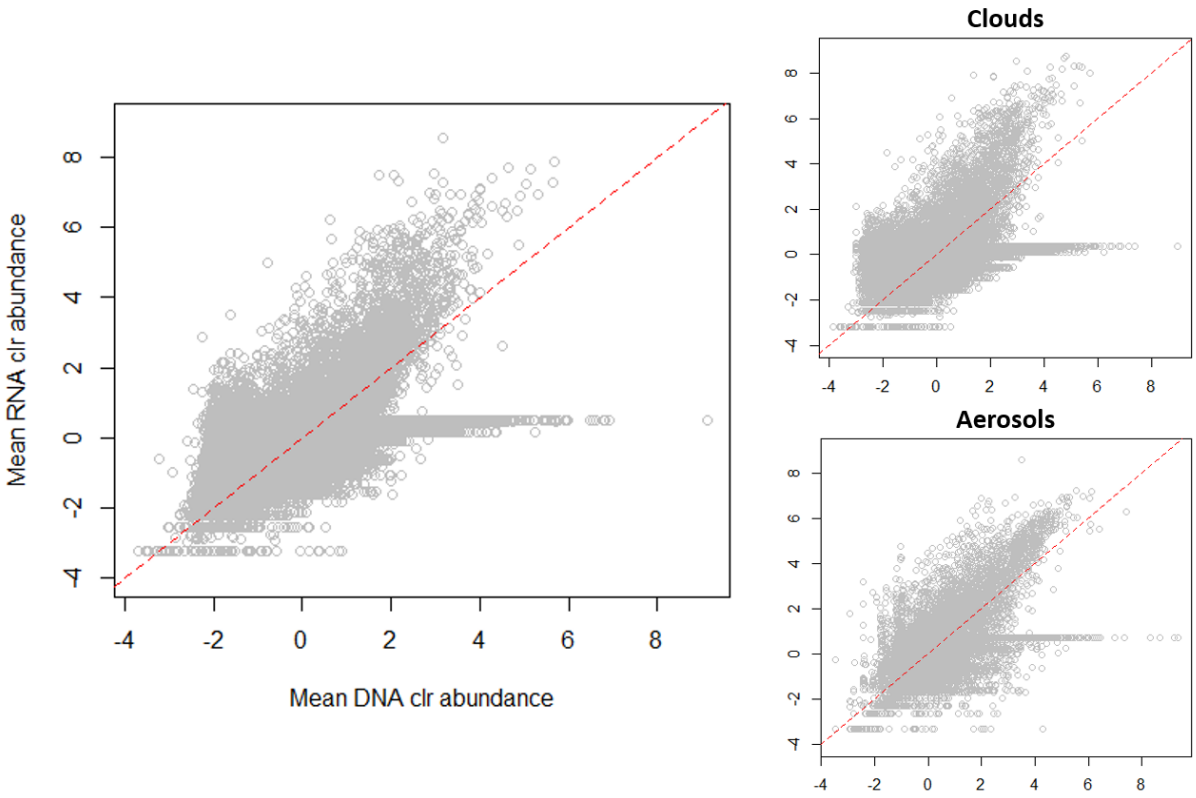
Supplementary Figure 10: Significantly differentially overrepresented bacterial and eukaryotic families (A) and genera (B) between cloud and aerosol samples. Values represent differential expression analysis coefficients from MTXmodel R package. Positive coefficient means taxon is more expressed in clouds.



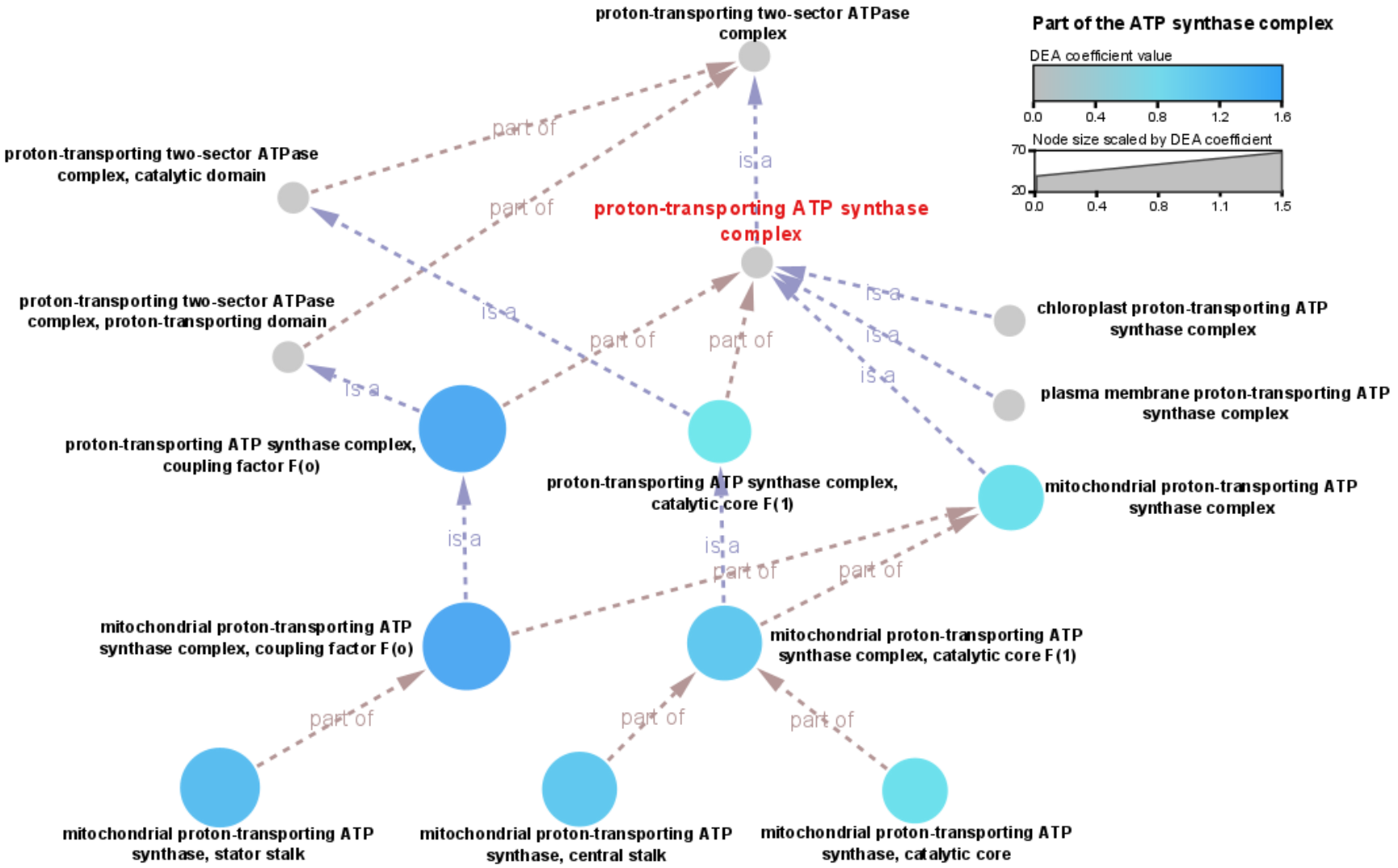
Supplementary Figure 11: Proportions (%) of overexpressed genes affiliated to eukaryota or bacteria and their respective phyla. Based on the 488 significantly expressed genes from differential expression analysis.



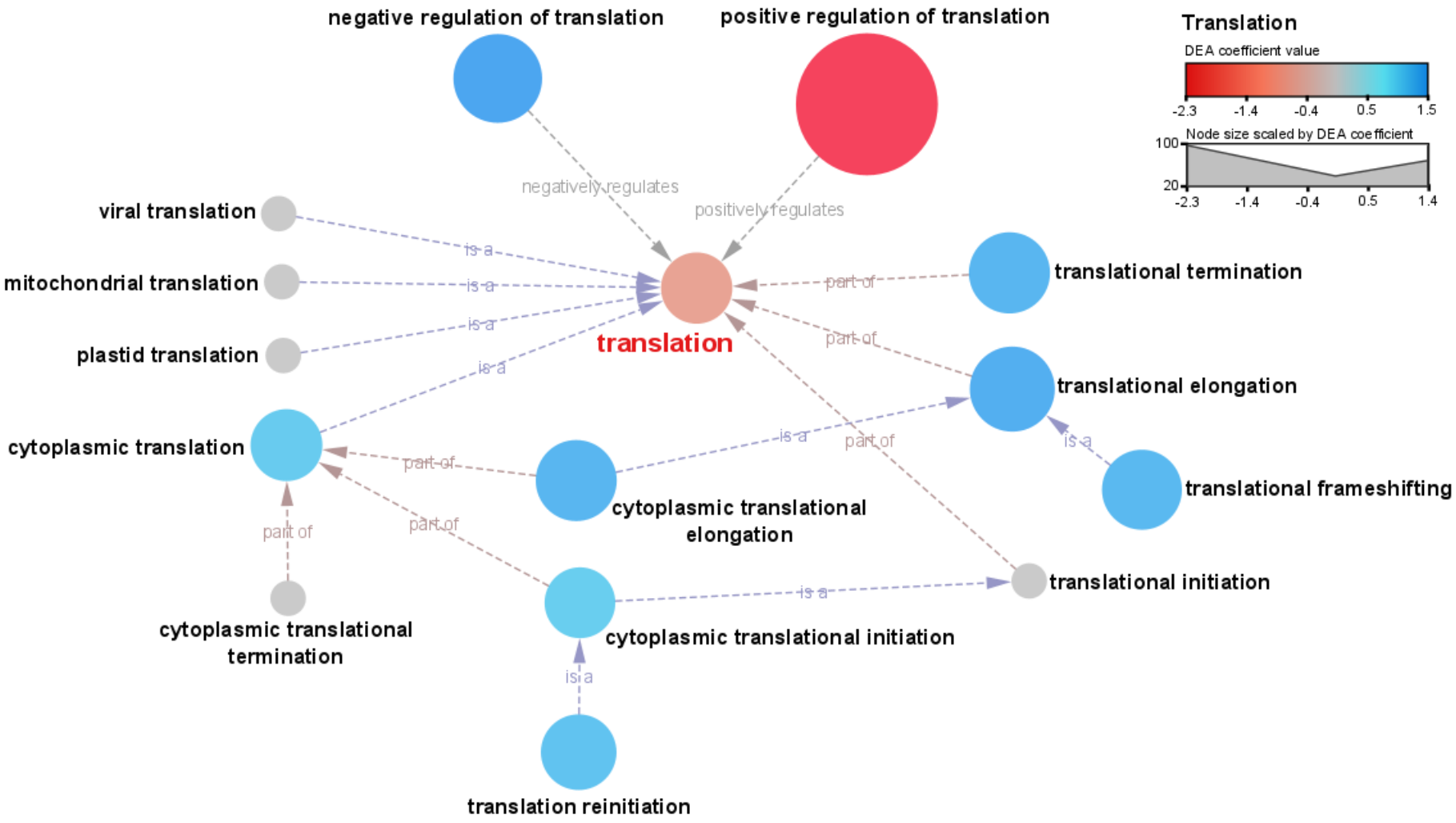
Supplementary Figure 12: RNA:DNA log ratios for each gene entries in cloud and aerosol samples. Only 8,627 genes are represented here (ratios were not calculable for the others). Red scale means values of RNA:DNA ratios in log. RNA and DNA data were first normalized as relative counts. EnvType: environment type. Sample name = "A" for aerosol or "C" for cloud and sampling date under the format "mmdd" (month and day).



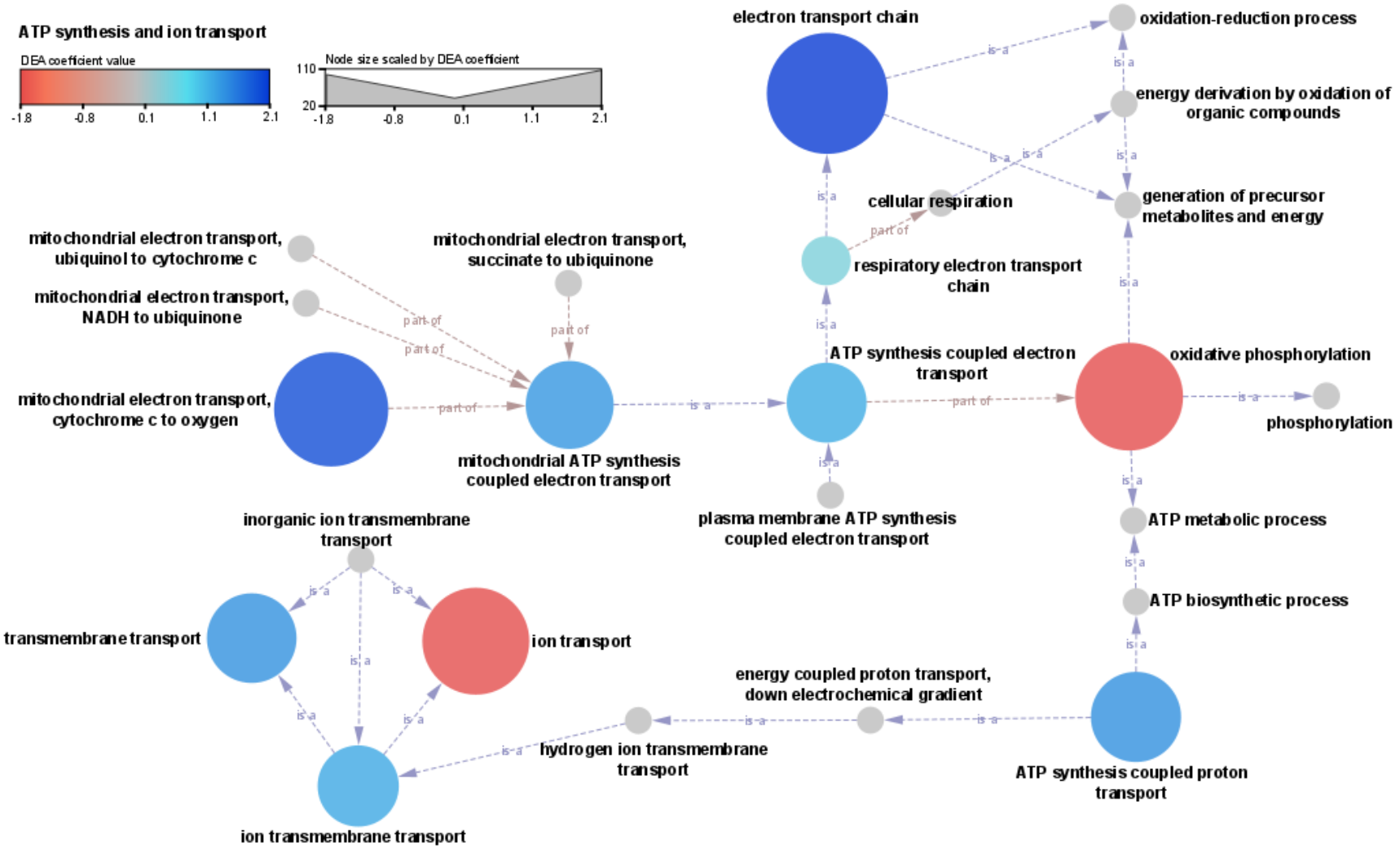
Supplementary Figure 13: RNA data against DNA data plots for all the samples, or clouds and aerosols only. Based on the 21,046 genes IDs recovered.



Supplementary Figure 14: Gene ontology (GO) relationship tree for Cellular Components related to the proton-transporting ATP synthase complex in clouds and aerosols. The blue shade scale represents the Differential Expression Analysis (DEA) coefficient value, with positive values indicating a significant overrepresentation in clouds as opposed to aerosols. The size of the nodes is scaled by the absolute value of the DEA coefficient.



Supplementary Figure 15: Gene ontology (GO) relationship tree for Biological Processes related to translation in clouds and aerosols. The red to blue color scale represents the Differential Expression Analysis (DEA) coefficient value, with negative values (red shades) indicating a significant overrepresentation in aerosols as opposed to positive values (blue shades) signifying an overrepresentation in clouds. The size of the nodes is scaled by the absolute value of the DEA coefficient.



Supplementary Figure 16: Gene ontology (GO) relationship tree for Biological Processes related to ATP synthesis and ion transport in clouds and aerosols. The red to blue colour scale represents the Differential Expression Analysis (DEA) coefficient value, with negative values (red shades) indicating a significant overrepresentation in aerosols as opposed to positive values (blue shades) signifying an overrepresentation in clouds. The size of the nodes is scaled by the absolute value of the DEA coefficient

4. Conclusions on the comparative functional analysis of clouds and aerosols

Key findings of **Article 4**:

- Microbial diversity is similar in both clouds and aerosols and only a small fraction is active in them, reflecting a demanding and selective environment and the fact that aerosols also play a role in atmospheric microbial activity.
- The clouds were more microbiologically active than the dry atmosphere (aerosols), indicating a potential "revival" of airborne microorganisms through the aqueous medium provided. Clouds are therefore specific microbial habitats for microorganisms, like oases in a desert atmosphere.
- Airborne microorganisms reacted and tried to acclimate to both atmospheric situations, giving clues of the specific selective pressures encountered in aerosols or clouds.

This functional study brings new horizons for microbial activity in the atmosphere with clouds like oases in such a hostile environment. These metagenomes and metatranscriptomes of clouds and aerosols were the first to be obtained (without prior amplification step) to our knowledge. To go further, it would be useful to have absolute quantification and complementary laboratory experiments to confirm specific activities. Moreover, a larger dataset with extended conditions (e.g. to include rain sample in the analysis, to sample at other site in the world) will permit to gain statistic power, to do correlations and to have a better view of the whole atmospheric cycle of microorganisms.

Many improvements can still be made concerning sampling protocol and data processing. There were differences in the sampling protocol (e.g., volume and concentration of the collection liquid) that can have induced some biases. Not all DNA and RNA extracts were concentrated prior to sequencing, but the sequencing data was processed in relative proportions which, we believe, compensated. Improvements concerning the taxonomic affiliation will be made in collaboration with the MEDIS team (Clermont-Ferrand) by reconstructing the rRNA sequences. The bioinformatics workflow will also be upgraded and implemented on the Galaxy Europe platform.

As a final point, metatranscriptomics is a powerful tool which generated a huge amount of information. The main challenge is to extract as much information as possible in relation to our scientific question. Generally, not all the available information is used. Moreover, many questions and problematics remained for the bioinformatics processing of these data such as the lack of environmental databases, the lack of standardized processing protocol and the multiple possibilities for the standardization of MT sequences count.

General Conclusion and Perspectives

1. Conclusion

As concluding remarks on this thesis work, cloudy situations were investigated in terms of microbial diversity and functional profile, in comparison to other atmospheric situations (i.e., precipitation and aerosols), in order to define the potential specificities of clouds as an atmospheric microbial habitat. This was the first time, to our knowledge, that **clouds were directly compared to associated rain events and to aerosols in terms of biology**.

A powerful experimental set up was developed for examining natural atmospheric conditions. A total of 10 clouds, 11 aerosols and 7 rain were collected over two years. The optimized protocol resulted in the recovery of sufficient quantities of nucleic acids to perform direct sequencing and ultimately obtained 9 MG and MT from clouds and 6 from aerosols. These are **the first non-amplified MG and MT from the atmosphere** to our knowledge and undoubtedly constitute a major advance in aeromicrobiology.

Investigation of bacteria in clouds and their precipitation evidenced that bacterial biomass in rain was mainly driven by the scavenging of airborne bacteria from the air column. This contradicts previous estimates in the literature (Moore et al., 2020a) and evidences the influence of local sources; of course, this must depends on the respective altitudes between cloud and precipitation sampling sites. In turn, a large proportion of the diversity in rain originated from the source clouds, which illustrated their important role in the dispersal of microbes over long distances.

A comparative analysis of microbial diversity between aerosols and clouds was performed through 16S rRNA gene amplicon sequencing and metagenomes, on the same samples. The same main abundant taxa were present in both cases, with the exception (among others) of the bacterial genera *Acidiphilium* and *Pelomonas* only detected by amplicon sequencing, and *Pasteurella* and *Cutibacterium* only identified in the metagenomes. These few differences underline the importance of approach in the description of microbial communities. All these differences influence the final results to some extent and should be kept in mind when interpreting the data. In addition, a seasonal effect on the distribution of bacterial diversity was detected in the amplicon study thanks to replicate sampling and analyses, but this was not clear in the metagenomic data. This points out the relevance and strengths of analyzing replicates from sampling to sequencing. The metagenomics studies presented other challenges (low biomass and direct sequencing) that required sacrificing the replicate aspect.

Analysis of the microbial functional profile of aerosols and clouds revealed potentially active communities in both atmospheric situations, with only a small fraction of the community surviving extreme atmospheric conditions. Microorganisms have previously been reported to be active when aerosolized (Krumins et al., 2014), but this is the first time specific expressed functions have been described in aerosols. This first observation brings a new advance in aeromicrobiology, with **aerosols**

supporting microbial activity, just like clouds, and also having an impact on atmospheric chemistry (e.g., H_2O_2 catabolism). However, clouds harbored a higher functional potential. Clouds have been previously described as containing active microbial communities and specific expressed functions (Amato et al., 2019), but here, comparison with dry atmospheric situations (i.e., without condensed water) have **highlighted clouds as specific microbial habitats in the atmosphere, likely to awaken the microbial metabolism**. Energy metabolism was clearly overexpressed in clouds compare to aerosols, and the multiple starvation responses observed in cloudy situations can be interpreted as a resurgence of cells activity and additional nutrient requirements.

To summarize the findings on the atmospheric cycle of microorganisms:

Microorganisms are aerosolized from variable sources and are mixed with air masses forming highly diverse assemblages. The small fraction that survive and remain active in aerosols are exposed to desiccation, direct sunlight and free radicals, and can potentially impact atmospheric chemistry. These airborne microorganisms can be transported to high altitudes and become part of clouds. The presence of condensed water provides more favorable living conditions to microbes, with protection from desiccation and direct sunlight. Nevertheless, clouds also expose cells to other selective pressures, such as osmotic shocks, freeze-thaw cycles, and perhaps increased UV exposure due to Mie scattering. Microbial metabolism is reactivated in clouds, for those cells that have survived so far, making the clouds a kind of oasis in a desert atmosphere as hypothesized. Meanwhile, microorganisms can affect the chemistry and physics of clouds. Multiple evaporation-condensation cycles will occur before the clouds precipitate, bringing the microbial cells back to the Earth's surface, some of which are "revived" and ready to colonize new ecosystems. Precipitation thus disseminates subsamples of cloud microbial diversity to local ecosystems, participating in genetic mixing between distant environments. Rainfalls also scavenges the air column, loading themselves with biomass and local emission sources, thereby limiting the vertical dispersion of many bioaerosols.

As a last point, a **bioinformatics workflow (BiW) has been developed to process MT data and their associated MG**. This new workflow is also an important contribution to the field of aeromicrobiology, but also to the analysis of environmental nucleic acids more generally. It will ultimately be publicly available on the Galaxy Europe platform. The **gene catalog constructed from atmospheric microbial communities** will serve as a reference for future work.

2. Perspectives

Additional efforts will be needed to **collect more atmospheric samples to allow for more robust statistics**. First, this would permit correlations between bacterial richness and biomass in precipitation as a function of dilution factor from clouds and perhaps also with precipitation duration and intensity. In addition, it would be very useful for environmental research to try **to calculate the upward and downward fluxes of microorganisms** (i.e., exchanges with the surface). Second, it would be interesting to collect aerosol and cloud events successively, on the same site, to observe the continuity between the two atmospheric situations and to have a better idea of the impact of the presence of condensed water on the activity of airborne microbial communities. Indeed, our question is not totally solved regarding the specificities of clouds compared to aerosols, because the samples came from different meteorological conditions during the year, and a seasonal effect can therefore not be dismissed. We can also think of sampling only the aqueous phase of the clouds and study the microbial activity in cloud water (as was done in Amato et al., 2019, with a cloud impactor) versus the dry aerosols outside the clouds.

It will also be of great interest to extend the MT and MG analyses on clouds and aerosols to other geographical locations in order to estimate **whether atmospheric microbial activity converges on a global scale** or depends mainly on the microbial sources composing the communities (i.e., depending on the local emission sources). In particular, urban areas, polar regions or sea surface air could be examined.

There are many challenges associated with the processing MT data, and complementary work could be done to improve the analyses presented. Taxonomic affiliation can be improved by first assembling the rRNA sequences. This is one of the objectives of the collaboration with the MEDIS team (Clermont-Ferrand). The choice of the database is one of the main limitations to taxonomic affiliations, and the richness of these databases is all the more important for environmental analyses. **Improving the content of databases should be a major concern for environmental research.** Potential gene prediction can also be enhanced in our BiW, especially for eukaryotes, by coupling a tool detecting prokaryotic genes with a tool identifying eukaryotic genes (e.g., MetaEuk). The same could be done for mapping, using tools such as STAR (Dobin et al., 2013) that can handle the presence of introns and exons in eukaryotic genomes. Another challenge is the standardization of MT counts and differential expression analysis (DEA). Indeed, the best way to normalize MT data is to do so based on the number of gene copies in MGs. Nevertheless, multiple standardization methods are used and each has its weaknesses. There is no standard procedure to do this and only one recent tool (R package MTXmodel) proposes to do the DEA and standardize the MT against the MG. Again, **efforts should be made to propose standardized and appropriate methods to normalize MT data**. Finally, there is a lack of

existing tools to do visualization suitable to MT, while this is one of the most important points to better explore and interpret such complex datasets. This specific point on visualization is currently under discussion and development with the Freiburg Galaxy team.

The unique dataset obtained (MT and MG from clouds and aerosols; gene catalog) will be used in on going and future projects in the group, such as the research of genes or activity in a project related to C1 compounds metabolism in clouds (ANR "METACLOUD" project in collaboration with research groups in Toulouse and Strasbourg, France), the presence of antibiotic genes resistance (NSERC "ARG" project, in collaboration with Univ. Laval, Canada), or again the use of nitrogen compounds by airborne bacteria.

Genome methylation (methylome) (Cohen et al., 2016; Pelizzola and Ecker, 2011; Sánchez-Romero et al., 2015) could also be a subject of study for examining ageing processes and evolution of biological aerosols during their transport. This could allow estimation of DNA ageing, i.e., the rate of damage inflicted to genomes (Gong and Miller, 2019; Grafstrom et al., 1984) exposed to atmospheric conditions, which could be related to time spent in the atmosphere. This could perhaps permit differentiation between microorganisms originating from local sources from those transported from long distances which spent longer times exposed to air.

Microbial activity and expressed functions have been detected in the atmosphere using metatranscriptomics. However, **quantification was only relative** and complementary methods such as quantitative PCR (qPCR) is needed to obtain absolute quantifications of genes and transcripts. It would be interesting to look closer to genes related to INA (e.g, Ina*W* gene, Ahern *et al*., 2007; Ina*Z* gene, Green and Warren, 1985), stress responses (e.g., hydrogen peroxide catabolism: ahp*C*, ahp*F*; Arts et al., 2015) or genes related to C1 compound metabolism, among others.

In addition, **the presence of RNA is not an absolute proof of activity** and further analysis or laboratory experiments will be needed to confirm specific functions of interest. One of many possibilities is to investigate catabolic (e.g., H_2O_2, polysaccharide, glycerol) or biosynthetic processes (e.g., fatty acid, phospholipids, acetyl-CoA) that have been detected in clouds and aerosols, and to confirm these metabolic functions through compound degradation or biosynthesis experiments under conditions that best mimic these atmospheric situations (i.e., using aerosols and cloud chambers). Enzymatic assays and characterization could also be performed (e.g., with fluorogenic molecules such as FDA and MUF, that could be coupled to flow cytometry). Estimation of enzymatic kinetics can also contribute to the development of atmospheric models.

Ultimately, metatranscriptomic analyses could be coupled with metabolomics, and proteomics to go further, in order to follow and confirm the expression of a gene and the presence of its final product.

As a final point, we can now investigate the next step of the microbial cycle in the atmosphere, i.e., the arrival in local ecosystems of active microbial cells falling from clouds with precipitation, or scavenged form the air column. To our knowledge, no study has yet resolved the question of the ability of airborne cells to colonize local surfaces and what motivates this phenomenon. We can imagine further research monitoring microbial potential activity, with MT, in precipitation and soil (or lake) collected at the same location, before and after rain.

References

Ahern, H.E., Walsh, K.A., Hill, T.C.J., and Moffett, B.F. (2007). Fluorescent pseudomonads isolated from Hebridean cloud and rain water produce biosurfactants but do not cause ice nucleation. Biogeosciences *4*, 115–124.

Aho, K., Weber, C.F., Christner, B.C., Vinatzer, B.A., Morris, C.E., Joyce, R., Failor, K., Werth, J.T., Bayless-Edwards, A.L.H., and Schmale III, D.G. (2019). Spatiotemporal patterns of microbial composition and diversity in precipitation. Ecol. Monogr. *0*, 1–26.

Almaguer, M., Aira, M.J., Rodríguez-Rajo, F.J., and Rojas, T.I. (2014). Temporal dynamics of airborne fungi in Havana (Cuba) during dry and rainy seasons: influence of meteorological parameters. Int. J. Biometeorol. *58*, 1459–1470.

Amato, P. (2013). Energy Metabolism at Low-temperature and Frozen Conditions in Cold-adapted Microorganisms. Cold-Adapted Microorg. 71–96.

Amato, P., Ménager, M., Sancelme, M., Laj, P., Mailhot, G., and Delort, A.M. (2005). Microbial population in cloud water at the Puy de Dôme: Implications for the chemistry of clouds. Atmos. Environ. *39*, 4143–4153.

Amato, P., Parazols, M., Sancelme, M., Laj, P., Mailhot, G., and Delort, A.M. (2007). Microorganisms isolated from the water phase of tropospheric clouds at the Puy de Dôme: Major groups and growth abilities at low temperatures. In FEMS Microbiology Ecology, pp. 242–254.

Amato, P., Joly, M., Besaury, L., Oudart, A., Taib, N., Moné, A.I., Deguillaume, L., Delort, A.M., and Debroas, D. (2017). Active microorganisms thrive among extremely diverse communities in cloud water. PLoS One *12*, 1–22.

Amato, P., Besaury, L., Joly, M., Penaud, B., Deguillaume, L., and Delort, A.M. (2019). Metatranscriptomic exploration of microbial functioning in clouds. Sci. Rep. *9*.

American Meteorological Society - AMS (2012). American Meteorological Society - Glossary of Meteorology.

André, F., Jonard, M., and Ponette, Q. (2007). Influence of meteorological factors and polluting environment on rain chemistry and wet deposition in a rural area near Chimay, Belgium. Atmos. Environ. *41*, 1426–1439.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Anesio, A.M., Hodson, A.J., Fritz, A., Psenner, R., and Sattler, B. (2009). High microbial activity on glaciers: Importance to the global carbon cycle. Glob. Chang. Biol. *15*, 955–960.

Arts, I.S., Gennaris, A., and Collet, J.F. (2015). Reducing systems protecting the bacterial cell envelope from oxidative damage. FEBS Lett. *589*, 1559–1568.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. *25*, 25.

Asmi, E., Freney, E., Hervo, M., Picard, D., Rose, C., Colomb, A., and Sellegri, K. (2012). Aerosol cloud activation in summer and winter at puy-de-Dôme high altitude site in France. Atmos. Chem. Phys. *12*, 11589–11607.

Aydogan, E.L., Moser, G., Müller, C., Kämpfer, P., and Glaeser, S.P. (2018). Long-term warming shifts the composition of bacterial communities in the phyllosphere of Galium album in a permanent grassland field-experiment. Front. Microbiol. *9*, 144.

Baldrian, P., Kolaiřík, M., Štursová, M., Kopecký, J., Valášková, V., Větrovský, T., Žifčáková, L., Šnajdr, J., Rídl, J., Vlček, Č., et al. (2012). Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. ISME J. *6*, 248–258.

Baray, J.L., Bah, A., Cacault, P., Sellegri, K., Pichon, J.M., Deguillaume, L., Montoux, N., Noel, V., Seze, G., Gabarrot, F., et al. (2019). Cloud occurrence frequency at puy de dome (France) deduced from an automatic camera image analysis: Method, validation, and comparisons with larger scale parameters. Atmosphere (Basel). *10*, 808.

Baray, J.L., Deguillaume, L., Colomb, A., Sellegri, K., Freney, E., Rose, C., Baelen, J. Van, Pichon, J.M., Picard, D., Fréville, P., et al. (2020). Cézeaux-Aulnat-Opme-Puy de Dôme: A multi-site for the long-term survey of the tropospheric composition and climate change. Atmos. Meas. Tech. *13*, 3413–3445.

Barberán, A., Ladau, J., Leff, J.W., Pollard, K.S., Menninger, H.L., Dunn, R.R., and Fierer, N. (2015). Continental-scale distributions of dust-associated bacteria and fungi. PNAS *112*, 5756–5761.

Barnard, R.L., Osborne, C.A., and Firestone, M.K. (2013). Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. ISME J. *7*, 2229–2241.

Bauer, H., Kasper-Giebl, A., Löflund, M., Giebl, H., Hitzenberger, R., Zibuschka, F., and Puxbaum, H. (2002). The contribution of bacteria and fungal spores to the organic carbon content of cloud water, precipitation and aerosols. Atmos. Res. *64*, 109–119.

Bauer, H., Giebl, H., Hitzenberger, R., Kasper-Giebl, A., Reischl, G., Zibuschka, F., and Puxbaum, H. (2003). Airborne bacteria as cloud condensation nuclei. J. Geophys. Res. Atmos. *108*, 4658.

Behzad, H., Gojobori, T., and Mineta, K. (2015). Challenges and opportunities of airborne metagenomics. Genome Biol. Evol. *7*, 1216–1226.

Bertolini, V., Gandolfi, I., Ambrosini, R., Bestetti, G., Innocente, E., Rampazzo, G., and Franzetti, A. (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.

Bertrand, G., Celle-Jeanton, H., Laj, P., Rangognio, J., and Chazot, G. (2008). Rainfall chemistry: Long range transport versus below cloud scavenging. A two-year study at an inland station (Opme, France). J. Atmos. Chem. *60*, 253–271.

Blanco-Alegre, C., Castro, A., Calvo, A.I., Oduber, F., Alonso-Blanco, E., Fernández-González, D., Valencia-Barrera, R.M., Vega-Maray, A.M., and Fraile, R. (2018). Below-cloud scavenging of fine and coarse aerosol particles by rain: The role of raindrop size. Q. J. R. Meteorol. Soc. *144*, 2715–2726.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bourcier, L., Masson, O., Laj, P., Chausse, P., Pichon, J.M., Paulat, P., Bertrand, G., and Sellegri, K. (2012). A new method for assessing the aerosol to rain chemical composition relationships. Atmos. Res. *118*, 295–303.

Bowers, R.M., McLetchie, S., Knight, R., and 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.

Bowers, R.M., Clements, N., Emerson, J.B., Wiedinmyer, C., Hannigan, M.P., and Fierer, N. (2013). Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. Environ. Sci. Technol. *47*, 12097–12106.

Bryan, N.C., Christner, B.C., Guzik, T.G., Granger, D.J., and Stewart, M.F. (2019). Abundance and survival of microbial aerosols in the troposphere and stratosphere. ISME J. *13*, 2789–2799.

Buchfink, B., Xie, C., and Huson, D.H. (2014). Fast and sensitive protein alignment using DIAMOND. Nat. Methods 2014 121 *12*, 59–60.

Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F., Rauf, P., Huettel, B., Reinhardt, R., Schmelzer, E., et al. (2012). Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. Nat. 2012 4887409 *488*, 91–95.

Burrows, S.M., Butler, T., Jöckel, P., Tost, H., Kerkweg, A., Pöschl, U., and Lawrence, M.G. (2009a). Bacteria in the global atmosphere-Part 2: Modeling of emissions and transport between different ecosystems.

Burrows, S.M., Elbert, W., Lawrence, M.G., and Pöschl, U. (2009b). Bacteria in the global atmosphere – Part 1: Review and synthesis of literature data for different ecosystems. Atmos. Chem. Phys. *9*, 9263–9280.

Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A.D. (2019). rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. Gigascience *8*, 1–13.

Button, D.K., and Robertson, B.R. (2001). Determination of DNA Content of Aquatic Bacteria by Flow Cytometry. Appl. Environ. Microbiol. *67*, 1636–1645.

Cáliz, J., Triadó-Margarit, X., Camarero, L., and Casamayor, E.O. (2018). A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. Proc. Natl. Acad. Sci. *115*, 12229–12234.

Camacho-Sanchez, M., Burraco, P., Gomez-Mestre, I., and Leonard, J.A. (2013). Preservation of RNA and DNA from mammal samples under field conditions. Mol. Ecol. Resour. *13*, 663–673.

Campbell, B.J., Yu, L., Heidelberg, J.F., and Kirchman, D.L. (2011). Activity of abundant and rare bacteria in a coastal ocean. Proc. Natl. Acad. Sci. U. S. A. *108*, 12776–12781.

Carbon, S., Douglass, E., Good, B.M., Unni, D.R., Harris, N.L., Mungall, C.J., Basu, S., Chisholm, R.L., Dodson, R.J., Hartline, E., et al. (2021). The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. *49*, D325–D334.

Carvalhais, L.C., Dennis, P.G., Tyson, G.W., and Schenk, P.M. (2012). Application of metatranscriptomics to soil environments. J. Microbiol. Methods *91*, 246–251.

Cha, S., Lee, D., Jang, J.H., Lim, S., Yang, D., and Seo, T. (2016). Alterations in the airborne bacterial community during Asian dust events occurring between February and March 2015 in South Korea. Nat. Publ. Gr. 1–9.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890.

Christodoulou, D., Link, H., Fuhrer, T., Kochanowski, K., Gerosa, L., and Sauer, U. (2018). Reserve Flux Capacity in the Pentose Phosphate Pathway Enables Escherichia coli's Rapid Response to Oxidative Stress. Cell Syst. *6*, 569-578.e7.

Cohen, N.R., Ross, C.A., Jain, S., Shapiro, R.S., Gutierrez, A., Belenky, P., Li, H., and Collins, J.J. (2016). A role for the bacterial GATC methylome in antibiotic stress survival. Nat. Genet. *48*, 581–586.

Collett, J., Oberholzer, B., and Staehelin, J. (1993). Cloud chemistry at Mt Rigi, Switzerland: Dependence on drop size and relationship to precipitation chemistry. Atmos. Environ. Part A, Gen. Top. *27*, 33–42.

Deguillaume, L., Charbouillot, T., Joly, M., Vaïtilingom, M., Parazols, M., Marinoni, A., Amato, P., Delort, A.M., Vinatier, V., Flossmann, A., et al. (2014). Classification of clouds sampled at the puy de Dôme (France) based on 10 yr of monitoring of their physicochemical properties. Atmos. Chem. Phys. *14*, 1485–1506.

DeLeon-Rodriguez, N., Lathem, T.L., Rodriguez-R, L.M., Barazesh, J.M., Anderson, B.E., Beyersdorf, A.J., Ziemba, L.D., Bergin, M., Nenes, A., and Konstantinidis, K.T. (2013). Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. Proc. Natl. Acad. Sci. *110*, 2575–2580.

Delort, A.M., Vaïtilingom, M., Amato, P., Sancelme, M., Parazols, M., Mailhot, G., Laj, P., and Deguillaume, L. (2010). A short overview of the microbial population in clouds: Potential roles in atmospheric chemistry and nucleation processes. Atmos. Res. *98*, 249–260.

Després, V.R., Alex Huffman, J., Burrows, S.M., Hoose, C., Safatov, A.S., Buryak, G., Fröhlich-Nowoisky, J., Elbert, W., Andreae, M.O., Pöschl, U., et al. (2012). Primary biological aerosol particles in the atmosphere: A review. Tellus, Ser. B Chem. Phys. Meteorol. *64*.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). Sequence analysis STAR: ultrafast universal RNA-seq aligner. *29*, 15–21.

Drautz-Moses, D.I., Luhung, I., Gusareva, E.S., Kee, C., Gaultier, N.E., Premkrishnan, B.N. V., Lee, C.F., Leong, S.T., Park, C., Yap, Z.H., et al. (2022). Vertical stratification of the air microbiome in the lower troposphere. Proc. Natl. Acad. Sci. *119*, e2117293119.

Dueker, M.E., Weathers, K.C., O'Mullan, G.D., Juhl, A.R., and Uriarte, M. (2011). Environmental controls on coastal coarse aerosols: Implications for microbial content and deposition in the near-shore environment. Environ. Sci. Technol. *45*, 3386–3392.

Dueker, M.E., O'Mullan, G.D., Weathers, K.C., Juhl, A.R., and Uriarte, M. (2012). Coupling of fog and marine microbial content in the near-shore coastal environment. Biogeosciences *9*, 803–813.

Edwards, K.J., Bond, P.L., Gihring, T.M., and Banfield, J.F. (2000). An Archaeal Iron-Oxidizing Extreme Acidophile Important in Acid Mine Drainage. Science (80-.). *287*, 1796–1799.

Elbert, W., Taylor, P.E., Andreae, M.O., and Pöschl, U. (2007). Contribution of fungi to primary biogenic aerosols in the atmosphere: Wet and dry discharged spores, carbohydrates, and inorganic ions. Atmos. Chem. Phys. *7*, 4569–4588.

Els, N., Baumann-Stanzer, K., Larose, C., Vogel, T.M., and Sattler, B. (2019). Beyond the planetary boundary layer: Bacterial and fungal vertical biogeography at Mount Sonnblick, Austria. Geo Geogr. Environ. *6*.

Ensign, S.A. (2006). Revisiting the glyoxylate cycle: Alternate pathways for microbial acetate assimilation. Mol. Microbiol. *61*, 274–276.

Ervens, B., and Amato, P. (2020). The global impact of bacterial processes on carbon mass. Atmos. Chem. Phys. Discuss. *20*, 1–25.

Evans, S.E., Dueker, M.E., Logan, J.R., and Weathers, K.C. (2019). The biology of fog: results from coastal Maine and Namib Desert reveal common drivers of fog microbial composition. Sci. Total Environ. *647*, 1547–1556.

Fankhauser, A.M., Antonio, D.D., Krell, A., Alston, S.J., Banta, S., and McNeill, V.F. (2019). Constraining the Impact of Bacteria on the Aqueous Atmospheric Chemistry of Small Organic Compounds. ACS Earth Sp. Chem. *3*, 1485–1491.

Fong, N.J.C., Burgess, M.L., Barrow, K.D., and Glenn, D.R. (2001). Carotenoid accumulation in the psychrotrophic bacterium Arthrobacter agilis in response to thermal and salt stress. Appl. Microbiol. Biotechnol. *56*, 750–756.

Forouzan, E., Shariati, P., Mousavi Maleki, M.S., Karkhane, A.A., and Yakhchali, B. (2018). Practical evaluation of 11 de novo assemblers in metagenome assembly. J. Microbiol. Methods *151*, 99–105.

Fraser, F.C., Corstanje, R., Deeks, L.K., Harris, J.A., Pawlett, M., Todman, L.C., Whitmore, A.P., and Ritz, K. (2016). On the origin of carbon dioxide released from rewetted soils. Soil Biol. Biochem. *101*, 1–5.

Freitas, T.A.K., Li, P.E., Scholz, M.B., and Chain, P.S.G. (2015). Accurate read-based metagenome characterization using a hierarchical suite of unique signatures. Nucleic Acids Res. *43*, e69–e69.

Fröhlich-Nowoisky, J., Pickersgill, D.A., Després, V.R., and Pöschl, U. (2009). High diversity of fungi in air particulate matter. Proc. Natl. Acad. Sci. *106*, 12814.

Fröhlich-Nowoisky, J., Burrows, S.M., Xie, Z., Engling, G., Solomon, P.A., Fraser, M.P., Mayol-Bracero, O.L., Artaxo, P., Begerow, D., Conrad, R., et al. (2012). Biogeography in the air: Fungal diversity over land and oceans. Biogeosciences *9*, 1125–1136.

Fröhlich-Nowoisky, J., Kampf, C.J., Weber, B., Huffman, J.A., Pöhlker, C., Andreae, M.O., Lang-Yona, N., Burrows, S.M., Gunthe, S.S., Elbert, W., et al. (2016). Bioaerosols in the Earth system: Climate, health, and ecosystem interactions. Atmos. Res. *182*, 346–376.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics *28*, 3150–3152.

Fuzzi, S., Mandrioli, P., and Perfetto, A. (1996). Fog droplets - An atmospheric source of secondary biological aerosol particles. Atmos. Environ. *31*, 287–290.

Fuzzi, S., Andreae, M.O., Huebert, B.J., Kulmala, M., Bond, T.C., Boy, M., Doherty, S.J., Guenther, A., Kanakidou, M., Kawamura, K., et al. (2006). Critical assessment of the current state of scientific knowledge, terminology, and research needs concerning the role of organic aerosols in the atmosphere, climate, and global change. Atmos. Chem. Phys. *6*, 2017–2038.

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

Gandolfi, I., Bertolini, V., Bestetti, G., Ambrosini, R., Innocente, E., Rampazzo, G., Papacchini, M., and Franzetti, A. (2015). Spatio-temporal variability of airborne bacterial communities and their correlation with particulate matter chemical composition across two urban areas. Appl. Microbiol. Biotechnol. *99*, 4867–4877.

Garratt, J.R. (1994). The atmospheric boundary layer. Earth-Science Rev. *37*, 89–134.

Ge, Z., Wexler, A.S., and Johnston, M. V (1998). Deliquescence behavior of multicomponent aerosols. J. Phys. Chem. A *102*, 173–180.

Ghobakhlou, A.F., Johnston, A., Harris, L., Antoun, H., and Laberge, S. (2015). Microarray transcriptional profiling of Arctic Mesorhizobium strain N33 at low temperature provides insights into cold adaption strategies. BMC Genomics *16*.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017). Microbiome datasets are compositional: And this is not optional. Front. Microbiol. *8*, 1–6.

de Goffau, M.C., Lager, S., Salter, S.J., Wagner, J., Kronbichler, A., Charnock-Jones, D.S., Peacock, S.J., Smith, G.C.S., and Parkhill, J. (2018). Recognizing the reagent microbiome. Nat. Microbiol. *3*, 851–853.

Gong, F., and Miller, K.M. (2019). Histone methylation and the DNA damage response. Mutat. Res. Mutat. Res. *780*, 37–47.

Goordial, J., Raymond-Bouchard, I., Zolotarov, Y., De Bethencourt, L., Ronholm, J., Shapiro, N., Woyke, T., Stromvik, M., Greer, C.W., Bakermans, C., et al. (2016). Cold adaptive traits revealed by comparative genomic analysis of the eurypsychrophile Rhodococcus sp. JG3 isolated from high elevation McMurdo

Dry Valley permafrost, Antarctica. FEMS Microbiol. Ecol. *92*, 154.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 2011 297 *29*, 644–652.

Grafstrom, R.H., Hamilton, D.L., and Yuan, R. (1984). DNA Methylation: DNA Replication and Repair. 111–126.

Graham, K.E., Prussin, A.J., Marr, L.C., Sassoubre, L.M., and Boehm, A.B. (2018). Microbial community structure of sea spray aerosols at three California beaches. FEMS Microbiol. Ecol. *94*, 5.

Green, R.L., and Warren, G.J. (1985). Physical and functional repetition in a bacterial ice nucleation gene. Nature *317*, 645–648.

Griffin, D.W. (2007). Atmospheric movement of microorganisms in clouds of desert dust and implications for human health. Clin. Microbiol. Rev. *20*, 459–477.

Hamilton, W.D., and Lenton, T.M. (1998). Spora and gaia : how microbes fly with their clouds. Ethol. Ecol. Evol. *10*, 1–16.

Heald, C.L., and Spracklen, D. V. (2009). Atmospheric budget of primary biological aerosol particles from fungal spores. Geophys. Res. Lett. *36*.

Hervàs, A., Camarero, L., Reche, I., and Casamayor, E.O. (2009). Viability and potential for immigration of airborne bacteria from Africa that reach high mountain lakes in Europe. Environ. Microbiol. *11*, 1612–1623.

Hoffmann, L., Günther, G., Li, D., Stein, O., Wu, X., Griessbach, S., Heng, Y., Konopka, P., Müller, R., Vogel, B., et al. (2019). From ERA-Interim to ERA5: The considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. Atmos. Chem. Phys. *19*, 3097–3214.

Hou, P., Wu, S., and McCarty, J. (2017). Sensitivity of atmospheric aerosol scavenging to precipitation intensity and frequency in the context of global climate change. Atmos. Chem. Phys. Discuss. 1–17.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. *47*, D309–D314.

Huffman, J.A., Prenni, A.J., Demott, P.J., Pöhlker, C., Mason, R.H., Robinson, N.H., Fröhlich-Nowoisky, J., Tobo, Y., Després, V.R., Garcia, E., et al. (2013). High concentrations of biological aerosol particles and ice nuclei during and after rain. Atmos. Chem. Phys. *13*, 6151–6164.

Hunt, D.E., Lin, Y., Church, M.J., Karl, D.M., Tringe, S.G., Izzo, L.K., and Johnson, Z.I. (2013). Relationship between abundance and specific activity of bacterioplankton in open ocean surface waters. Appl. Environ. Microbiol. *79*, 177–184.

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics *11*, 1–11.

Imshenetsky, A.A., Lysenko, S. V., and Kazakov, G.A. (1978). Upper boundary of the biosphere. Appl. Environ. Microbiol. *35*, 1–5.

Iovieno, P., and Bååth, E. (2008). Effect of drying and rewetting on bacterial growth rates in soil. FEMS Microbiol. Ecol. *65*, 400–407.

Jaing, C., Thissen, J., Morrison, M., Dillon, M.B., Waters, S.M., Graham, G.T., Be, N.A., Nicoll, P., Verma, S., Caro, T., et al. (2020). Sierra Nevada sweep: metagenomic measurements of bioaerosols vertically distributed across the troposphere. Sci. Rep. *10*, 12399.

Joly, M., Amato, P., Sancelme, M., Vinatier, V., Abrantes, M., Deguillaume, L., and Delort, A.M. (2015). Survival of microbial isolates from clouds toward simulated atmospheric stress factors. Atmos. Environ. *117*, 92–98.

Jones, B.E., Grant, W.D., Duckworth, A.W., Owenson, G.G., Horikoshi, K., Jones, B.E., Grant, W.D., Duckworth, · A W, and Owenson, · G G (1998). Microbial diversity of soda lakes. Extrem. 1998 23 *2*, 191–200.

Joung, Y.S., Ge, Z., and Buie, C.R. (2017). Bioaerosol generation by raindrops on soil. Nat. Commun. *8*, 1–10.

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *28*, 27–30.

Karin, E.L., Mirdita, M., and Söding, J. (2020). MetaEuk-sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. Microbiome *8*, 1–15.

Khaled, A., Zhang, M., Amato, P., Delort, A.M., and Ervens, B. (2021). Biodegradation by bacteria in clouds: An underestimated sink for some organics in the atmospheric multiphase system. Atmos. Chem. Phys. *21*, 3123–3141.

Kieft, T.L., and Ahmadjian, V. (1989). Biological ice nucleation activity in lichen mycobionts and photobionts. Lichenol. *21*, 355–362.

Kieft, T.L., Soroker, E., and Firestone, M.K. (1987). Microbial biomass response to a rapid increase in water potential when dry soil is wetted. Soil Biol. Biochem. *19*, 119–126.

Kim, D., Song, L., Breitwieser, F.P., and Salzberg, S.L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. *26*, 1721–1729.

Klein, A.M., Bohannan, B.J.M., Jaffe, D.A., Levin, D.A., and Green, J.L. (2016). Molecular evidence for metabolically active bacteria in the atmosphere. Front. Microbiol. *7*.

Klingenberg, H., and Meinicke, P. (2017). How to normalize metatranscriptomic count data for

differential expression analysis. PeerJ *2017*.

Kobziar, L.N., Vuono, D., Moore, R., Christner, B.C., Dean, T., Betancourt, D., Watts, A.C., Aurell, J., and Gullett, B. (2022). Wildland fire smoke alters the composition, diversity, and potential atmospheric function of microbial life in the aerobiome. ISME Commun. *2*, 1–9.

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics *28*, 3211–3217.

Krumins, V., Mainelis, G., Kerkhof, L.J., and Fennell, D.E. (2014). Substrate-Dependent rRNA Production in an Airborne Bacterium. Environ. Sci. Technol. Lett. *1*, 376–381.

Ladino, L., Stetzer, O., Hattendorf, B., Günther, D., Croft, B., and Lohmann, U. (2011). Experimental Study of Collection Efficiencies between Submicron Aerosols and Cloud Droplets. J. Atmos. Sci. *68*, 1853–1864.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lavrinienko, A., Jernfors, T., Koskimäki, J.J., Pirttilä, A.M., and Watts, P.C. (2021). Does Intraspecific Variation in rDNA Copy Number Affect Analysis of Microbial Communities? Trends Microbiol. *29*, 19–27.

Lazaridis, M. (2019). Bacteria as Cloud Condensation Nuclei (CCN) in the Atmosphere. Atmosphere (Basel). *10*, 786.

Lee, M., Heikes, B.G., and O'Sullivan, D.W. (2000). Hydrogen peroxide and organic hydroperoxide in the troposphere: A review. Atmos. Environ. *34*, 3475–3494.

Leizeaga, A., Meisner, A., Rousk, J., and Bååth, E. (2022). Repeated drying and rewetting cycles accelerate bacterial growth recovery after rewetting. Biol. Fertil. Soils *58*, 365–374.

Leung, H.C.M., Yiu, S.M., Parkinson, J., and Chin, F.Y.L. (2013). IDBA-MT: De Novo Assembler for Metatranscriptomic Data Generated from Next-Generation Sequencing Technology. Https://Home.Liebertpub.Com/Cmb *20*, 540–550.

Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2014). IDBA-MTP: A Hybrid MetaTranscriptomic Assembler Based on Protein Information. Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) *8394 LNBI*, 160–172.

Li, H., and Durbin, R. (2009). Fast and accurate short-read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659.

Li, D., Liu, C.M., Luo, R., Sadakane, K., and Lam, T.W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics *31*, 1674–1676.

Li, H., Zhou, X.Y., Yang, X.R., Zhu, Y.G., Hong, Y.W., and Su, J.Q. (2019). Spatial and seasonal variation of the airborne microbiome in a rapidly developing city of China. Sci. Total Environ. *665*, 61–68.

Li, X., Chen, H., and Yao, M. (2020). Microbial emission levels and diversities from different land use types. Environ. Int. *143*, 105988.

Lighthart, B. (1997). The ecology of bacteria in the alfresco atmosphere. FEMS Microbiol. Ecol. *23*, 263–274.

Lighthart, B., and Shaffer, B.T. (1995). Viable bacterial aerosol particle size distributions in the midsummer atmosphere at an isolated location in the high desert chaparral. Aerobiologia (Bologna). *11*, 19–25.

Lighthart, B., and Shaffer, B.T. (1997). Increased airborne bacterial survival as a function of particle content and size. Aerosol Sci. Technol. *27*, 439–446.

Lindemann, J., and Upper, C.D. (1985). Aerial Dispersal of Epiphytic Bacteria. Appl. Environ. Microbiol. *50*, 1229–1232.

Lindemann, J., Constantinidou, H.A., Barchet, W.R., and Upper, C.D. (1982). Plants as sources of airborne bacteria, including ice nucleation-active bacteria. Appl. Environ. Microbiol. *44*, 1059–1063.

Lindow, S.E., and Brandl, M.T. (2003). Microbiology of the phyllosphere. Appl. Environ. Microbiol. *69*, 1875–1883.

Lo, C.C., and Chain, P.S.G. (2014). Rapid evaluation and quality control of next generation sequencing data with FaQCs. BMC Bioinformatics *15*, 1–8.

Lofgren, L.A., Uehling, J.K., Branco, S., Bruns, T.D., Martin, F., and Kennedy, P.G. (2019). Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. Mol. Ecol. *28*, 721–730.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 1–21.

Maki, T., Hara, K., Iwata, A., Lee, K.C., Kawai, K., Kai, K., Kobayashi, F., Pointing, S.B., Archer, S., Hasegawa, H., et al. (2017). Variations in airborne bacterial communities at high altitudes over the Noto Peninsula (Japan) in response to Asian dust events. Atmos. Chem. Phys. *17*, 11877–11897.

Manirajan, B.A., Maisinger, C., Ratering, S., Rusch, V., Schwiertz, A., Cardinale, M., and Schnell, S. (2018). Diversity, specificity, co-occurrence and hub taxa of the bacterial-fungal pollen microbiome. FEMS Microbiol. Ecol. *94*, 1–11.

Marinoni, A., Laj, P., Sellegri, K., and Mailhot, G. (2004). Cloud chemistry at the Puy de Dôme: variability

and relationships with environmental factors.

Maron, P.A., Mougel, C., David, D.P., Carvalho, E., Bizet, K., Marck, G., Cubito, N., Lemanceau, P., and Ranjard, L. (2006). Temporal variability of airborne bacterial community structure in an urban area. Atmos. Environ. *40*, 8074–8080.

Matthias-Maser, S., Gruber, S., and Jaenicke, R. (2000). The size distribution of primary biological aerosol particles in cloud water on the mountain Kleiner Feldberg/Taunus (FRG). Atmos. Res. *54*, 1–13.

Medinger, R., Nolte, V., Pandey, R.V., Jost, S., Ottenwälder, B., Schlötterer, C., and Boenigk, J. (2010). Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Mol. Ecol. *19*, 32–40.

Menke, S., Gillingham, M.A.F., Wilhelm, K., and Sommer, S. (2017). Home-made cost effective preservation buffer is a better alternative to commercial preservation methods for microbiome research. Front. Microbiol. *8*.

Michaud, J.M., Thompson, L.R., Kaul, D., Espinoza, J.L., Richter, R.A., Xu, Z.Z., Lee, C., Pham, K.M., Beall, C.M., Malfatti, F., et al. (2018). Taxon-specific aerosolization of bacteria and viruses in an experimental ocean-atmosphere mesocosm. Nat. Commun. *9*.

Mikhailov, E.F., Pöhlker, M.L., Reinmuth-Selzle, K., Vlasenko, S.S., Krüger, O.O., Fröhlich-Nowoisky, J., Pöhlker, C., Ivanova, O.A., Kiselev, A.A., Kremper, L.A., et al. (2021). Water uptake of subpollen aerosol particles: Hygroscopic growth, cloud condensation nuclei activation, and liquid-liquid phase separation. Atmos. Chem. Phys. *21*, 6999–7022.

Miles, N.L., Verlinde, J., and Clothiaux, E.E. (2002). Cloud Droplet Size Distributions in Low-Level Stratiform Clouds. J. Atmos. Sci. *57*, 295–311.

Mircea, M., Stefan, S., and Fuzzi, S. (2000). Precipitation scavenging coefficient: influence of measured aerosol and raindrop size distributions. Atmos. Environ. *34*, 5169–5174.

Möhler, O., DeMott, P.J., Vali, G., and Levin, Z. (2007). Microbiology and atmospheric processes: The role of biological particles in cloud physics. Biogeosciences *4*, 1059–1071.

Moore, D., Robson, G.D., and Trinci, A.P.J. (2011). 21st Century Guidebook to Fungi (Cambridge University Press).

Moore, R.A., Hanlon, R., Powers, C., Schmale, D.G., and Christner, B.C. (2020a). Scavenging of Sub-Micron to Micron-Sized Microbial Aerosols during Simulated Rainfall. Atmos. Chem. Phys 1–13.

Moore, R.A., Bomar, C., Kobziar, L.N., and Christner, B.C. (2020b). Wildland fire as an atmospheric source of viable microbial aerosols and biological ice nucleating particles. ISME J.

Morita, R.Y. (1975). Psychrophilic Bacteria. Am. Soc. Microbiol. *39*, 144–167.

Morris, C.E., Georgakopoulos, D.G., and Sands, D.C. (2004). Ice nucleation active bacteria and their potential role in precipitation. J. Phys. IV *121*, 87–103.

Morris, C.E., Sands, D.C., Vinatzer, B.A., Glaux, C., Guilbaud, C., Buffière, A., Yan, S., Dominguez, H., and Thompson, B.M. (2008). The life history of the plant pathogen Pseudomonas syringae is linked to the water cycle. ISME J. *2*, 321–334.

Mueller, D.R., Vincent, W.F., Bonilla, S., and Laurion, I. (2005). Extremotrophs, extremophiles and broadband pigmentation strategies in a high arctic ice shelf ecosystem. FEMS Microbiol. Ecol. *53*, 73–87.

Mykytczuk, N.C., Foote, S.J., Omelon, C.R., Southam, G., Greer, C.W., and Whyte, L.G. (2013). Bacterial growth at -15 °C; molecular insights from the permafrost bacterium Planococcus halocryophilus Or1. ISME J. *7*, 1211–1226.

Noguchi, H., Taniguchi, T., and Itoh, T. (2008). MetaGeneAnnotator: Detecting Species-Specific Patterns of Ribosomal Binding Site for Precise Gene Prediction in Anonymous Prokaryotic and Phage Genomes. DNA Res. *15*, 387–396.

Noirmain, F., Baray, J., Tridon, F., Cacault, P., Billard, H., Voyard, G., Baelen, J. Van, and Latour, D. (2022). Interdisciplinary strategy to survey phytoplankton dynamics of a eutrophic lake under rain forcing: description of the instrumental set-up and first results. Biogeosciences.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome Res. *27*, 824–834.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745.

Park, C., Shin, B., and Park, W. (2019). Alternative fate of glyoxylate during acetate and hexadecane metabolism in Acinetobacter oleivorans DR1. Sci. Rep. *9*, 1–12.

Pascault, N., Loux, V., Derozier, S., Martin, V., Debroas, D., Maloufi, S., Humbert, J.F., and Leloup, J. (2015). Technical challenges in metatranscriptomic studies applied to the bacterial communities of freshwater ecosystems. Genetica *143*, 157–167.

Pasteur, L., Chamberland, C., and Joubert, J. (1878). Théorie des germes et ses applications à la médecine et à la chirurgie. *7*, 107–115.

Péguilhan, R., Besaury, L., Rossi, F., Enault, F., Baray, J., Deguillaume, L., and Amato, P. (2021). Rainfalls sprinkle cloud bacterial diversity while scavenging biomass. FEMS Microbiol. Ecol. 1–15.

Pelizzola, M., and Ecker, J.R. (2011). The DNA methylome. FEBS Lett. *585*, 1994–2000.

Peng, Y., Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics *28*, 1420–1428.

Peng, Y., Leung, H.C.M., Yiu, S.M., Lv, M.J., Zhu, X.G., and Chin, F.Y.L. (2013). IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. Bioinformatics *29*, i326–i334.

Petrenchuk, O.P., and Drozdova, V.M. (1966). On the chemical composition of cloud water. Tellus *18*, 280–286.

Petters, M.D., and Kreidenweis, S.M. (2007). A single parameter representation of hygroscopic growth and cloud condensation nucleus activity. Atmos. Chem. Phys. *7*, 1961–1971.

Prass, M., Andreae, M.O., Araùjo, A.C. De, Artaxo, P., and Ditas, F. (2021). Bioaerosols in the Amazon rain forest: Temporal variations and vertical profiles of Eukarya, Bacteria and Archaea. Biogeosciences Discuss. 1–23.

Pruppacher, H.R., and Jaenicke, R. (1995). The processing of water vapor and aerosols by atmospheric clouds, a global estimate. Atmos. Res. *38*, 283–295.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. *41*, 590–596.

Radke, L.F., Hobbs, P. V., and Eltgroth, M.W. (1980). Scavenging of aerosol particles by precipitation. Am. Meteorol. Soc. *19*, 715–722.

Rastogi, G., Coaker, G.L., and Leveau, J.H.J. (2013). New insights into the structure and function of phyllosphere microbiota through high-throughput molecular approaches. FEMS Microbiol. Lett. *348*, 1–10.

Reche, I., D'Orta, G., Mladenov, N., Winget, D.M., and Suttle, C.A. (2018). Deposition rates of viruses and bacteria above the atmospheric boundary layer. ISME J. *12*, 1154–1162.

Renard, P., Bianco, A., Baray, J.L., Bridoux, M., Delort, A.M., and Deguillaume, L. (2020). Classification of clouds sampled at the puy de Dôme station (France) based on chemical measurements and air mass history matrices. Atmosphere (Basel). *11*, 732.

Rho, M., Tang, H., and Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res. *38*, e191–e191.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47–e47.

Robinson, C.H. (2001). Cold adaptation in Arctic and Antarctic fungi. New Phytol. *151*, 341–353.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Romano, S., Di Salvo, M., Rispoli, G., Alifano, P., Perrone, M.R., and Talà, A. (2019). Airborne bacteria in the Central Mediterranean: Structure and role of meteorology and air mass transport. Sci. Total Environ. *697*.

Rothschild, L.J., and Mancinelli, R.L. (2001). Life in extreme environments. Nature *409*, 1092–1101.

Ruiz-Gil, T., Acuña, J.J., Fujiyoshi, S., Tanaka, D., Noda, J., Maruyama, F., and Jorquera, M.A. (2020). Airborne bacterial communities of outdoor environments and their associated influencing factors. Environ. Int. *145*, 106156.

Sajjad, W., Din, G., Rafiq, M., Iqbal, A., Khan, S., Zada, S., Ali, B., and Kang, S. (2020). Pigment production by cold-adapted bacteria and fungi: colorful tale of cryosphere with wide range applications. Extremophiles *24*, 447–473.

Salazar, G., Paoli, L., Alberti, A., Huerta-Cepas, J., Ruscheweyh, H.J., Cuenca, M., Field, C.M., Coelho, L.P., Cruaud, C., Engelen, S., et al. (2019). Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. Cell *179*, 1068-1083.e21.

Samaké, A., Bonin, A., Jaffrezo, J.L., Taberlet, P., Weber, S., Uzu, G., Jacob, V., Conil, S., and Martins, J.M.F. (2020). High levels of primary biogenic organic aerosols are driven by only a few plant-associated microbial taxa. Atmos. Chem. Phys. *20*, 5609–5628.

Sánchez-Romero, M.A., Cota, I., and Casadesús, J. (2015). DNA methylation in bacteria: From the methyl group to the methylome. Curr. Opin. Microbiol. *25*, 9–16.

Sands D.C (1982). The association between bacteria and rain and possible resultant meteorological implications.

Šantl-Temkiv, T., Finster, K., Hansen, B.M., Pašić, L., and Karlson, U.G. (2013). Viable methanotrophic bacteria enriched from air and rain can oxidize methane at cloud-like conditions. Aerobiologia (Bologna). *29*, 373–384.

Šantl-Temkiv, T., Amato, P., Gosewinkel, U., Thyrhaug, R., Charton, A., Chicot, B., Finster, K., Bratbak, G., and Löndahl, J. (2017). High-Flow-Rate Impinger for the Study of Concentration, Viability, Metabolic Activity, and Ice-Nucleation Activity of Airborne Bacteria. Environ. Sci. Technol. *51*, 11224–11234.

Šantl-Temkiv, T., Gosewinkel, U., Starnawski, P., Lever, M., and Finster, K. (2018). Aeolian dispersal of bacteria in southwest Greenland: Their sources, abundance, diversity and physiological states. FEMS Microbiol. Ecol. *94*, 1–10.

Šantl-Temkiv, T., Sikoparija, B., Maki, T., Carotenuto, F., Amato, P., Yao, M., Morris, C.E., Schnell, R., Jaenicke, R., Pöhlker, C., et al. (2020). Bioaerosol field measurements: Challenges and perspectives in outdoor studies. Aerosol Sci. Technol. *54*, 520–546.

Šantl-Temkiv, T., Amato, P., Casamayor, E.O., Lee, P.K.H., and Pointing, S.B. (2022). Microbial ecology of the atmosphere. FEMS Microbiol. Rev.

Sattler, B., Puxbaum, H., and Psenner, R. (2001). Bacterial growth in supercooled cloud droplets. Geophys. Res. Lett. *28*, 239–242.

Schleper, C., Puehler, G., Holz, I., Gambacorta, A., Janekovic, D., Santarius, U., Klenk, H.P., and Zillig, W. (1995). Picrophilus gen. nov., fam. nov.: a novel aerobic, heterotrophic, thermoacidophilic genus and family comprising archaea capable of growth around pH 0. J. Bacteriol. *177*, 7050–7059.

Sellegri, K., Laj, P., Marinoni, A., Dupuy, R., Legrand, M., and Preunkert, S. (2003). Contribution of gaseous and particulate species to droplet solute composition at the Puy de Dôme, France. Atmos. Chem. Phys. *3*, 1509–1522.

Shakya, M., Lo, C.C., and Chain, P.S.G. (2019). Advances and challenges in metatranscriptomic analysis. Front. Genet. *10*.

Slekar, K.H., Kosman, D.J., and Culotta, V.C. (1996). The Yeast Copper/Zinc Superoxide Dismutase and the Pentose Phosphate Pathway Play Overlapping Roles in Oxidative Stress Protection *. J. Biol. Chem. *271*, 28831–28836.

Smets, W., Moretti, S., Denys, S., and Lebeer, S. (2016). Airborne bacteria in the atmosphere: Presence, purpose, and potential. Atmos. Environ. *139*, 214–221.

Smith, D.J., Timonen, H.J., Jaffe, D.A., Griffin, D.W., Birmele, M.N., Perry, K.D., Ward, P.D., and Roberts, M.S. (2013). Intercontinental dispersal of bacteria and archaea by transpacific winds. Appl. Environ. Microbiol. *79*, 1134–1139.

Smith, D.J., Ravichandar, J.D., Jain, S., Griffin, D.W., Yu, H., Tan, Q., Thissen, J., Lusby, T., Nicoll, P., Shedler, S., et al. (2018). Airborne bacteria in earth's lower stratosphere resemble taxa detected in the troposphere: Results from a new NASA Aircraft Bioaerosol Collector (ABC). Front. Microbiol. *9*, 1–20.

Sonwani, S., and Kulshrestha, U.C. (2019). PM10 carbonaceous aerosols and their real-time wet scavenging during monsoon and non-monsoon seasons at Delhi, India. J. Atmos. Chem. *76*, 171–200.

Stoddard, S.F., Smith, B.J., Hein, R., Roller, B.R.K., and Schmidt, T.M. (2015). rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Res. *43*, D593–D598.

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One *6*, 21800.

The UniProt Consortium (2019). UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res. *47*, D506–D515.

Tignat-perrier, R., Dommergue, A., Thollot, A., Keuschnig, C., Magand, O., Vogel, T.M., and Larose, C. (2019). Global airborne microbial communities controlled by surrounding landscapes and wind conditions. Sci. Rep. 1–11.

Tignat-Perrier, R., Dommergue, A., Thollot, A., Magand, O., Amato, P., Joly, M., Sellegri, K., Vogel, T.M., and Larose, C. (2020). Seasonal shift in airborne microbial communities. Sci. Total Environ. 137129.

Till, A., Lakhani, R., Burnett, S.F., and Subramani, S. (2012). Pexophagy: The Selective Degradation of Peroxisomes. Int. J. Cell Biol. *2012*.

Triadó-Margarit, X., Caliz, J., Reche, I., and Casamayor, E.O. (2019). High similarity in bacterial bioaerosol compositions between the free troposphere and atmospheric depositions collected at high-elevation mountains. Atmos. Environ. 79–86.

Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., and Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat. Methods *12*, 902–903.

Vaïtilingom, M., Amato, P., Sancelme, M., Laj, P., Leriche, M., and Delort, A.M. (2010). Contribution of microbial activity to carbon chemistry in clouds. Appl. Environ. Microbiol. *76*, 23–29.

Vaïtilingom, M., Deguillaume, L., Vinatier, V., Sancelme, M., Amato, P., Chaumerliac, N., and Delort, A.-M. (2013). Potential impact of microbial activity on the oxidant capacity and organic carbon budget in clouds. Proc. Natl. Acad. Sci. *110*, 559–564.

Vinatier, V., Wirgot, N., Joly, M., Sancelme, M., Abrantes, M., Deguillaume, L., and Delort, A.M. (2016). Siderophores in cloud waters and potential impact on atmospheric chemistry: Production by microorganisms isolated at the puy de Dôme station. Environ. Sci. Technol. *50*, 9315–9323.

Wex, H., Stratmann, F., Topping, D., and McFiggans, G. (2008). The Kelvin versus the raoult term in the köhler equation. J. Atmos. Sci. *65*, 4004–4016.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998). Prokaryotes: The unseen majority. Proc. Natl. Acad. Sci. U. S. A. *95*, 6578–6583.

Wilhelm, L., Besemer, K., Fasching, C., Urich, T., Singer, G.A., Quince, C., and Battin, T.J. (2014). Rare but active taxa contribute to community dynamics of benthic biofilms in glacier-fed streams. Environ. Microbiol. *16*, 2514–2524.

Willis, P.T., and Tattelman, P. (1989). Drop-size distribution associated with intense rainfall. J. Appl. Meteorol. *28*, 3–15.

Wirgot, N., Vinatier, V., Deguillaume, L., Sancelme, M., and Delort, A.M. (2017). H2O2 modulates the energetic metabolism of the cloud microbiome. Atmos. Chem. Phys. *17*, 14841–14851.

Womack, A.M., Bohannan, B.J.M., and Green, J.L. (2010). Biodiversity and biogeography of the atmosphere. Philos. Trans. R. Soc. B Biol. Sci. *365*, 3645–3653.

Womack, A.M., Artaxo, P.E., Ishida, F.Y., Mueller, R.C., Saleska, S.R., Wiedemann, K.T., Bohannan, B.J.M., and Green, J.L. (2015). Characterization of active and total fungal communities in the atmosphere over the Amazon rainforest. Biogeosciences *12*, 6337–6349.

Woo, C., and Yamamoto, N. (2020). Falling bacterial communities from the atmosphere. Environ. Microbiomes *15*, 22.

Wood, D.E., and Salzberg, S.L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. Genome Biol. *15*, 1–12.

Xu, C., Wei, M., Chen, J., Zhu, C., Li, J., Lv, G., Xu, X., Zheng, L., Sui, G., Li, W., et al. (2017). Fungi diversity in PM2.5 and PM1 at the summit of Mt. Tai: Abundance, size distribution, and seasonal variation. Atmos. Chem. Phys. *17*, 11247–11260.

Yan, D., Zhang, T., Su, J., Zhao, L.L., Wang, H., Fang, X.M., Zhang, Y.Q., Liu, H.Y., and Yu, L.Y. (2016). Diversity and composition of airborne fungal community associated with particulate matters in Beijing during haze and non-haze days. Front. Microbiol. *7*, 487.

Yang, K., Li, L., Wang, Y., Xue, S., Han, Y., and Liu, J. (2018). Airborne bacteria in a wastewater treatment plant: Emission characterization, source analysis and health risk assessment. Water Res. *149*, 596–606.

Ye, Y., and Tang, H. (2016). Utilizing de Bruijn graph of metagenome assembly for metatranscriptome analysis. Bioinformatics *32*, 1001–1008.

Ye, L., Wu, X., Tan, X., Shi, X., Li, D., Yu, Y., Zhang, M., and Kong, F. (2010). Cell Lysis of Cyanobacteria and Its Implications for Nutrient Dynamics. Int. Rev. Hydrobiol. *95*, 235–245.

Yilmaz, S., Allgaier, M., and Hugenholtz, P. (2010). Multiple displacement amplification compromises quantitative analysis of metagenomes. Nat. Methods *7*, 943–944.

Zhang, M., Khaled, A., Amato, P., Delort, A.M., and Ervens, B. (2021a). Sensitivities to biological aerosol particle properties and ageing processes: Potential implications for aerosol-cloud interactions and optical properties. Atmos. Chem. Phys. *21*, 3699–3724.

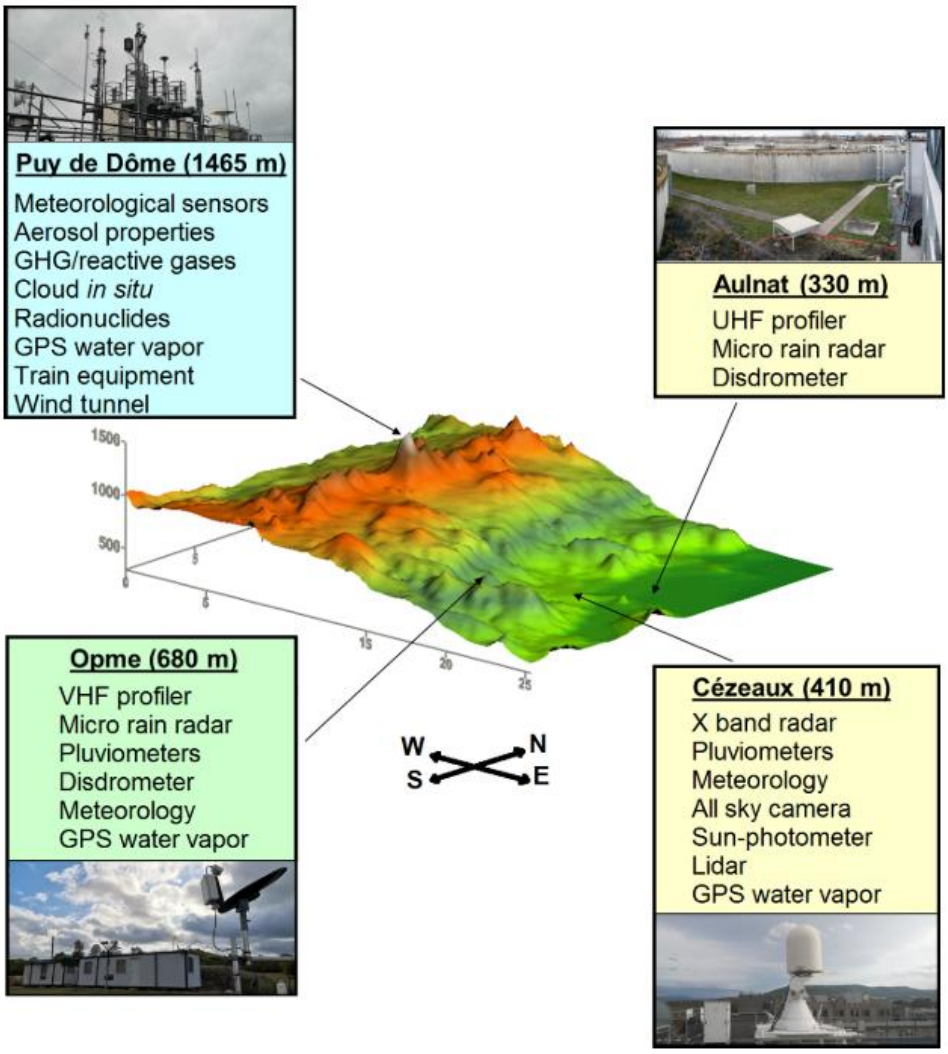
Zhang, N., Castlebury, L.A., Miller, A.N., Huhndorf, S.M., Schoch, C.L., Seifert, K.A., Rossman, A.Y., Rogers, J.D., Kohlmeyer, J., Volkmann-Kohlmeyer, B., et al. (2007). An overview of the systematics of the Sordariomycetes based on a four-gene phylogeny. Mycologia *98*, 1076–1087.

Zhang, Y., Zhao, Z., Dai, M., Jiao, N., and Herndl, G.J. (2014). Drivers shaping the diversity and biogeography of total and active bacterial communities in the South China Sea. Mol. Ecol. *23*, 2260–2274.

Zhang, Y., Thompson, K.N., Huttenhower, C., and Franzosa, E.A. (2021b). Statistical approaches for differential expression analysis in metatranscriptomics. Bioinformatics *37*, I34–I41.

Annexes

<u>Annex 1</u>: Instruments and measurement sites of the CO-PDD atmospheric research station (from Baray et al., 2020).



Annex 2: Oral and poster presentations.

Oral:

EGU General Assembly 2020 (online discussion): "Partitioning of microbial cells between clouds and precipitation"; Raphaëlle Péguilhan, Ludovic Besaury, Florent Rossi, Jean-Luc Baray, Thibaud Mas, Laurent Deguillaume, Barbara Ervens, and Pierre Amato.
Online Abstract: https://meetingorganizer.copernicus.org/EGU2020/EGU2020-2876.html

17ème rencontre des Microbiologistes du Pôle Clermontois 2021 (17th meeting of microbiologists from Clermont-Ferrand) (online presentation): "Les bactéries dans le cycle de l'eau atmosphérique, des nuages aux précipitations"; Raphaëlle Péguilhan, Ludovic Besaury, Florent Rossi, François Enault, Jean-Luc Baray, Laurent Deguillaume, and Pierre Amato.

Journée AuBi Bioinformatique-NGS 2021 (Bioinformatics day for NGS of the AuBi network) (oral presentation): "Développement d'un workflow pour l'analyse de données métagénomiques et métatranscriptomiques dans un contexte environnemental"; Raphaëlle Péguilhan, Florent Rossi, Engy Nasr, Bérénice Batut, Laurent Deguillaume, François Enault, and Pierre Amato.

Séminaire au LMGE le 12 juillet 2022 (seminar at LMGE) : « Les nuages : Oasis de l'atmosphère » ; présentation des résultats majeurs de mon travail de thèse / *presentation of the major results of my thesis work.*

Poster :

World Microbe Forum 2021 (ASM and FEMS Collaboration) (Online interactive Poster): "Clouds as atmospheric oases for microorganisms"; Raphaëlle Péguilhan, Florent Rossi, François Enault, Jean-Luc Baray, Laurent Deguillaume, and Pierre Amato.
Online Poster : https://wmf2021-asm.ipostersessions.com/?s=D3-C6-3F-87-20-C4-6F-AD-60-AB-84-BB-AB-3A-08-4E

JOBIM 2021 (Open Days in Biology, Computer Science and Mathematics) (Online Poster): "Metagenomic and metatranscriptomic analysis for the study of clouds and aerosols"; Raphaëlle Péguilhan, Florent Rossi, François Enault, Laurent Deguillaume, and Pierre Amato.

Résumé étendu en français

1. Contexte

Les microorganismes sont connus depuis longtemps pour être présents en suspension dans l'atmosphère (Lighthart, 1997; Lighthart and Shaffer, 1995). On les retrouve même à très haute altitude (Bowers et al., 2009; Smith et al., 2018) jusque dans les nuages (Amato et al., 2005; Sattler et al., 2001). Ils peuvent ainsi être transportés sur de longues distances à l'échelle régionale et continentale (Smith et al., 2013). La biomasse microbienne présente dans l'atmosphère est de l'ordre de 10^3 à 10^6 cellules par m^3 d'air (Bauer et al., 2002; Burrows et al., 2009) suivant la proximité avec la source d'émission, ce qui en fait un des environnements les plus pauvres en biomasse. Les sources d'émission de microorganismes dans l'atmosphère sont très diverses et variées et sont principalement les sols nus et la végétation (Després et al., 2012). Ces sources vont être mixées sur de longues distances et former des mosaïques complexes et variables de microorganismes dans l'atmosphère. Les groupes taxonomiques bactériens communément retrouvés dans l'atmosphère sont les Protéobactéries, Actinobactéries et Bactéroidetes, dont de nombreux genres bactériens associés à la phyllosphère comme *Pseudomonas*, *Sphingomonas*, *Methylobacterium* et *Massilia* (Bowers et al., 2011; Burrows et al., 2009; Thompson et al., 2017). Concernant les champignons, ceux sont principalement des Ascomycètes et des Basidiomycètes que l'on retrouve (Fröhlich-Nowoisky et al., 2009; Womack et al., 2015).

Les microorganismes présents dans l'atmosphère vont être exposés à des conditions extrêmes, tel qu'une forte dessiccation, des températures très basses, l'exposition aux UVs et radicaux libres (stress oxydatif). Lorsqu'ils vont intégrer les gouttelettes de nuages ceux-ci vont être protégés contre la dessiccation, mais exposés aux chocs osmotiques et cycles de gel-dégel (Joly et al., 2015; Smets et al., 2016) (**Figure 1**). Malgré ces conditions considérées comme extrêmes, des cellules microbiennes ont été détectées comme viables et actives dans les aérosols (l'atmosphère sèche) et les nuages (Amato et al., 2017, 2019; Krumins et al., 2014; Šantl-Temkiv et al., 2018). Les microorganismes dans les nuages pourraient donc impacter la chimie et la physique des nuages par leur activité métabolique (Khaled et al., 2021; Möhler et al., 2007; Wirgot et al., 2017). Des bactéries sont même soupçonnées d'être capable d'induire des précipitations, ce qui impacterait entre autres la durée de vie du nuage (Morris et al., 2004, 2010). Cependant, encore peu de choses sont connus sur les activités exprimées et les potentiels spécificités métaboliques de ces communautés aéroportées. Ces processus sont également d'une grande importance à étudier pour comprendre l'impact de ces communautés atmosphériques disséminées avec les précipitations sur les écosystèmes de surface (Noirmain et al., 2022) et sur l'écologie microbienne des cellules aéroportées. La diversité microbienne atteignant les sols n'est pas seulement le résultat de la diversité présente dans les nuages, mais aussi des cellules lessivées (« Wash-out » ou « scavenging ») par les précipitations dans la colonne d'air (**Figure 1**).

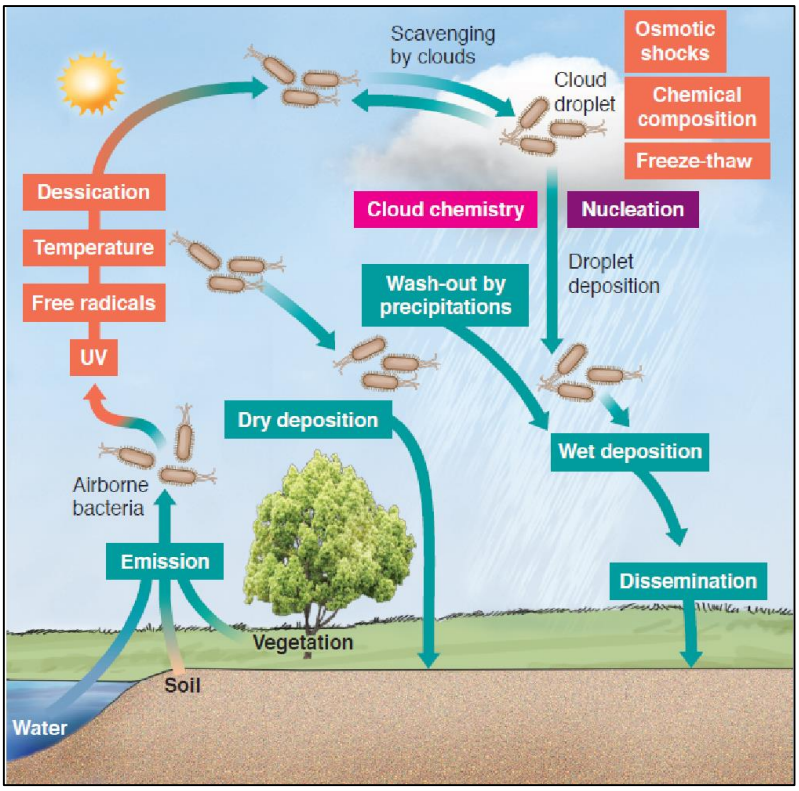


Figure 1 : Schéma du cycle des microorganismes dans l'atmosphère. Amato 2012: Clouds provide atmospheric oasis for microbes. Microbe Magazine 7, 3, 119-123

2. Objectifs

C'est dans ce contexte que nous nous sommes intéressés à la particularité des nuages en tant que potentielles oasis pour les microorganismes aéroportés au milieu d'un désert atmosphérique. Dans ce but, nous avons comparé, tout d'abord, les nuages aux précipitations en termes de diversité bactérienne, puis les nuages aux aérosols en termes de diversité microbienne et de profile fonctionnel.

Pour les profils de biodiversité, c'est le séquençage de régions variables du gène de l'ARNr 16S (métagénétique) qui a été utilisé, avec des analyses complémentaires biologiques (quantification ATP et nombre total de cellules), chimiques (quantification des ions) et météorologiques (vitesse et direction du vent, température, …). Concernant les profils fonctionnels, la métatranscriptomique (MT) couplée à la métagénomique (MG) ont été utilisées, avec pour objectif de ne pas recourir à des étapes de pré-amplification (qui pourraient biaiser l'aspect quantitatif). Pour atteindre notre objectif, dans ce contexte atmosphérique où la biomasse est faible, plusieurs améliorations et validations des protocoles d'échantillonnage et de traitement des échantillons ont été réalisés.

Cette étude a nécessité de multiples collaborations entre plusieurs disciplines (biologie, chimie et physique) et a demandé diverses compétences et connaissances tel que : l'échantillonnage de terrain, la biologie moléculaire, la bioinformatique, mais aussi de la chimie et physique de l'atmosphère.

Ce travail de thèse est découpé en 4 parties résumées ci-après :

- Mise en place de la méthodologie (contrôles et validation des protocoles)
- Comparaison de la diversité bactérienne entre la pluie et les nuages
- Comparaison de la diversité bactérienne entre les aérosols et les nuages
- Comparaison du potentiel fonctionnel des communautés microbiennes dans les aérosols et les nuages

3. Matériels et méthodes

3.1.L'échantillonnage

L'échantillonnage s'est effectué en collaboration avec l'OPGC en utilisant les stations météorologiques du réseau Cézeaux-Aulnat-Opme-Puy de Dôme (CO-PDD) (Baray et al., 2020). L'échantillonnage de nuages et d'aérosols s'est effectué à la station PUY (puy de Dôme, France ; 1 465 m), tandis que l'échantillonnage de pluie a été réalisé à la station OPM (Opme, France ; 660 m) (Péguilhan et al., 2021).

Le High-Flow-Rate impinger (HFRi) a été choisi pour l'échantillonnage des nuages et des aérosols. Il s'agit d'un aspirateur commercial Kärcher DS5600 ou DS6 (Kärcher SAS ; Bonneuil sur Marne, France), pouvant contenir jusqu'à 2 litres de liquide de collecte et ayant un débit d'aspiration de 2 m^3 m^{-1} (Šantl-Temkiv et al., 2017). Cet échantillonneur à l'avantage de pouvoir utiliser un liquide de préservation des acides nucléiques (NAP buffer) comme liquide de collecte, d'avoir un débit d'aspiration assez élevé pour collecter suffisamment de biomasse en un temps court (entre 2 et 6 heures de collecte) et de pouvoir être utilisé à la fois pour les nuages et les aérosols, limitant ainsi les biais liés au collecteur. Trois HFRi dédiés à l'analyse des acides nucléiques, et contenant du NAP buffer, étaient utilisés en simultané, avec en supplément un autre HFRi dédié aux analyses complémentaires et contenant de l'H_2O (**Figure 2**). Un blanc de collecte été réalisé avant chaque échantillonnage en laissant du NAP buffer 10 min dans une cuve de HFRi. Le protocole de décontamination des cuves d'HFRi a également été testé en contaminant volontairement les cuves avec de l'ATP ou une souche bactérienne (*Pseudomonas syringae* 32b-74).

La pluie était collectée avec un échantillonneur automatique réfrigéré (NSA 181/KHS ; Eigenbrodt ; Königsmoor, Germany) (Pouzet et al., 2017) pendant des créneaux de 24 h.

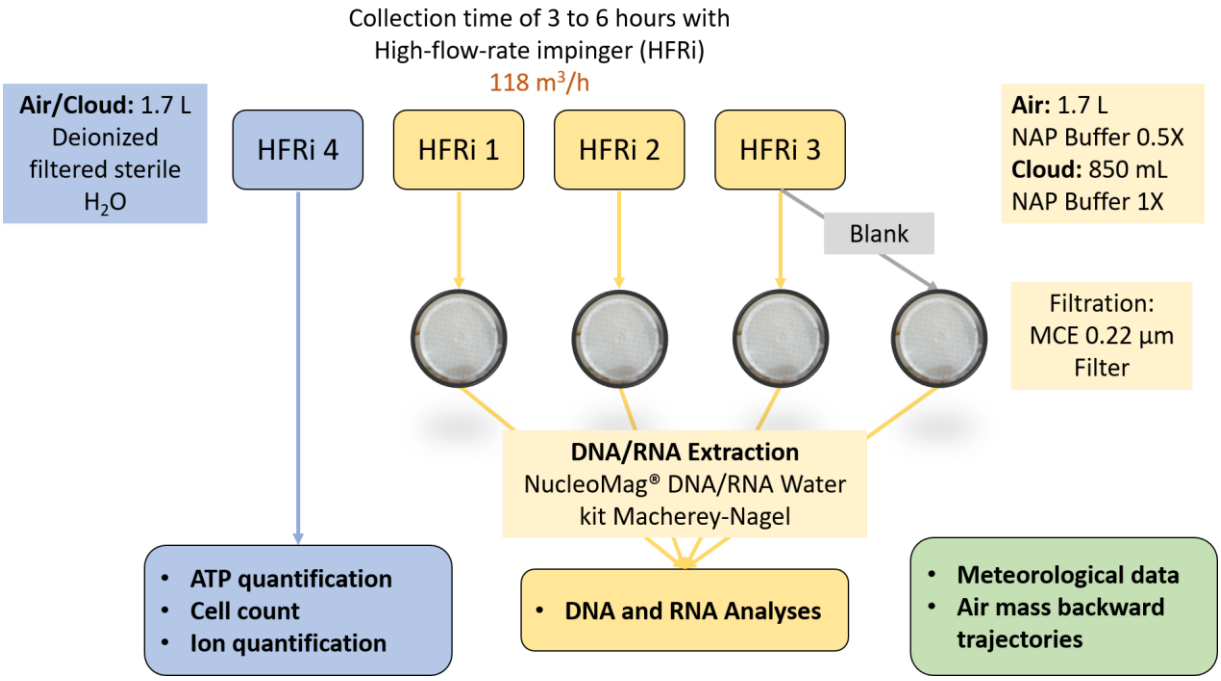


Figure 2 : Procédure d'échantillonnage des aérosols et des nuages avec le High-Flow-Rate impinger (HRFi) comme collecteur. NAP buffer : Nucleic Acid Preservation buffer ; MCE : Mixed Cellulose Esters; Air : aérosols.

3.2.L'extraction des acides nucléiques

Trois kits d'extraction des acides nucléiques ont été comparés. C'est finalement le kit NucleoMag® DNA/RNA Water kit (Macherey-Nagel) qui a été retenu car les quantités extraites d'ADN étaient égales ou supérieures aux deux autres kits et il permettait d'extraire à la fois l'ADN et l'ARN.

Les échantillons et les blancs étaient filtrés sur des membranes en Mixed Cellulose Ester (MCE) de porosité 0.22 µm directement après l'échantillonnage et séparément pour chacun des 3 HRFi (réplicas d'échantillonnage) (**Figure 2**). Les filtres étaient conservés à -80°C avec du buffer de lyse dans les tubes de billes du kit NucleoMag, avant de procéder à l'extraction des acides nucléiques. Nous avions donc pour chaque échantillon un triplicata d'extraits ADN et ARN.

3.3.Le séquençage (métagénétique, MG et MT)

Dans le cadre de l'étude de comparaison de la diversité bactérienne dans les nuages et la pluie, les régions V5, V6 et V7 du gène de l'ARNr 16S ont été amplifiées par PCR (primers 799f-1193r). Les réplicas ont été poolés par échantillon avant l'amplification PCR. Pour la comparaison de la diversité bactérienne dans les nuages et les aérosols, c'est la région variable V4 qui a été amplifiée (515f-806r) et chaque réplica a été séquencé. Le séquençage a été réalisé par Illumina MiSeq (2x250 pb) dans les deux cas (GenoScreen ; Lille, France).

Pour les MG et MT, les ADN et ADNc (ARN retro-transcrit) totaux ont été poolés pour un même échantillon et séquencés directement, sans amplification au préalable, par Illumina HiSeq (2x150 pb) (GenoScreen ; Lille, France). Dans le cas des ARN, les ADNc ont dû être pré-concentrés avant la préparation des librairies de séquençage.

3.4. Le traitement des données de séquençage métagénétiques (profil de biodiversité)

Les données de séquençage nuages-pluie ont été traitées avec la plateforme Galaxy du service bioinformatique AuBi (Auvergne Bioinformatique) et les ressources de calcul du Mesocentre Clermont Auvergne. Les séquences ont été regroupées en OTU (Unité Taxonomique Opérationnelle) et affiliées avec FROGS (Bernard et al., 2021; Escudié et al., 2018) et la base de données Silva 16S v 132 (Quast et al., 2013). Les données nuages-aérosols, quant à elles, ont été traitées jusqu'à l'affiliation par l'équipe MIGALE de l'INRAE (Olivier Rué) dans le cadre d'une collaboration. Les séquences ont été regroupées en ASV (Variant de Séquence d'Amplicon) avec DADA2 (Callahan et al., 2016) et affiliées avec FROGS et la base de données Silva 16S v 138.1. Pour ces deux études, les données ont ensuite été traitées en tant que données de type compositionnel comme décrit dans Gloor *et al*. (2017).

3.5. Contrôle qualité du séquençage d'amplicon

Pour s'assurer de la qualité du séquençage des profils de biodiversité, des biais de quantification et contaminants potentiels, des communautés synthétiques (mock) ont été construites à partir de 6 souches bactériennes à des concentrations connues, et traitées comme les échantillons de l'étude nuages-aérosols (amplification de la région V4 de l'ARNr 16S et regroupement en ASV). Les 6 souches utilisées étaient : *Pseudomonas syringae* 32b-74 (GenBank ID : HQ256872), *Bacillus sp.* 5b-1 (DQ512749), *Sphingomonas sp.* 32b-11 (HQ256831), *Rhodococcus enclensis* 23b-28 (DOVD00000000), *Staphylococcus equorum* 5b-16 (DQ512761) et *Flavobacterium sp.* 57b-18 (KR922118.1) (Amato et al., 2007; Lallement et al., 2017; Vaïtilingom et al., 2012).

3.6. Le développement d'un workflow pour le traitement des données de séquençage MG et MT

Plusieurs difficultés sont liées au traitement des données de séquençage MG et MT, surtout pour le traitement des MT. Il n'existe pas de protocole standard de traitement et chaque étude à tendance à adapter la méthode aux contraintes de son jeu de données. Néanmoins, certains workflows informatiques publiques existent mais, soit n'ont pas de bases de données adaptées à notre jeu de données environnemental, soit ne proposent pas de traiter les MG et MT en même temps. Pour ces

raisons, nous avons développé notre propre workflow informatique à partir d'outils existants (**Figure 3**). Le workflow a été développé sur la plateforme Galaxy Aubi (avec les ressources du Mesocentre) et a donné lieu à une collaboration avec l'équipe Galaxy de l'Université de Freiburg (Bérénice Batut et Engy Nasr) pour l'améliorer et le rendre publiquement accessible, à terme, sur Galaxy Europe.

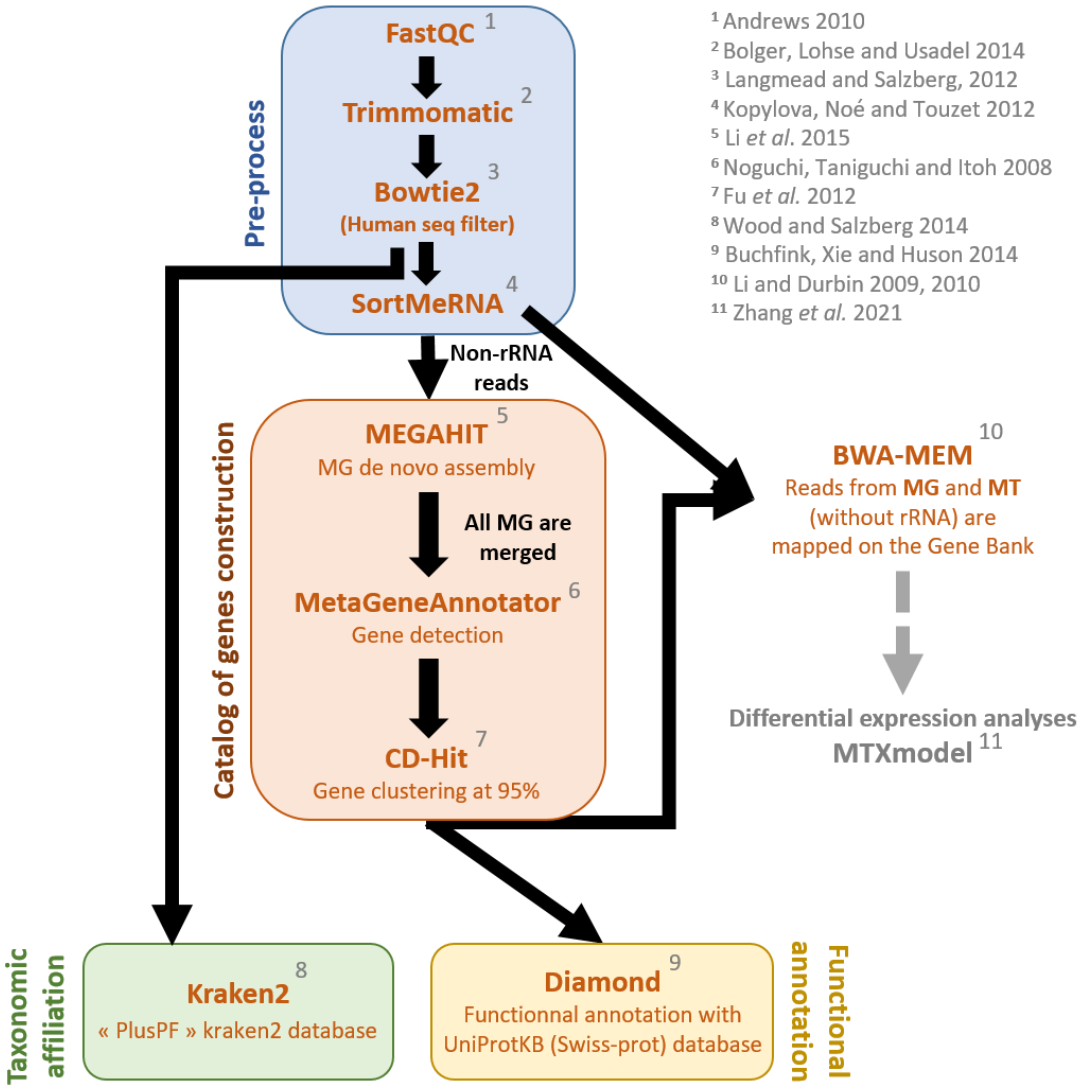


Figure 3 : Workflow bioinformatique pour le traitement de métagénomes et métatranscriptomes environnementaux.

3.7.Autres caractérisations biologiques et chimiques

En complément de l'analyse des acides nucléiques, le nombre total de cellules a été quantifié par cytométrie en flux et l'ATP a été quantifié par bioluminescence. Les principaux ions (Na^+, NH_4^+, K^+, Mg^{2+}, Ca^{2+}, Cl^-, NO_3^- et SO_4^{2-}) ont également été quantifiés par chromatographie ionique. Ces mesures ont été prises pour chaque échantillon à partir du quatrième HFRi rempli d'H_2O comme liquide de collecte (**Figure 2**) ou de l'eau de pluie avant filtration pour l'analyse des acides nucléiques.

3.8. Les données météorologiques et rétro-trajectoires des masses d'air.

Grâce aux plateformes météorologiques instrumentées de la CO-PDD, les données météorologiques sont publiquement accessibles sur le centre de données de l'OPGC : https://www.opgc.fr/data-center/public/data/copdd/pdd. La hauteur de la couche limite atmosphérique (ABL) et des rétro-trajectoires de 72 heures des masses d'air peuvent être obtenues à partir de la ré-analyse des données de l'ECMWF ERA5 (Jean-Luc Baray) (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5; Hoffmann et al. 2019). De plus, à partir de la composition chimique des nuages, leur source/origine majoritaire peut être déterminée et classée en catégories : pollué, marin, hautement marin (Deguillaume et al., 2014; Renard et al., 2020).

4. Résultats et discussion

4.1. Contrôles et validation des protocoles

De nombreuse difficultés sont liées à l'échantillonnage atmosphérique telles que : la faible biomasse (~10^4 cellules m^{-3} d'air), la temps d'échantillonnage (temporalité de la masse d'air) et la préservation de l'état *in situ* de l'échantillon pendant le temps de collecte (plusieurs heures) (Šantl-Temkiv et al., 2020). L'un de nos objectifs principaux étant de collecter assez de biomasse pour obtenir des quantités suffisantes d'ADN et d'ARN pour ne pas recourir à des étapes de pré-amplification avant le séquençage des MGs et MTs. De plus, avec l'arrivée des nouvelles technologies de séquençage NGS, d'autres difficultés s'ajoutent, comme le contrôle et la détection des contaminants et des biais et artefacts de séquençage (de Goffau et al., 2018).

Nous avons ainsi mis en place plusieurs contrôles négatifs au cours de l'échantillonnage (blanc d'échantillonnage) et du traitement des échantillons (blanc H_2O). Le protocole de décontamination des HFRi a également été testé en contaminant intentionnellement les cuves (avec de l'ATP ou des cellules bactériennes). Les résultats ont montré que les cuves étaient correctement décontaminées, avec des concentrations en ATP et en cellules après décontamination identiques aux blancs. De même, le séquençage des blancs d'échantillonnage et d'H_2O ont montré l'absence de contaminations communes, et donc une décontamination et des précautions lors du traitement des échantillons efficaces. La réalisation de contrôles négatifs tout au long de l'échantillonnage et du traitement ainsi qu'une décontamination efficace du matériel est essentielle à toute étude ; et l'est d'autant plus dans le cas d'un environnement pauvre en biomasse comme l'atmosphère, où les contaminations peuvent être plus facilement confondues avec la biodiversité réelle (Šantl-Temkiv et al., 2020).

Des contrôles positifs pour le séquençage d'amplicon (communautés mocks) ont également été mis en place. Nous avons ainsi pu détecter des soucis de regroupement avec l'approche par OTU et avons mis en place une collaboration avec l'équipe MIGALE (INRAE) pour analyser les données par regroupement en ASV. Cela a également permis de mettre en évidence la présence d'artefacts de séquençage ou de contaminants provenant certainement du kit d'extraction (8 ASVs attendues sur 22). Peu d'études utilisent des contrôles positifs lors du séquençage, or cela est recommandé pour améliorer la détection des contaminants et artefacts de séquençages (de Goffau et al., 2018) ainsi que des biais de quantifications ou d'affiliations. Les primers de séquençage utilisés n'ont en effet pas tous la même affinité avec toutes les séquences et vont amplifier préférentiellement certains groupes taxonomiques (Parada et al., 2016; Rajendhran and Gunasekaran, 2011; Reysenbach et al., 1992).

Finalement, les échantillons de 9 nuages et 11 aérosols collectés ont été séquencés en triplicata pour vérifier la répétabilité des échantillonneurs pour un même évènement. Trois échantillons d'aérosols collectés à des jours consécutifs en triplicata ont été présentés dans cette étude. Les résultats de clustérisassions hiérarchiques ont montré que chaque réplica groupait par évènement. Les HFRi donnent donc des résultats reproductibles, et les réplicas pourront être poolés pour un même échantillon dans le cas d'analyses demandant de plus grandes quantités d'acides nucléiques (MG et MT). Les quantités d'ADN récupérées avec la procédure expérimentale mise en place étaient suffisantes pour le séquençage Illumina (minimum 5 ng ; Dommergue *et al.*, 2019; Quick *et al.*, 2017) : 0.05 à 9.01 ng μL^{-1} pour les aérosols et 0.25 à 0.50 ng μL^{-1} pour les nuages. Les quantités d'ADN obtenues ainsi que l'ensemble des contrôles négatifs et positifs effectués valide donc notre protocole expérimental pour les analyses atmosphériques basées sur les acides nucléiques.

Les résultats résumés ci-dessus sont présentés dans l'**Article 1** du **Chapitre 2** de ce manuscrit.

5. Les nuages et les précipitations n'abritent pas les mêmes communautés bactériennes : les nuages sont des réservoirs de biodiversité tandis que la pluie lessive la colonne d'air et se charge en biomasse.

Les communautés microbiennes atteignant la surface terrestre avec les précipitations ne sont pas exactement les mêmes que celles présentes dans les nuages. En effet, les précipitations vont lessiver la colonne d'air lors de leur passage et se charger en composés chimiques et biologiques (Bourcier et al., 2012; Jaffrezo and Colin, 1988). Les communautés microbiennes atteignant le sol vont donc être le résultat de la diversité présente dans les nuages mais aussi dans la colonne d'air, et donc influencer par les sources d'émission plus locales. Néanmoins, encore peu de choses sont connues sur la proportion de microorganismes lessivés et sur les potentielles différences de lessivage entre groupes taxonomiques. Les communautés microbiennes dans les nuages et les précipitations ont déjà été

étudiées (Aho et al., 2019; Amato et al., 2017), mais jamais de manière coordonnée. L'objectif de cette étude était ainsi d'étudier la diversité bactérienne dans les nuages et les aérosols ainsi que leur composition chimique dans le but d'établir des liens directs entre ces deux compartiments atmosphériques et de mieux comprendre l'impact du lessivage (**Figure 4**).

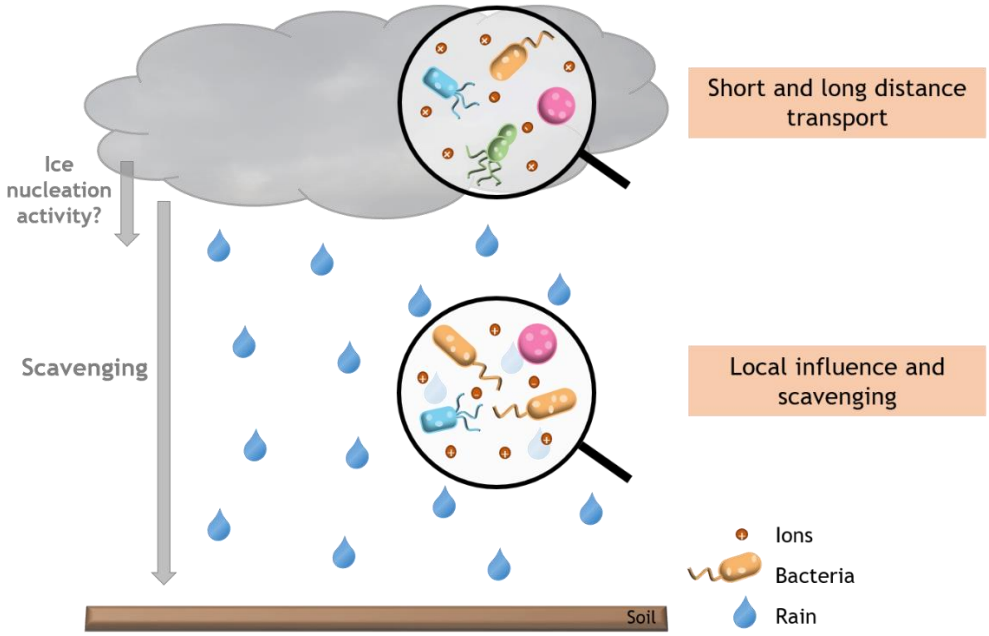


Figure 4 : Objectif d'étude de l'analyse des communautés microbiennes dans les nuages et la pluie (Péguilhan et al., 2021).

Un total de 4 nuages et 11 précipitations (10 pluies et 1 neige) ont été collectés aux stations PUY et OPM. À trois reprises, des échantillons de pluie et de nuages ont pu être collectés le même jour, ou à un jour d'intervalle, et sont donc directement associés à la même masse d'air (désignés comme les évènements *a*, *b* et *c*).

Premièrement, des concentrations en ions étaient corrélées indiquant une origine commune, tel que les ions Na^+, Cl^-, Mg^{2+}, SO_4^{2-} et K^+ (corrélation de Spearman, $P > 0.05$) témoignant d'une source océanique, et d'une source continentale pour les ions N_4^+ et NO_3^- (Deguillaume et al., 2014), en accord avec les rétro-trajectoires des masses d'air. Les ions d'origines océaniques étaient plus concentrés dans les nuages (**Figure 5**), mettant en évidence la dilution de ces ions dans la pluie. En effet, la colonne d'air, sous le nuage au niveau de la région puy de Dôme, ne contient aucune source d'origine océanique. La pluie va donc se charger en ions d'origines plus continentales et diluer les concentrations en ions d'origines marines. À partir de ce facteur de dilution, nous pouvons estimer les concentrations attendues en cellules dans la pluie à partir des concentrations présentes dans les nuages, et calculer d'après la différence entre les deux concentrations estimée et réelle, un taux de lessivage de la colonne d'air par la pluie. Les concentrations en cellules dans la pluie ont donc été normalisées par la

concentration en ion Na^+ pour les trois évènements liés (*a*, *b* et *c*). Ainsi, il a été estimé que 95% des cellules présentes dans la pluie proviennent du lessivage de la colonne d'air. La pluie se chargerait donc massivement en biomasse lors du lessivage de la colonne d'air, contrairement à ce qui était estimé dans Moore *et al.* (2020). Cependant, la localisation (puy de Dôme, France ; ou la Louisiane, USA), le nombre et la taille des aérosols et gouttelettes d'eau, ainsi que l'intensité et la durée des précipitations peuvent contribuer à l'obtention de grandes différences en termes d'estimation.

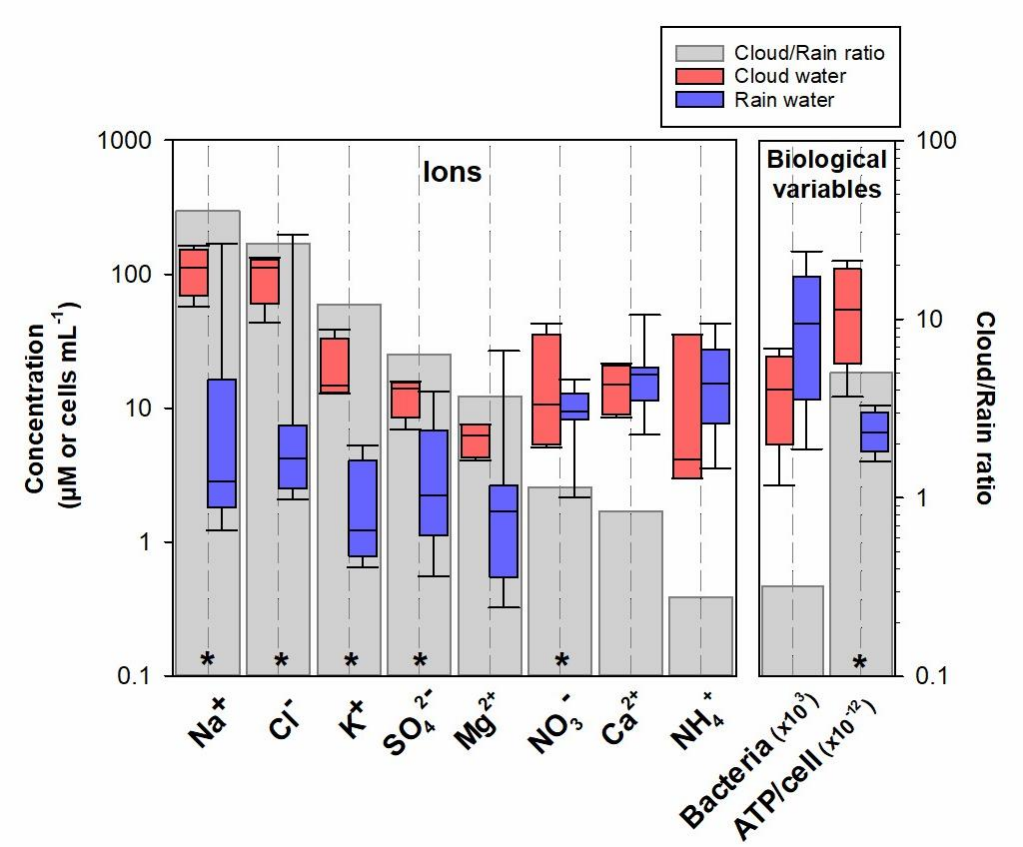


Figure 5 : Concentrations absolues (axe de gauche) et médiane des ratios (axe de droite) des principaux ions et des variables biologiques dans les nuages et les précipitations. Les axes y sont à échelle logarithmique. Les astérisques indiquent une différence significative entre les concentrations dans les nuages ou la pluie (test de Mann-Whitney, $P < 0,05$). Figure provenant de Péguilhan *et al.* (2021).

Deuxièmement, concernant la diversité bactérienne, les phylums les plus abondants étaient les Protéobactéries (en particulier les Betaproteobactériales, Pseudomonadales, Sphingomonadales et Rhizobiales), Actinobatéries (Micrococcales, Corynebacteriales et Frankiales), Firmicutes (Bacillales et Lactobacillales), Bactéroidetes et Deinococcus-Thermus. Environs 75 % des genres bactériens détectés étaient partagés par les nuages et les pluies ce qui montrent le lien fort entre ces deux étapes consécutives du cycle de l'eau atmosphérique. Certains genres bactériens étaient significativement plus abondants dans les nuages, comme *Undibacterium*, *Staphylococcus*, *Bacillus*, *Streptococcus* and *Corynebacterium*. D'autres étaient plus présents dans la pluie tels que *Massilia*, *Rhodococcus*, *Curtobacterium*, *Frondihabitans*, *Sphingomonas* et *Deinococcus* (**Figure 6**). Les genres bactériens retrouvés préférentiellement dans la pluie étaient principalement des taxa associés à la phyllosphère, ce qui corrèle avec de précédentes observations concernant leur capacité limitée en terme de dispersion atmosphérique verticale (Els et al., 2019). Concernant les taxa majoritairement présents en condition nuageuse, le fait qu'ils aient pu éviter la précipitation semble irréaliste, cependant, à l'inverse, certains genres ont été reportés comme pouvant induire des précipitations et donc être préférentiellement précipités, telle que la bactérie *Pseudomonas syringae* (Morris et al., 2004). De plus, les nuages contenaient une richesse en genre bactérien plus importante que les précipitations. Cela peut être le résultat des multiples sources d'émission de microorganismes mixées sur de longues distances (Després et al., 2007; Smith et al., 2013). Les nuages seraient donc un réservoir de biodiversité, tandis que les pluies sont plutôt des sous échantillons de ceux-ci, récupérant de la biomasse lors de leur passage dans la colonne d'air jusqu'à la surface.

Enfin, le contenu en ATP par cellule était plus important dans les nuages (**Figure 5**), ce qui peut mettre en évidence une plus grande proportion de cellules actives dans les nuages, ou une activité métabolique plus importante, soutenant les nuages comme habitats microbiens (Ervens and Amato, 2020; Sattler et al., 2001).

L'étude résumée ci-dessus a été publiée (Péguilhan et al., 2021) et est présentée dans le **Chapitre 3** de ce manuscrit.

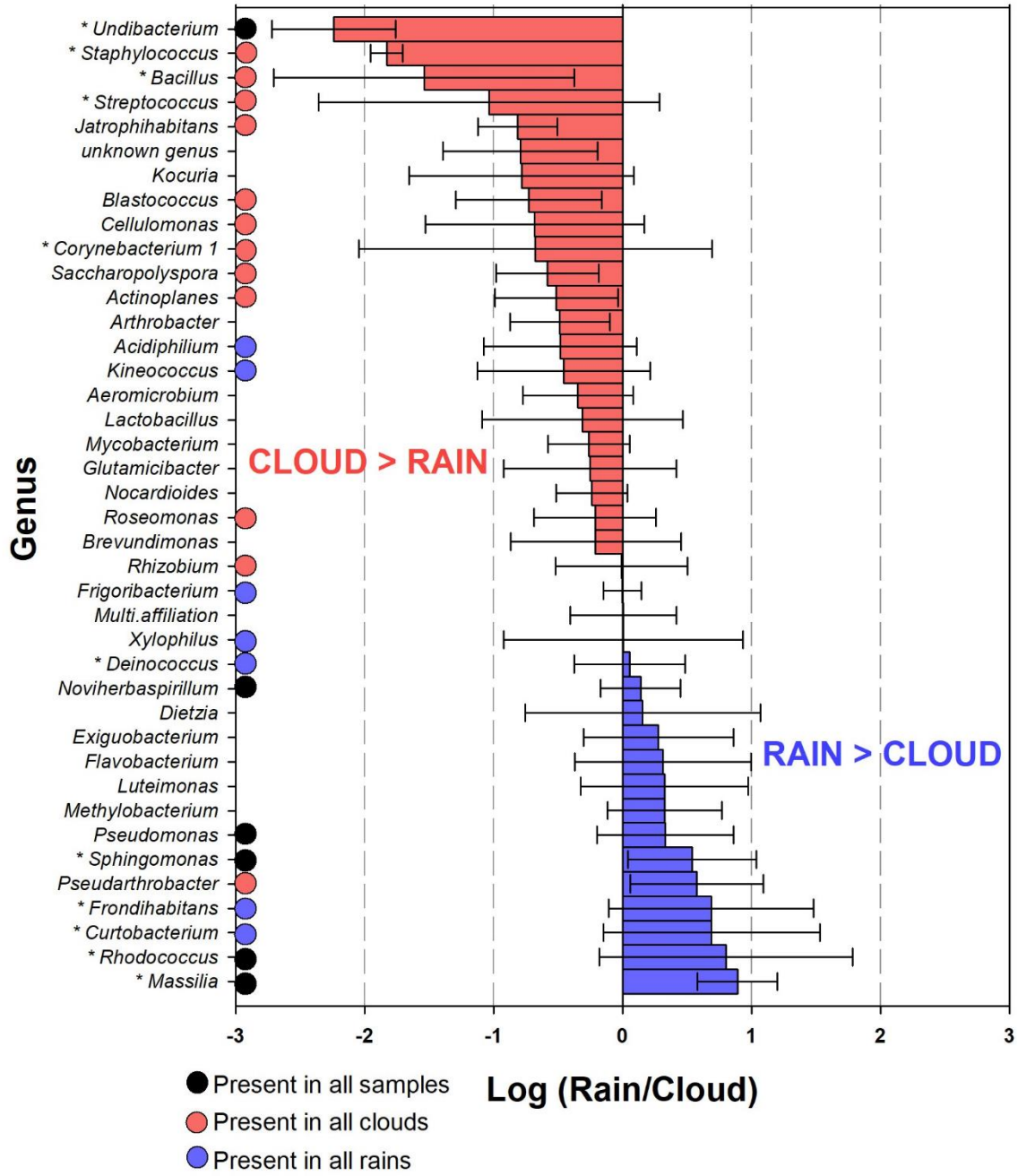


Figure 6 : Moyenne des ratios pluies sur nuages (log) dans les trois évènements *a*, *b* et *c* pour les 40 genres bactériens représentant plus de 100 séquences (parmi 135 genres distinctes). Les genres plus présents dans les nuages par rapport aux pluies sont représentés en rouge, et les genres plus présents dans les pluies sont représentés en bleu. Les barres d'erreur indiquent la déviation standard par rapport à la moyenne des ratios pluies sur nuages. Les genres présents dans tous les échantillons, dans tous les nuages ou toutes les pluies sont indiqués par un rond noir, rouge ou bleu respectivement. Les astérisques sur les noms de genre indiquent une différence significative de l'abondance entre les nuages et les pluies, considérant tous les échantillons de l'étude (test de Kruskal-Wallis, P-value < 0,05). Figure provenant de Péguilhan *et al.* (2021).

6. Comparaison de la diversité bactérienne dans les nuages et les aérosols : un effet saison important.

Dans le but d'étudier les nuages en tant qu'entités biologiques spécifiques dans l'atmosphère, nous nous sommes tout d'abord intéressé à la diversité bactérienne présente dans la phase nuageuse (atmosphère humide) et les aérosols (atmosphère sèche). L'objectif de cette étude était également de vérifier la répétabilité des HFRi pour un même évènement en analysant chaque réplica d'échantillon. Un nuage est un ensemble de gouttelettes d'eau en suspension dans l'atmosphère qui vont se former autour de noyau de condensation (CCN), c'est-à-dire des aérosols, lorsque l'eau sera en sursaturation dans l'air. La capacité à devenir CCN d'un aérosol est déterminée par sa taille et son rayon de courbure (équation de Köhler) (Wex et al., 2008), tous les aérosols ne sont donc pas forcement des CCNs, mais une particule d'environ 1 µm ne peut y échapper, d'après les connaissances théoriques. Les microorganismes peuvent donc être des CCNs et la diversité présente dans les aérosols devrait être retrouvée dans les nuages (Bauer et al., 2003; Lazaridis, 2019; Möhler et al., 2007). L'hypothèse ici est donc qu'il n'y a pas de différence en termes de diversité bactérienne entre les nuages et les aérosols.

Pour cette étude l'ensemble des 8 et 11 échantillons de nuages et d'aérosols ont été séquencés avec leur réplicas. La diversité bactérienne détectée était en accord avec les groupes taxonomiques communéments retrouvés dans l'atmosphère (même phylums dominants que pour l'analyse des pluies et des nuages) (Amato et al., 2017; Bowers et al., 2013; Tignat-Perrier et al., 2020). Parmi cette biodiversité, les genres bactériens les plus abondants étaient : *Sphingomonas*, *Pseudomonas*, *Methylobacterium*, *Hymenobacter*, *Acidiphilium*, *Massilia*, *Bacillus* et *Staphylococcus* (**Figure 7**). Des genres bactériens étaient également significativement plus présents en condition nuageuse ou dans les aérosols secs, tel que *Staphylococcus*, *Acinetobacter*, *Paracoccus*, *Kocuria*, *Corynebacterium*, *Enhydrobacter*, *Streptococcus* et *Aerococcus*, ou les genres *Bacillus*, *Rubellimicrobium* et *Blastococcus* respectivement (**Figure 7**). Le fait que des bactéries aient pu éviter d'intégrer les gouttelettes de nuages parait peu probable compte tenu de leur taille et de l'équation théorique de Köhler. Cependant, plusieurs processus peuvent expliquer une différence de biodiversité comme : (i) la forme (indice de courbure) de l'aérosol ; (ii) l'hydrophobicité cellulaire ou de la particule organique ou inorganique à laquelle elle est attachée (bien qu'ici nous échantillonnons à la fois l'eau de nuage et l'air) ; (iii) la croissance cellulaire dans les gouttelettes de nuages (bien qu'estimée négligeable ; Ervens and Amato, 2020) ; (iv) ou encore le choc osmotique lié au passage en phase aqueuse qui aurait éliminé une partie des communautés bactériennes présentes dans les aérosols. Néanmoins, la présence d'eau condensé ou non n'explique pas seule les différences de biodiversité observées, l'effet de la saison est également important (**Figure 8**). En effet, les aérosols étaient en moyenne collectés en été et automne, et les nuages en conditions plus hivernales (automne et hiver). La richesse spécifique était d'ailleurs

plus importante dans les aérosols et en été. Ces observations sont appuyées par Tignat-Perrier *et al.* (2020) avec une concentration cellulaire supérieure en été par rapport aux autres saisons.

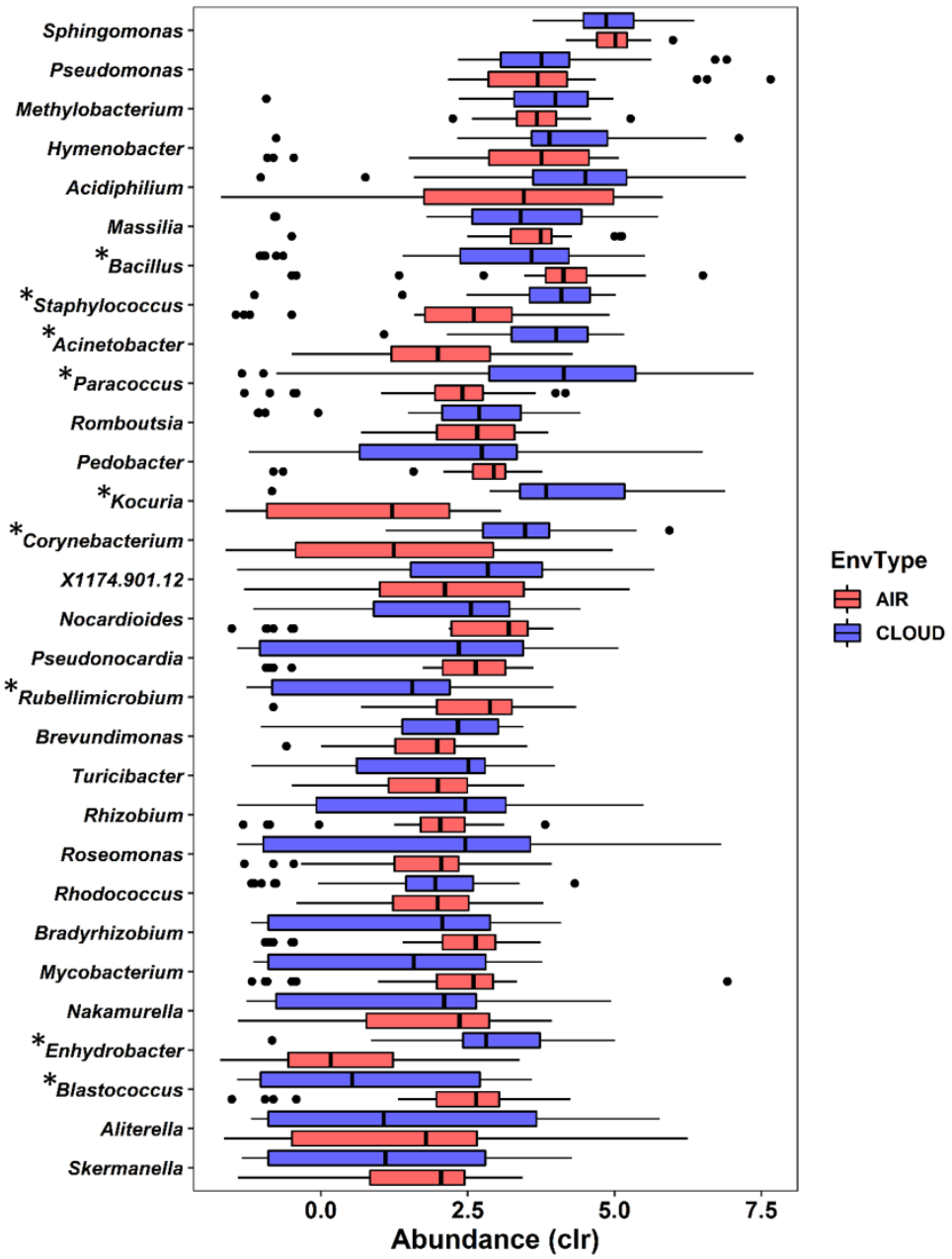


Figure 7 : Les 30 genres bactériens les plus abondants dans les nuages et les aérosols. L'axe x représente le nombre de séquence par genre transformé en centered-log ratio (clr). EnvType : type environnemental ; * : genre ayant une abondance significativement différente entre les nuages et les aérosols, test de Kruskal-Wallis avec une p-value < 0,05.

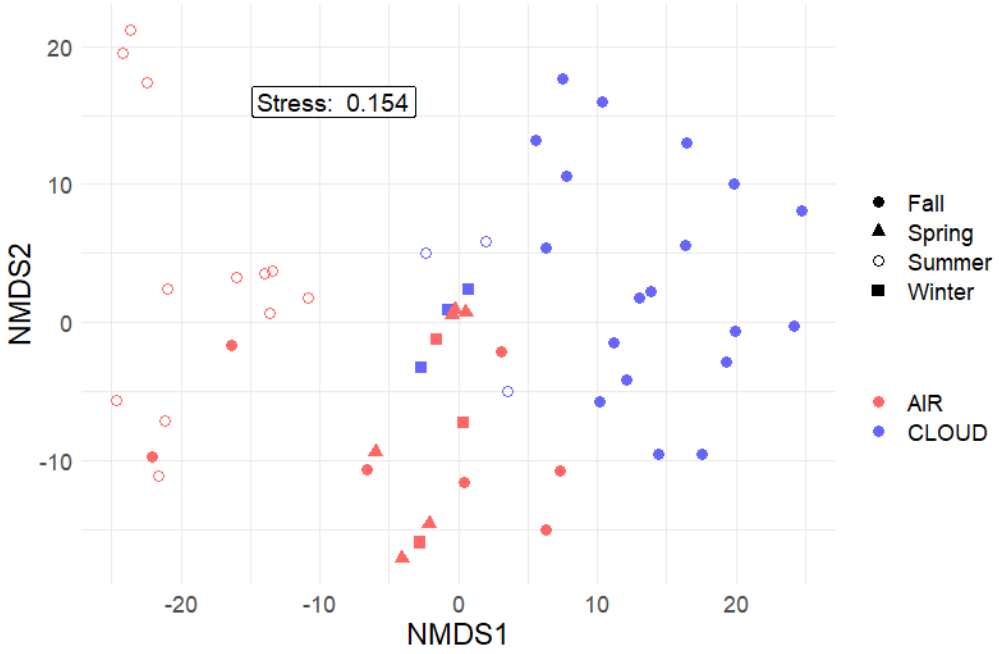


Figure 8 : Etalonnage multidimensionnel non-métrique (NMDS) basé sur l'abondance des 862 ASVs présentes dans les nuages et les aérosols (distance Euclidienne). Chaque réplica est représenté.

Pour finir, la majorité des réplicas se regroupait par échantillon lors de l'analyse par regroupement hiérarchique. Cela met en avant l'homogénéité de collecte des échantillonneurs HFRi pour un même évènement. La présence de réplicas d'échantillonnage est importante dans toute étude, bien que peu appliqué dans le contexte environnemental, pour l'aspect statistique mais aussi être informé de la répétabilité de l'échantillonneur utilisé.

7. Les microorganismes aéroportés sont actifs dans les aérosols et les nuages, mais leur potentiel fonctionnel est supérieur en condition nuageuse.

Toujours dans l'objectif d'étudier la spécificité des nuages dans l'atmosphère, nous nous sommes intéressés à l'aspect fonctionnel des communautés microbiennes en condition nuageuse ou non. Des approches de métagénomique et métatranscriptomique ont été utilisés pour explorer de manière non ciblée ces communautés complexes. Ceux sont les premiers MG et MT non-amplifiés de nuages et d'aérosols obtenus à notre connaissance.

Premièrement, concernant les taxa actifs, seulement une faible proportion de la diversité totale présente dans les nuages et les aérosols était active (~20%), témoignant d'un environnement exigeant. Une majorité d'eucaryote était détectée dans les nuages (85% des séquences affiliées dans MT, principalement des champignons), et une majorité de bactérie dans les aérosols (70% des séquences affiliées dans MT). Le fait d'être en condition nuageuse (donc humidité plus importante) a peut-être favorisé le développement ou la survie des champignons. Cependant, la richesse en groupe actif était

grandement supérieure chez les bactéries avec un total de 84 familles bactériennes actives contre seulement 18 familles d'eucaryote, et un niveau d'activité pour les groupes les plus actifs clairement supérieur chez les bactéries. Cela met en valeur les bactéries comme le domaine du vivant majoritairement actif dans ces communautés aéroportées. La diversité microbienne et les taxons actifs de ces communautés n'étaient pas significativement différents entre les nuages et les aérosols, et assez variable entre échantillon. La composition des communautés microbiennes et leur partie active s'expliquerait donc plutôt par les apports spécifiques des différentes sources d'émission qui les constituent, que par les spécificités seules de l'environnement (nuageux ou sec). Parmi les groupes les plus actifs, nous pouvons citer : les Halomonadaceae (Gamma-proteobactéries), Rickettsiaceae (Alpha-proteobactéries), Mycoplasmataceae, Clostridiaceae, Peptoniphilaceae, Peptostreptococcaceae (Firmicutes), Amoebophilaceae (Bactéroidetes), Chroococcaceae (Cyanobactérie) et Treponemataceae (Spirochètes) chez les bactéries ; et les Hemiselmidaceae, Geminigeraceae, Cryptomonadaceae (Cryptophytes), Phaeodactylaceae (Bacillariophytes, diatomées) et Plasmodiidae (Apicomplexes) pour les eucaryotes. De nombreux taxa photosynthétiques étaient donc actifs en conditions nuageuse ou sèche (cyanobactéries et micro-algues). Néanmoins, certains taxa étaient significativement plus actifs dans un compartiment atmosphérique ou dans l'autre, tels que les Parachlamydiaceae, Halomonadaceae, Legionellaceae (familles bactériennes), Glomerellaceae, Cryptomonadaceae et Trichomonadaceae (familles d'eucaryotes) dans les nuages ; et les Bryobacteraceae, Pirellulaceae, Ornithinimicrobiaceae (familles bactériennes), Dipodascaceae et Theileriidae (familles d'eucaryotes) dans les aérosols. De plus, le genre *Pseudomonas* semblait actif seulement dans les nuages, ce qui corrèle avec les observations faites dans (Amato et al., 2017) et (Šantl-Temkiv et al., 2018) montrant cette bactérie comme potentiellement active dans les nuages et la pluie mais pas dans les aérosols. Cela soutient le lien supposé entre cette bactérie et le cycle de l'eau ainsi que la théorie des bioprécipitations (Morris et al., 2004, 2008).

Deuxièmement, concernant le profil fonctionnel des communautés microbiennes, les gènes ont été regroupés en ontologie de gène (GO ; c'est-à-dire en famille de gène par fonction) qui sont elles-mêmes subdivisées en trois catégories : les composants cellulaires (CC), les processus biologiques (PB) et les fonctions moléculaires (FM). Parmi les CC les plus surreprésentés en conditions nuageuses et sèches, nous avions principalement des composants de membrane cellulaire et du photosystème. Pour les PB, c'étaient également des processus liés à la photosynthèse qui étaient surexprimés, avec de nombreux processus liés au métabolisme central (métabolisme des carbohydrate, glycolyse, cycle des acides tricarboxyliques, …), à la synthèse de protéine (translation), au métabolisme énergétique (synthèse d'ATP, chaîne respiratoire de transport d'électron, …) (**Figure 9A**) et au cycle cellulaire (division cellulaire, réplication de l'ADN, division du noyau). Cela met en évidence la présence d'un métabolisme minimum pour la survie des cellules dans les nuages mais aussi les aérosols. Différents

processus liés à une réponse à un environnement extrême étaient aussi surreprésentés comme : réponse aux stresses osmotiques et oxydatifs, catabolisme du peroxyde d'hydrogène, régulation du pH, réparation de l'ADN, autophagie et réponses à un manque de nutriments. La dégradation du peroxyde d'hydrogène soutient les observations de Vaïtilingom *et al.* (2013) et Wirgot *et al.* (2017), et supporte l'impact potentiel des communautés microbiennes sur la chimie atmosphérique. Enfin, les FM surreprésentées témoignaient, de la même manière, d'une activité photosynthétique (activité productrice d'oxygène et voie de transport d'électrons liée à l'activité photosynthétique) et d'un milieu où les nutriments se font rare avec la surexpression de fonctions liées à l'isocitrate lyase qui est la clé d'entrée dans le cycle du glyoxylate, permettant de métaboliser des sucres plus simples comme l'acétate. Le cycle du glyoxylate peut également être lié au stress oxydatif (Ensign, 2006; Park et al., 2019).

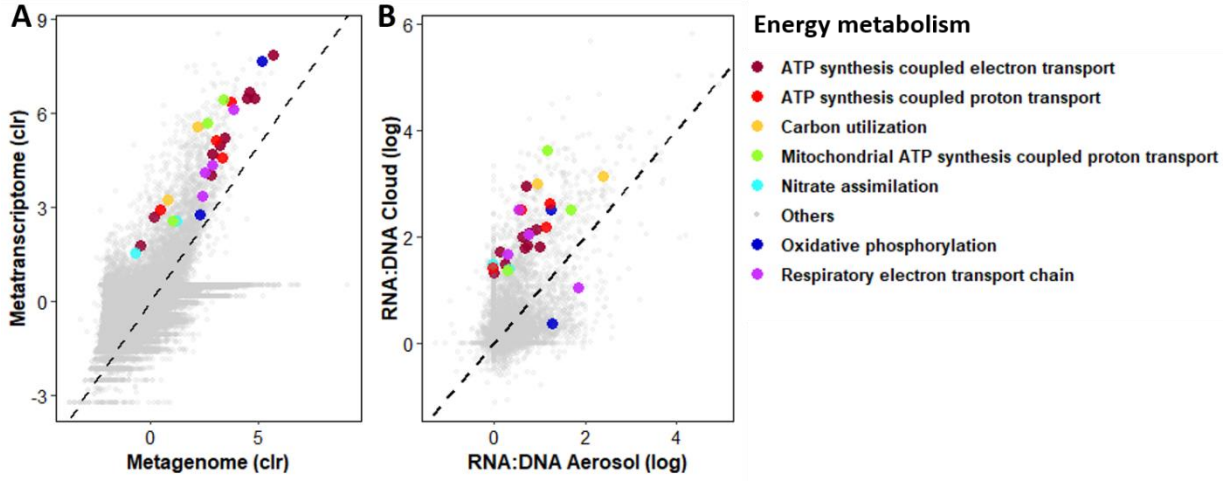


Figure 9 : Niveau d'expression global (A) ou dans les nuages et les aérosols (B) de gènes liés au métabolisme énergétique. A : l'abondance dans les métagénomes et métatranscriptomes a été transformée en centered-log ratio (clr) pour la normalisation ; **B** : l'échelle des ratios ARN:ADN est représentée en log.

Au milieu de cette activité globale des communautés dans les aérosols et les nuages, plusieurs gènes étaient significativement surexprimés dans l'un ou l'autre de ces situations atmosphériques. Les résultats de regroupement hiérarchique et de NMDS montraient également une distinction entre le fonctionnement des communautés actives en conditions nuageuse et sèche. Dans les aérosols, la phosphorylation oxydative et le catabolisme du glucose étaient significativement surreprésentés. Le catabolisme de l'hydrogène peroxyde, les réponses liées à un stimulus de température, à des dommages de l'ADN et au stress oxydatif ainsi que la réponse SOS étaient également surreprésentés dans les aérosols. Cela indique que les microorganismes présents dans l'atmosphère sèche seraient plus exposés aux radicaux libres que dans les nuages. En effet, la concentration atmosphérique en H_2O_2

est liée à la température (aérosols collectés en été principalement) et aux radiations solaires (Lee et al., 2000). En condition nuageuse les processus liés au métabolisme énergétique étaient fortement surexprimés (**Figure 9B**) ainsi que la translation cytoplasmique, les métabolismes du glucose et des carbohydrates et le catabolisme des polysaccharides. De même, différentes réponses à l'environnement étaient surreprésentées telle que l'autophagie des peroxysomes et autres processus d'autophagie, les réponses au manque de nutriment, au stress osmotique, au UV, au stress nitrosatif, la régulation intracellulaire du pH et la détoxification des composés nitrogénés. Les réponses liées au stress osmotique et à la régulation du pH semblent logiques compte tenu de l'environnement aqueux. L'autophagie des peroxysomes est un des principaux moyens de régulation de ces organites intracellulaires, ayant pour fonction, entre autre, la détoxification de l'H_2O_2 (Till et al., 2012). La régulation négative des peroxysomes indiquerait donc un besoin moins important de détoxification en raison de concentrations moins élevées en radicaux libres, ce qui corrèle avec les observations faites pour les aérosols.

Par ailleurs, les ratios ARN:ADN étaient en moyenne significativement plus importants dans les nuages que dans les aérosols, indiquant un potentiel fonctionnel plus élevé en condition nuageuse. Le nombre de GO significativement surexprimés était également plus important dans les nuages pour chaque catégorie (**Figure 10**) mettant en avant un plus grand nombre de fonctions exprimées, ou des fonctions plus fortement exprimées, dans les nuages par rapport aux aérosols.

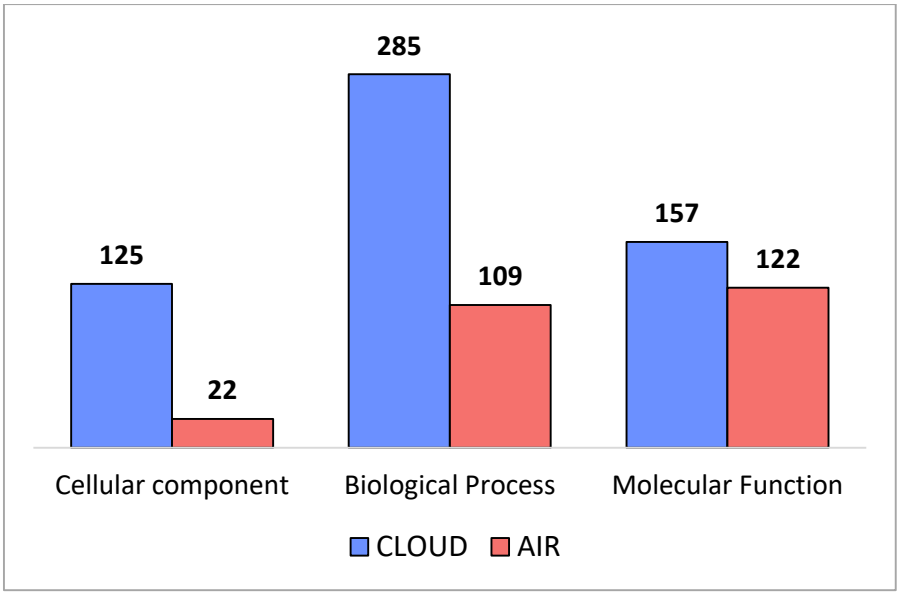


Figure 10 : Nombre d'ontologie de gène (GO) significativement surexprimée dans les nuages ou les aérosols par catégories de GO.

Les communautés microbiennes aéroportées sont donc potentiellement actives à la fois dans les aérosols et les nuages. Elles expriment de nombreuses fonctions en lien avec le métabolisme central et en réponse à leur environnement témoignant de leur viabilité et tentative d'acclimatation aux conditions atmosphériques. Cependant, les communautés microbiennes semblent avoir un potentiel fonctionnel plus important en condition nuageuse, surement en raison du milieu aqueux fournis (**Figure 11**). Les nombreuses réponses liées au manque de nutriment peuvent être interprétées comme un regain d'activité des cellules et donc la nécessiter d'une plus grande quantité de nutriment.

Cette étude fonctionnelle est développée dans le **Chapitre 4** de ce manuscrit.

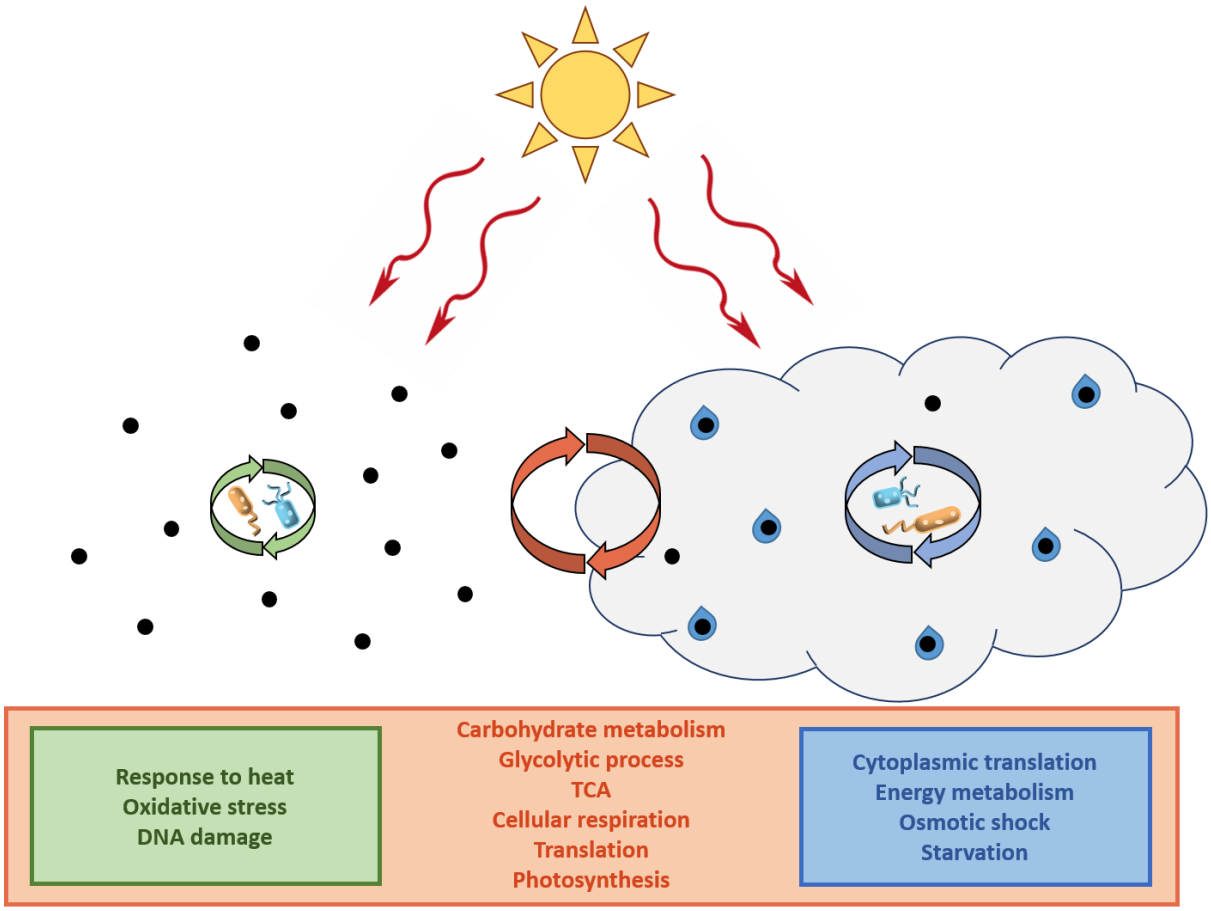


Figure 11 : Synthèse des résultats d'analyse fonctionnelle des nuages et des aérosols. Encadré orange : ontologie de gène (GO) surexprimée dans les deux conditions ; encadré vert : GO surexprimée dans les aérosols ; encadré bleu : GO surexprimée en condition nuageuse.

8. Conclusions et perspectives

Pour conclure, les procédures expérimentales misent en place au cours de ce travail de thèse ont permis non seulement de pouvoir échantillonner en simultané des nuages et les précipitations associées, mais aussi d'obtenir des quantités d'acide nucléique suffisantes à partir d'échantillons de nuages, aérosols et précipitations pour pouvoir recourir à du séquençage Illumina HiSeq sans étape de pré-amplification. Les protocoles d'échantillonnage et de traitement des échantillons ont été optimisés et ont pu être validés grâce à l'utilisation de contrôles négatifs, de contrôles qualité du séquençage et du traitement de réplicas d'échantillonnage.

L'étude de la diversité bactérienne des nuages et précipitations en simultanée a montré que les nuages sont des mosaïques de microorganismes provenant de sources d'émissions variées, mixées sur de longues distances, et peuvent être ainsi considérés comme des réservoirs atmosphériques de souches qui seront disséminées sur les écosystèmes de surface avec les précipitations. La pluie, quant à elle, constitue un sous-échantillon des nuages et se charge en biomasse et composés chimiques lors du lessivage de la colonne d'air, limitant le transport vertical de certains groupes taxonomiques. En complément, l'étude de la diversité bactérienne dans les nuages et les aérosols a mis en valeur des différences en termes de communauté. La diversité présente dans les nuages devrait provenir de celle contenue dans les aérosols, et il parait peu probable que des taxa bactériens puissent éviter d'intégrer les gouttelettes de nuage. Un effet saison a été démontré et explique au moins en partie les différences observées. Cependant, ces distinctions peuvent aussi être le résultat de phénomènes physiques comme les chocs osmotiques et cycles de gel-dégel qui élimineraient une partie des cellules ayant survécu jusqu'ici.

Enfin, les approches NGS non ciblées de métagénomiques et métatranscriptomiques ont permis d'explorer et comparer pour la première fois, à notre connaissance, le profil fonctionnel des nuages et des aérosols. Il en ressort que ces deux situations atmosphériques abritent des communautés microbiennes potentiellement actives, exprimant de nombreuses fonctions liées au métabolisme central et à diverses réponses et acclimatations à leur environnement. Cependant, les nuages se distinguent par un potentiel fonctionnel plus élevé que dans les aérosols, mettant en avant les nuages comme un habitat potentiel et spécifique dans l'atmosphère.

Pour résumer l'ensemble de résultats présentés, les microorganismes sont aérosolisés et mixés sur de longues distances. Une faible proportion seulement restera viable et active dans l'atmosphère sec (aérosols) et sera exposée à des conditions de vie extrêmes avec entre autres de fortes expositions aux UVs et radicaux libres. Les microorganismes aéroportés qui atteindront de hautes altitudes et seront intégrés aux nuages devront faire face aux chocs osmotiques et cycles de gel-dégel, mais auront en échange un microenvironnement aqueux les « revivifiant ». Des cycles de déshydratation-

réhydratation successifs des sols ont même été montrés comme favorisant une réactivation plus rapide du métabolisme des cellules en présence d'eau (Leizeaga et al., 2022). Si on applique cela à l'atmosphère, les microorganismes ayant survécu jusqu'ici seraient donc revivifiés et préparés à être plus réactif avant d'être redéposés sur les écosystèmes de surface avec les précipitations. Les précipitations dissémineraient ainsi des sous-échantillons de souches microbiennes provenant des nuages et prêtes à coloniser de nouveaux écosystèmes, jouant un rôle potentiellement important sur l'écologie des communautés locales.

En perspectives, il serait des plus intéressant de pouvoir lier directement des échantillons de nuages, de pluie mais aussi d'aérosols, avec également un plus grand nombre d'échantillon. Il serait alors peut-être possible de suivre des souches microbiennes entre ces compartiments atmosphériques. Il faudrait également échantillonner en continue le passage d'une atmosphère sèche à un nuage pour mieux comprendre l'impact de la présence d'eau condensée sur les communautés aéroportées. Pour aller encore plus loin, l'idéal serait de pouvoir collecter des aérosols et nuages dans différentes zones géographiques pour observer l'activité microbienne atmosphérique converge à l'échelle globale.

Finalement, la métatranscriptomique seule ne permet pas d'avoir une preuve d'activité ou de pouvoir quantifier de manière absolue une activité spécifique. Le passage par la PCR quantitative permettrait de confirmer et quantifier certaines activités d'intérêts. Des approches de métabolomique et de métaprotéomique pourraient également être envisagées en complément des analyses réalisées pour avoir l'ensemble des étapes menant à la présence d'une protéine et de son activité.

9. References

Aho, K., Weber, C.F., Christner, B.C., Vinatzer, B.A., Morris, C.E., Joyce, R., Failor, K., Werth, J.T., Bayless-Edwards, A.L.H., and Schmale III, D.G. (2019). Spatiotemporal patterns of microbial composition and diversity in precipitation. Ecol. Monogr. *0*, 1–26.

Amato, P., Ménager, M., Sancelme, M., Laj, P., Mailhot, G., and Delort, A.M. (2005). Microbial population in cloud water at the Puy de Dôme: Implications for the chemistry of clouds. Atmos. Environ. *39*, 4143–4153.

Amato, P., Parazols, M., Sancelme, M., Laj, P., Mailhot, G., and Delort, A.M. (2007). Microorganisms isolated from the water phase of tropospheric clouds at the Puy de Dôme: Major groups and growth abilities at low temperatures. In FEMS Microbiology Ecology, pp. 242–254.

Amato, P., Joly, M., Besaury, L., Oudart, A., Taib, N., Moné, A.I., Deguillaume, L., Delort, A.M., and Debroas, D. (2017). Active microorganisms thrive among extremely diverse communities in cloud water. PLoS One *12*, 1–22.

Amato, P., Besaury, L., Joly, M., Penaud, B., Deguillaume, L., and Delort, A.M. (2019). Metatranscriptomic exploration of microbial functioning in clouds. Sci. Rep. *9*.

Baray, J.L., Deguillaume, L., Colomb, A., Sellegri, K., Freney, E., Rose, C., Baelen, J. Van, Pichon, J.M., Picard, D., Fréville, P., et al. (2020). Cézeaux-Aulnat-Opme-Puy de Dôme: A multi-site for the long-term survey of the tropospheric composition and climate change. Atmos. Meas. Tech. *13*, 3413–3445.

Bauer, H., Kasper-Giebl, A., Löflund, M., Giebl, H., Hitzenberger, R., Zibuschka, F., and Puxbaum, H. (2002). The contribution of bacteria and fungal spores to the organic carbon content of cloud water, precipitation and aerosols. Atmos. Res. *64*, 109–119.

Bauer, H., Giebl, H., Hitzenberger, R., Kasper-Giebl, A., Reischl, G., Zibuschka, F., and Puxbaum, H. (2003). Airborne bacteria as cloud condensation nuclei. J. Geophys. Res. Atmos. *108*, 4658.

Bernard, M., Rué, O., Mariadassou, M., and Pascal, G. (2021). FROGS: a powerful tool to analyse the diversity of fungi with special management of internal transcribed spacers. Brief. Bioinform. *22*.

Bourcier, L., Masson, O., Laj, P., Chausse, P., Pichon, J.M., Paulat, P., Bertrand, G., and Sellegri, K. (2012). A new method for assessing the aerosol to rain chemical composition relationships. Atmos. Res. *118*, 295–303.

Bowers, R.M., Lauber, C.L., Wiedinmyer, C., Hamady, M., Hallar, A.G., Fall, R., Knight, R., and Fierer, N. (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, R.M., McLetchie, S., Knight, R., and 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.

Bowers, R.M., Clements, N., Emerson, J.B., Wiedinmyer, C., Hannigan, M.P., and Fierer, N. (2013). Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. Environ. Sci. Technol. *47*, 12097–12106.

Burrows, S.M., Elbert, W., Lawrence, M.G., and 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.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 2016 137 *13*, 581–583.

Deguillaume, L., Charbouillot, T., Joly, M., Vaïtilingom, M., Parazols, M., Marinoni, A., Amato, P., Delort, A.M., Vinatier, V., Flossmann, A., et al. (2014). Classification of clouds sampled at the puy de Dôme (France) based on 10 yr of monitoring of their physicochemical properties. Atmos. Chem. Phys. *14*, 1485–1506.

Després, V.R., Nowoisky, J.F., Klose, M., Conrad, R., Andreae, M.O., and Pöschl, U. (2007). Characterization of primary biogenic aerosol particles in urban, rural, and high-alpine air by DNA sequence and restriction fragment analysis of ribosomal RNA genes. Biogeosciences *4*, 1127–1141.

Després, V.R., Alex Huffman, J., Burrows, S.M., Hoose, C., Safatov, A.S., Buryak, G., Fröhlich-Nowoisky, J., Elbert, W., Andreae, M.O., Pöschl, U., et al. (2012). Primary biological aerosol particles in the atmosphere: A review. Tellus, Ser. B Chem. Phys. Meteorol. *64*.

Dommergue, A., Amato, P., Tignat-Perrier, R., Magand, O., Thollot, A., Joly, M., Bouvier, L., Sellegri, K., Vogel, T., Sonke, J.E., et al. (2019). Methods to investigate the global atmospheric microbiome. Front. Microbiol. *10*.

Els, N., Baumann-Stanzer, K., Larose, C., Vogel, T.M., and Sattler, B. (2019). Beyond the planetary boundary layer: Bacterial and fungal vertical biogeography at Mount Sonnblick, Austria. Geo Geogr. Environ. *6*.

Ensign, S.A. (2006). Revisiting the glyoxylate cycle: Alternate pathways for microbial acetate assimilation. Mol. Microbiol. *61*, 274–276.

Ervens, B., and Amato, P. (2020). The global impact of bacterial processes on carbon mass. Atmos. Chem. Phys. Discuss. *20*, 1–25.

Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., Maman, S., Hernandez-Raquet, G., Combes, S., and Pascal, G. (2018). FROGS: Find, Rapidly, OTUs with Galaxy Solution. Bioinformatics *34*, 1287–1294.

Fröhlich-Nowoisky, J., Pickersgill, D.A., Després, V.R., and Pöschl, U. (2009). High diversity of fungi in air particulate matter. Proc. Natl. Acad. Sci. *106*, 12814.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017). Microbiome datasets are compositional: And this is not optional. Front. Microbiol. *8*, 1–6.

de Goffau, M.C., Lager, S., Salter, S.J., Wagner, J., Kronbichler, A., Charnock-Jones, D.S., Peacock, S.J., Smith, G.C.S., and Parkhill, J. (2018). Recognizing the reagent microbiome. Nat. Microbiol. *3*, 851–853.

Hoffmann, L., Günther, G., Li, D., Stein, O., Wu, X., Griessbach, S., Heng, Y., Konopka, P., Müller, R., Vogel, B., et al. (2019). From ERA-Interim to ERA5: The considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. Atmos. Chem. Phys. *19*, 3097–3214.

Jaffrezo, J.L., and Colin, J.L. (1988). Rain-aerosol coupling in urban area: Scavenging ratio measurement and identification of some transfer processes. Atmos. Environ. *22*, 929–935.

Joly, M., Amato, P., Sancelme, M., Vinatier, V., Abrantes, M., Deguillaume, L., and Delort, A.M.

(2015). Survival of microbial isolates from clouds toward simulated atmospheric stress factors. Atmos. Environ. *117*, 92–98.

Khaled, A., Zhang, M., Amato, P., Delort, A.M., and Ervens, B. (2021). Biodegradation by bacteria in clouds: An underestimated sink for some organics in the atmospheric multiphase system. Atmos. Chem. Phys. *21*, 3123–3141.

Krumins, V., Mainelis, G., Kerkhof, L.J., and Fennell, D.E. (2014). Substrate-Dependent rRNA Production in an Airborne Bacterium. Environ. Sci. Technol. Lett. *1*, 376–381.

Lallement, A., Besaury, L., Eyheraguibel, B., Amato, P., Sancelme, M., Mailhot, G., and Delort, A.M. (2017). Draft Genome Sequence of Rhodococcus enclensis 23b-28, a Model Strain Isolated from Cloud Water. Genome Announc. *5*.

Lazaridis, M. (2019). Bacteria as Cloud Condensation Nuclei (CCN) in the Atmosphere. Atmosphere (Basel). *10*, 786.

Lee, M., Heikes, B.G., and O'Sullivan, D.W. (2000). Hydrogen peroxide and organic hydroperoxide in the troposphere: a review. Atmos. Environ. *34*, 3475–3494.

Leizeaga, A., Meisner, A., Rousk, J., and Bååth, E. (2022). Repeated drying and rewetting cycles accelerate bacterial growth recovery after rewetting. Biol. Fertil. Soils *58*, 365–374.

Lighthart, B. (1997). The ecology of bacteria in the alfresco atmosphere. FEMS Microbiol. Ecol. *23*, 263–274.

Lighthart, B., and Shaffer, B.T. (1995). Viable bacterial aerosol particle size distributions in the midsummer atmosphere at an isolated location in the high desert chaparral. Aerobiologia (Bologna). *11*, 19–25.

Möhler, O., DeMott, P.J., Vali, G., and Levin, Z. (2007). Microbiology and atmospheric processes: The role of biological particles in cloud physics. Biogeosciences *4*, 1059–1071.

Moore, R.A., Hanlon, R., Powers, C., Schmale, D.G., and Christner, B.C. (2020). Scavenging of Sub-Micron to Micron-Sized Microbial Aerosols during Simulated Rainfall. Atmos. Chem. Phys 1–13.

Morris, C.E., Georgakopoulos, D.G., and Sands, D.C. (2004). Ice nucleation active bacteria and their potential role in precipitation. J. Phys. IV *121*, 87–103.

Morris, C.E., Sands, D.C., Vinatzer, B.A., Glaux, C., Guilbaud, C., Buffière, A., Yan, S., Dominguez, H., and Thompson, B.M. (2008). The life history of the plant pathogen Pseudomonas syringae is linked to the water cycle. ISME J. *2*, 321–334.

Morris, C.E., Sands, D.C., Vanneste, J.L., Montarry, J., Oakley, B., Guilbaud, C., and Glaux, C. (2010). Inferring the Evolutionary History of the Plant Pathogen Pseudomonas syringae from Its Biogeography in Headwaters of Rivers in North America, Europe, and New Zealand. MBio *1*.

Noirmain, F., Baray, J., Tridon, F., Cacault, P., Billard, H., Voyard, G., Baelen, J. Van, and Latour, D. (2022). Interdisciplinary strategy to survey phytoplankton dynamics of a eutrophic lake under rain forcing: description of the instrumental set-up and first results. Biogeosciences.

Parada, A.E., Needham, D.M., and Fuhrman, J.A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ. Microbiol. *18*, 1403–1414.

Park, C., Shin, B., and Park, W. (2019). Alternative fate of glyoxylate during acetate and hexadecane metabolism in Acinetobacter oleivorans DR1. Sci. Rep. *9*, 1–12.

Péguilhan, R., Besaury, L., Rossi, F., Enault, F., Baray, J., Deguillaume, L., and Amato, P. (2021). Rainfalls sprinkle cloud bacterial diversity while scavenging biomass. FEMS Microbiol. Ecol. 1–15.

Pouzet, G., Peghaire, E., Aguès, M., Baray, J.L., Conen, F., and Amato, P. (2017). Atmospheric processing and variability of biological ice nucleating particles in precipitation at Opme, France. Atmosphere (Basel). *8*, 18–20.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. *41*, 590–596.

Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T.F., Beutler, N.A., et al. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat. Protoc. *12*, 1261–1266.

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

Renard, P., Bianco, A., Baray, J.L., Bridoux, M., Delort, A.M., and Deguillaume, L. (2020). Classification of clouds sampled at the puy de Dôme station (France) based on chemical measurements and air mass history matrices. Atmosphere (Basel). *11*, 732.

Reysenbach, A.L., Giver, L.J., Wickham, G.S., and Pace, N.R. (1992). Differential amplification of rRNA genes by polymerase chain reaction. Appl. Environ. Microbiol. *58*, 3417–3418.

Šantl-Temkiv, T., Amato, P., Gosewinkel, U., Thyrhaug, R., Charton, A., Chicot, B., Finster, K., Bratbak, G., and Löndahl, J. (2017). High-Flow-Rate Impinger for the Study of Concentration, Viability, Metabolic Activity, and Ice-Nucleation Activity of Airborne Bacteria. Environ. Sci. Technol. *51*, 11224–11234.

Šantl-Temkiv, T., Gosewinkel, U., Starnawski, P., Lever, M., and Finster, K. (2018). Aeolian dispersal of bacteria in southwest Greenland: Their sources, abundance, diversity and physiological states. FEMS Microbiol. Ecol. *94*, 1–10.

Šantl-Temkiv, T., Sikoparija, B., Maki, T., Carotenuto, F., Amato, P., Yao, M., Morris, C.E., Schnell, R., Jaenicke, R., Pöhlker, C., et al. (2020). Bioaerosol field measurements: Challenges and perspectives in outdoor studies. Aerosol Sci. Technol. *54*, 520–546.

Sattler, B., Puxbaum, H., and Psenner, R. (2001). Bacterial growth in supercooled cloud droplets. Geophys. Res. Lett. *28*, 239–242.

Smets, W., Moretti, S., Denys, S., and Lebeer, S. (2016). Airborne bacteria in the atmosphere: Presence, purpose, and potential. Atmos. Environ. *139*, 214–221.

Smith, D.J., Timonen, H.J., Jaffe, D.A., Griffin, D.W., Birmele, M.N., Perry, K.D., Ward, P.D., and Roberts, M.S. (2013). Intercontinental dispersal of bacteria and archaea by transpacific winds. Appl. Environ. Microbiol. *79*, 1134–1139.

Smith, D.J., Ravichandar, J.D., Jain, S., Griffin, D.W., Yu, H., Tan, Q., Thissen, J., Lusby, T., Nicoll, P., Shedler, S., et al. (2018). Airborne bacteria in earth's lower stratosphere resemble taxa detected in

the troposphere: Results from a new NASA Aircraft Bioaerosol Collector (ABC). Front. Microbiol. *9*, 1–20.

Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., Prill, R.J., Tripathi, A., Gibbons, S.M., Ackermann, G., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. Nature *551*, 457–463.

Tignat-Perrier, R., Dommergue, A., Thollot, A., Magand, O., Amato, P., Joly, M., Sellegri, K., Vogel, T.M., and Larose, C. (2020). Seasonal shift in airborne microbial communities. Sci. Total Environ. 137129.

Till, A., Lakhani, R., Burnett, S.F., and Subramani, S. (2012). Pexophagy: The Selective Degradation of Peroxisomes. Int. J. Cell Biol. *2012*.

Vaïtilingom, M., Attard, E., Gaiani, N., Sancelme, M., Deguillaume, L., Flossmann, A.I., Amato, P., and Delort, A.M. (2012). Long-term features of cloud microbiology at the puy de Dôme (France). Atmos. Environ. *56*, 88–100.

Vaïtilingom, M., Deguillaume, L., Vinatier, V., Sancelme, M., Amato, P., Chaumerliac, N., and Delort, A.-M. (2013). Potential impact of microbial activity on the oxidant capacity and organic carbon budget in clouds. Proc. Natl. Acad. Sci. *110*, 559–564.

Wex, H., Stratmann, F., Topping, D., and McFiggans, G. (2008). The Kelvin versus the raoult term in the köhler equation. J. Atmos. Sci. *65*, 4004–4016.

Wirgot, N., Vinatier, V., Deguillaume, L., Sancelme, M., and Delort, A.M. (2017). H2O2 modulates the energetic metabolism of the cloud microbiome. Atmos. Chem. Phys. *17*, 14841–14851.

Womack, A.M., Artaxo, P.E., Ishida, F.Y., Mueller, R.C., Saleska, S.R., Wiedemann, K.T., Bohannan, B.J.M., and Green, J.L. (2015). Characterization of active and total fungal communities in the atmosphere over the Amazon rainforest. Biogeosciences *12*, 6337–6349.

-Abstract-

The outdoor atmosphere is a complex and dynamic environment harboring microbial assemblages of various airborne microorganisms from local and distant sources (mainly soils and vegetation). They can be transported to high altitudes and incorporate clouds. The presence of condensed water can provide potentially more viable conditions for cells. When in clouds, microbes can be redeposited on the local surface with precipitation, ending their atmospheric cycle and potentially impacting surface ecosystems. These communities have previously been demonstrated to be viable and active when aerosolized, in dry aerosols and in clouds. However, little is yet known about what they can do. The objective of this thesis work was therefore to explore the specificities of clouds as particular microbial habitats, like oases in a desert atmosphere. In this end, microbial diversity and functional profile were studied in clouds compared to other atmospheric situations such as dry aerosols and precipitation. Clouds and aerosols were collected at high altitude at the top of puy de Dôme (1,465 m a.s.l.; France), and rain was sampled at the Opme station (680 m a.s.l.; France), near the first site.

First, a comparative analysis of bacterial diversity in clouds versus precipitation was performed. It appeared that the bacterial diversity in rain was mainly to air column scavenging (biomass loading). Clouds were rather seed banks disseminating their bacterial richness with precipitation on the ecosystems.

Then, in a first approach, bacterial diversity was studied in clouds and aerosols. Communities were similar both atmospheric situations, and sequencing of sampling replicates allowed the detection of a seasonal effect.

Finally, a comparative analysis of the functional profile of clouds and aerosols was performed using metatranscriptomics and metagenomics. These are the first non-amplified metatranscriptomes and metagenomes of the outdoor atmosphere that have been obtained. A bioinformatics workflow was built to process this unique dataset. Aerosols were demonstrated to harbor microbial metabolism just like clouds. However, microbial communities were more active in clouds, highlighting the fact that clouds are specific atmospheric habitats that “revive” cells through the presence of condensed water.

-Résumé-

L'atmosphère extérieure est un environnement complexe et dynamique abritant des assemblages microbiens de divers micro-organismes aérosolisés provenant de sources locales et lointaines. Ils peuvent être transportés jusqu'à de hautes altitudes et incorporer des nuages. La présence d'eau condensée peut offrir des conditions potentiellement plus viables aux cellules. Lorsqu'ils sont dans les nuages, les microbes peuvent être redéposés à la surface avec les précipitations, ayant un impact potentiel sur les écosystèmes locaux. Il a été démontré précédemment que ces communautés étaient viables et actives lorsqu'elles étaient aérosolisées, dans des aérosols secs et dans des nuages. Cependant, on sait encore peu de choses sur ce qu'elles peuvent y faire. L'objectif de ce travail de thèse était donc d'explorer les spécificités des nuages comme habitats microbiens particuliers, comme des oasis dans une atmosphère désertique. Dans ce but, la diversité et le profil fonctionnel des communautés microbiennes dans les nuages ont été étudiés par rapport à d'autres situations atmosphériques, telles que les aérosols secs et les précipitations. Les nuages et les aérosols ont été collectés en hautes altitudes au sommet du puy de Dôme (1 465 m d'altitude ; France), et les précipitations ont été échantillonnées à la station d'Opme (680 m d'altitude, France), près du premier site.

Tout d'abord, une analyse comparative de la diversité bactérienne dans les nuages par rapport aux précipitations a été réalisée. Il est apparu que la diversité bactérienne dans la pluie provenait principalement du balayage de la colonne d'air (chargement en biomasse). Les nuages étaient plutôt des banques de souches disséminant leur richesse bactérienne avec les précipitations sur les écosystèmes.

Ensuite, dans une première approche, la diversité bactérienne a été étudiée dans les nuages et les aérosols. Les communautés étaient similaires dans les deux situations atmosphériques, et le séquençage des réplicats d'échantillonnage a permis de détecter un effet saisonnier non négligeable.

Enfin, une analyse comparative du profil fonctionnel des nuages et des aérosols a été réalisée à l'aide de la métatranscriptomique et de la métagénomique. Ceux sont les premiers métatranscriptomes et métagénomes non amplifiés de l'atmosphère extérieure qui ont été obtenus. Un flux de travail bioinformatique a été construit pour traiter cet ensemble de données unique. Il a été démontré que les aérosols abritent un métabolisme microbien tout comme les nuages. Cependant, les communautés microbiennes étaient plus actives dans les nuages, ce qui met en évidence le fait que les nuages sont des habitats spécifiques dans l'atmosphère qui "revivifient" les cellules grâce à la présence d'eau condensée.