Structuring factors of microbial communities in the atmospheric boundary layer

Romie Tignat-Perrier

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Sciences de la Terre et Univers, Environnement**

Arrêté ministériel : 25 mai 2016

Présentée par

Romie TIGNAT-PERRIER

Thèse dirigée par **Aurélien DOMMERGUE, Maître de Conférences, Institut des Géosciences de l'Environnement, Université Grenoble Alpes**, et
Catherine LAROSE, Chargée de Recherche, Laboratoire Ampère de l'Ecole Centrale de Lyon

Thèse préparée au sein de l'**Institut des Géosciences de l'Environnement à Grenoble**
dans **l'École Doctorale Terre Univers Environnement**

Facteurs de structuration des communautés microbiennes de la couche limite atmosphérique

Thèse soutenue publiquement le **22 novembre 2019**,
devant le jury composé de :

Dr. Viviane DESPRES
HDR, Max Planck Institute for Chemistry Biochemistry Department – Mainz (Allemagne), Rapportrice
Dr. Andrea FRANZETTI
HDR, University of Milano Department of Earth and Environmental Sciences – Milan (Italie), Rapporteur
Dr. Barbara D'ANNA
Directrice de Recherche, Laboratoire Chimie Environnement Université Aix Marseille – Marseille (France), Examinatrice
Dr. Pierre AMATO
Chargé de Recherche, Institut de Chimie de Clermont-Ferrand – Clermont-Ferrand (France), Examinateur
Dr. Jean MARTINS
Directeur de Recherche, Institut des Géosciences de l'Environnement Université Grenoble Alpes – Grenoble (France), Président du jury
Pr. Timothy M. VOGEL
Professeur, Laboratoire Ampère Université Lyon 1 – Lyon (France), Examinateur
Dr. Catherine LAROSE
Chargée de Recherche, Laboratoire Ampère Ecole Centrale de Lyon – Lyon (France), Directrice de thèse
Dr. Aurélien DOMMERGUE
Maître de conférences, Institut des Géosciences de l'Environnement Université Grenoble Alpes – Grenoble (France), Directeur de thèse

Communauté
UNIVERSITÉ Grenoble Alpes

PHD THESIS

Toward the title of

DOCTOR OF SCIENCE OF THE GRENOBLE ALPES UNIVERSITY COMMUNITY

Speciality: Earth and Universe Sciences

Arrêté ministériel : 25 mai 2016

Presented by

Romie TIGNAT-PERRIER

PhD thesis supervised by **Aurélien DOMMERGUE, Maître de Conférences, Institut des Géosciences de l'Environnement, Université Grenoble Alpes** and
Catherine LAROSE, Chargée de Recherche, Laboratoire Ampère de l'Ecole Centrale de Lyon

PhD thesis prepared in the **Institut des Géosciences de l'Environnement in Grenoble**
and the **Terre Univers Environnement** doctoral school

Structuring factors of microbial communities in the atmospheric boundary layer

PhD defense on the **22th of November, 2019**, in Grenoble
In front of the following jury members:

Dr. Viviane DESPRES
HDR, Max Planck Institute for Chemistry Biochemistry Department – Mainz (Germany), Reviewer
Dr. Andrea FRANZETTI
HDR, University of Milano Department of Earth and Environmental Sciences – Milan (Italy), Reviewer
Dr. Barbara D'ANNA
DR, Laboratoire Chimie Environnement Université Aix Marseille – Marseille (France), Examinator
Dr. Pierre AMATO
CR, Institut de Chimie de Clermont-Ferrand – Clermont-Ferrand (France), Examinator
Dr. Jean MARTINS
DR, Institut des Géosciences de l'Environnement Université Grenoble Alpes – Grenoble (France), President of the jury
Pr. Timothy M. VOGEL
Professor, Laboratoire Ampère Université Lyon 1 – Lyon (France), Examinator
Dr. Catherine LAROSE
Chargée de Recherche, Laboratoire Ampère Ecole Centrale de Lyon – Lyon (France), Supervisor
Dr. Aurélien DOMMERGUE
Maître de conférences, Institut des Géosciences de l'Environnement Université Grenoble Alpes – Grenoble (France), Supervisor

Acknowledgements

First of all, I wish to thank my supervisors Catherine Larose and Aurélien Dommergue, as well as my adviser Timothy M. Vogel. I thank you Tim for being one of my best university lecturer. I am grateful to you for telling me about this project on atmospheric microbial communities. Thank you Cath and Aurélien for being supporting and trusting me. I loved discussing with you three about my work, I thank you for everything I learnt from you. These three years have been intense and the PhD a real training work where I developed many skills. I really appreciated to have the opportunity to teach at the University of Lyon 1, to supervise a master student (Myriam Spajer), to do oral and poster presentations in international conferences and to develop English oral and written skills.

I would like to thank the members of my annual comities, *i.e.* Pierre Amato (Institut de Chimie de Clermont-Ferrand), Barbara d'Anna (LCE de Marseille), Alexandre Poulain (University of Ottawa) and Eric Bapteste (IBPS de Paris) for having accepted to discuss about my work and for giving me advices. I thank the members of my PhD jury, Jean Martins (IGE Grenoble), Barbara d'Anna, Pierre Amato, Andrea Franzetti (University of Milano, Italy) and Viviane Després (Max Planck Institute for Chemistry in Mainz, Germany) for having accepted to evaluate my work. I thank the Région Auvergne-Rhône Alpes for having financed my PhD.

I thank the team of the IGE in Grenoble for having welcomed me. I thank particularly Jean-Luc Jaffrezo, Jean Martins and the PhD student Abdoulaye for our discussions on atmospheric chemistry and atmospheric microorganisms. It was nice to be in regular contact with you Abdoulaye.

I could really not have hoped for better coworkers at the Laboratoire Ampère, Ecole Centrale de Lyon: Mia, Concepcion, Rose, Benoît, Arthur and Adrien (all PhD students) as well as Christoph, Cécile, Graeme, Christina, Laure and Pascal. It was so nice to work with you and I would like to keep you all and bring you with me to my next lab. To PhD students: it was really nice to share this PhD experience with you, you were a real every day motivation and by far an important element that made my PhD succeed. I would like to thank particularly Mia: we spent so much time together in the lab during these three years, your presence was so motivating and reassuring, you helped me a lot.

It was really nice to meet Nora from Innsbruck and Frederik from Copenhagen, both PhD students invited in our lab. We also spent good times in conferences.

I thank Laurent Pouilloux, a competent and nice computer scientist from the ECL, for helping me on some of my scripts and for always answering me rapidly when I got computing problems.

I will finish by thanking my family as well as my life partner Robin who made big concessions to be at my sides in Lyon and gave me endless support.

“All models are wrong, but some are useful.”

George Box, statistician at the University of Wisconsin

Résumé (French)

La couche limite planétaire est la couche atmosphérique la plus basse qui est en interaction directe et constante avec les surfaces terrestres et marines sur lesquelles se concentrent les activités humaines, les cultures et divers écosystèmes. Comprendre l'origine de sa composition à la fois chimique et microbiologique est fondamental dans notre étude approfondie de la biosphère. Alors que les microorganismes de la couche limite planétaire – retrouvés jusqu'à 10^6 cellules par mètre cube d'air – semblent varier significativement à l'échelle spatiale et temporelle en termes de concentration et de diversité, ils restent largement méconnus. L'objectif principal de cette thèse est de comprendre comment se structurent les communautés microbiennes dans la troposphère, et en particulier dans la couche limite planétaire, ainsi que d'identifier les facteurs de contrôle majeurs. En travaillant sur des échantillons collectés pendant plusieurs semaines sur neuf sites répartis sur la planète, et en utilisant les technologies de séquençage ADN haut-débit, nous avons étudié la composition taxonomique et fonctionnelle des communautés microbiennes de la phase gazeuse et solide de l'atmosphère (c'est-à-dire non associés aux nuages).

Nos premiers résultats sur la taxonomie des communautés microbiennes révèlent que les surfaces proches des sites sont les contributeurs principaux de distribution des communautés microbiennes atmosphériques, malgré l'occurrence potentielle du transport longue-distance des microorganismes atmosphériques. Egalement, les conditions météorologiques combinées à la diversité des surfaces locales terrestres ou océaniques jouent un rôle important dans la variation temporelle de la structure des communautés microbiennes de la couche limite planétaire. Une deuxième étude nous a permis d'étudier davantage la variation temporelle des communautés microbiennes atmosphériques sur un site continental montagneux en France (1465 m d'altitude) sur une année complète. Cette étude révèle l'importance des conditions de surface des paysages aux alentours dans la composition taxonomique des communautés atmosphériques. L'évolution au cours de l'année des terres agricoles et de la végétation, qui composaient en majeure partie le paysage du site, était responsable du changement temporel observé dans la composition taxonomique des communautés microbiennes atmosphériques. Finalement, nous avons étudié la composition fonctionnelle des communautés microbiennes de la couche limite planétaire afin d'identifier si les conditions physiques et chimiques de l'atmosphère jouaient un rôle dans la sélection ou

adaptation microbienne des microorganismes atmosphériques. L'analyse comparative de données métagénomiques ne révèle pas de signature atmosphérique spécifique du potentiel fonctionnel des communautés microbiennes atmosphériques. La composition fonctionnelle semble avant tout liée aux écosystèmes locaux. Toutefois, nous avons observé que les champignons étaient plus dominants relativement aux bactéries dans l'air comparativement aux autres écosystèmes. Ce résultat suggère un processus de sélection des champignons durant l'aérosolisation et/ou le transport aérien. Les champignons pourraient survivre davantage l'aérosolisation et le transport aérien comparativement aux bactéries du fait de leur résistance naturelle aux conditions physiques stressantes de l'atmosphère. Nos résultats ont apporté une meilleure compréhension des facteurs déterminants (c'est-à-dire les surfaces locales, les sources distantes, les conditions météorologiques locales, les conditions physiques stressantes de l'atmosphère) et de leur contribution dans la structuration des communautés microbiennes de la couche limite atmosphérique. Nos investigations constituent une base importante pour de nouvelles études sur la prévision et le contrôle des communautés microbiennes atmosphériques, afin de répondre à des questions majeures dans les domaines de la santé publique et de l'agronomie.

Mots clefs : communautés microbiennes de la couche limite planétaire, microorganismes atmosphériques, transport longue-distance, potentiel fonctionnel, métagénomique comparative, séquençage ADN haut-débit, séquençage MiSeq Illumina, sélection physique, adaptation microbienne, aérosolisation

Abstract

Up to 10^6 microbial cells per cubic meter are found in suspension in the planetary boundary layer, the lowest part of the atmosphere. Direct influences of the planetary boundary layer on humans, crops and diverse ecosystems like soils and oceans make the full understanding of its composition, both chemical and microbiological, of utmost importance. While microbial communities of the planetary boundary layer vary significantly at different temporal and spatial scales, they remain largely unexplored. The main goal of this thesis was to understand how airborne microbial communities are structured in the troposphere with special emphasis on the planetary boundary layer and to identify their main controlling factors. We investigated both the taxonomic and functional composition of airborne microbial communities in the dry phase (*i.e.* not cloud-associated) over time at nine different geographical sites around the world using high throughput sequencing technologies.

Our investigation that focused on microbial taxonomy showed that local landscapes were the main contributors to the global distribution of airborne microbial communities despite the potential occurrence of long-range transport of airborne microorganisms. We also observed that meteorology and the diversity of the surrounding landscapes played major roles in the temporal variation of the microbial community structure in the planetary boundary layer. We further explored the temporal variation of airborne microbial communities at a continental and mountainous site in France (1465 m above sea level) over a full-year. This study demonstrated the importance of the surface conditions (*i.e*. vegetation, snow cover *etc.*) of the surrounding landscapes on the taxonomic composition of airborne microorganisms. The seasonal changes in agricultural and vegetated areas, which represented a significant part of the site's surrounding landscape, were correlated to the shifts in the taxonomic composition of airborne microbial communities during the year. Finally, we investigated the functional composition of microbial communities of the planetary boundary layer to identify whether the physical and chemical conditions of the atmosphere played a role in selection or microbial adaptation of airborne microorganisms. The comparative metagenomic analysis did not show a specific atmospheric signature in the functional potential of airborne microbial communities. To the contrary, their functional composition was mainly correlated to the underlying ecosystems. However, we also showed that fungi were more dominant relatively to bacteria in air as compared to other (planetary bound) ecosystems. This result suggested a selective

process for fungi during aerosolization and/or aerial transport and that fungi might likely survive aerosolization and/or aerial transport better than bacteria due to their innate resistance to stressful physical conditions (*i.e.* UV radiation, desiccation *etc*.). Our results provide a clearer understanding of the factors (*i.e.* surrounding landscapes, distant sources, local meteorology, and stressful physical atmospheric conditions) that control the distribution of microbial communities in the atmospheric boundary layer. Our investigations provide a basis for further studies on the prediction and even control of airborne microbial communities that would be of interest for public health and agriculture.

Keywords: airborne microbial communities, atmospheric microorganisms, planetary boundary layer, long-range transport, functional potential, high throughput sequencing, MiSeq Illumina sequencing, physical selection, microbial adaptation, aerosolisation, comparative metagenomics

Table of content

List of abbreviations

ANOVA= analysis of variance

BLAST= basic local alignment search tool

DNA= deoxyribonucleic acid

EC=elemental carbon

MG-RAST= metagenomic rapid annotations using subsystems technology

MODIS= moderate resolution imaging spectroradiometer

NR= non-redundant protein database

OC= organic carbon

PANDAseq= paired-end assembler for Illumina sequences

PCoA= principal coordinates analysis

PM= particulate matter

HYSPLIT= hybrid single particle lagrangian integrated trajectory model

qPCR= quantitative polymerase chain reaction

RDA= redundancy analysis

RNA= ribonucleic acid

ROS= reactive oxygen species

RDP= ribosomal database project

UV= ultra-violet

WHO= World Health Organization

List of figures and tables

Figure S3: Hierarchical cluster analysis (average method) on the Euclidean distance matrix calculated on the PM10 chemistry data.
Figure S4: Distance-based RDA analyses.
Figure S5: Wind roses covering the sampling time at each site.
Figure S6. Backward trajectories calculated over 3 days at each site using HYSPLIT.
Figure S7: Q-Q plots of the multiple linear regressions.
Table S1: Information on samples.
Table S2: Presentation of the different MODIS land covers from Friedl *et al.*, 2002.
Table S3: Estimation of mean bacterial cell concentration per cubic meter of air in near-surface air above the different landscapes reported in Burrows *et al.*, 2009.
Table S4: Total abundance (number of annotated sequences), contribution in each site and relative abundance per site of the first fifty most abundant bacterial genera.
Table S5: Total abundance (number of annotated sequences), contribution in each site and relative abundance per site of the first fifty most abundant fungal species.
Table S6: Bacterial genera and fungal species characterizing the different sites or groups of sites identified using hierarchical cluster analyses based on both bacterial and fungal community structures.
Table S7: Average concentration (and standard deviation) of the chemicals per site in ng/m^3 of air.
Table S8: Multiple linear regression results.

Chapter 4
Figure 1: Central position of Puy-de-Dôme in France and relative surfaces of the landscapes surrounding the site in a perimeter of 50 km based on the MODIS land surfaces.
Figure 2: Log10 of the bacterial and fungal cell concentration estimated by the number of 16S and 18S rRNA gene copies per cubic meter of air and per sample.
Figure 3: Hierarchical cluster analysis (average method) on the airborne bacterial community structure and airborne fungal community structure based on the Bray-Curtis dissimilarity matrices.
Figure 4: Trophic mode of the fungal species.
Figure 5: Temporal evolution of the relative abundance of several fungal species and bacterial genera over the year.
Figure 6: Temporal evolution of the concentration of chemical species over the year.
Figure S1: Monthly NASA satellite images of Puy-de-Dôme surrounding surfaces.
Figure S2: Rarefaction curves of the number of bacterial genera and fungal species per season.
Figure S3: Venn diagrams showing the number of shared and unique bacterial genera and fungal species between the samples of each season after rarefaction.
Figure S4: Temporal evolution of the relative abundance of four fungal phytopathogens over the year.
Figure S5: Heatmap showing the temporal evolution of the relative abundances of the fifty most abundant bacterial genera and fungal species in the dataset.
Figure S6: Hierarchical cluster analysis (average method) on the particulate matter chemical concentrations based on the Euclidean distance matrix.
Figure S7: Constrained Analysis of Principal Coordinates.
Figure S8: Temporal evolution of the temperature and relative humidity over the year.
Figure S9: Meteorological conditions per season.
Figure S10: Air mass backtrajectories over 3 days at each season.

Table S1: Information on the samples.
Table S2: Bacterial and fungal concentration and richness averaged per season.
Table S3: Top 25 of the bacterial genera and fungal species observed in Puy-de-Dôme, and average percentages of these bacterial genera and fungal species per season.
Table S4: Weekly variability of the bacterial and fungal composition per season
Table S5: Richness in fungal species annotated as pathotroph, symbiotroph and saprotroph averaged per season and according to FUNGuild.
Table S6: Grouping of the fifty most abundant bacterial genera and fungal species based of the trend of their relative abundance over the year.
Table S7: Chemical concentrations averaged per season.

Chapter 5
Figure 1: Map showing the geographical location and elevation from sea level of our nine sampling sites and the geographical position of whose public metagenomes come from.
Figure 2: Percentage of fungi-associated sequences and bacteria-associated sequences over the total number of sequences annotated as belonging to fungal and bacterial genomes in the metagenomes.
Figure 3: PCoA analysis of the Bray-Curtis dissimilarity matrix based on the functional potential structure of each site.
Figure 4: Average number of hits of sporulation related functions per 10000 annotated sequences from all sequences, fungi-associated sequences and bacteria-associated sequences per site.
Figure 5: Average number of hits of UV protection related functions per 10000 annotated sequences from all sequences, fungi-associated sequences and bacteria-associated sequences per site.
Figure 6: Average number of hits of oxidative stress cell response related functions per 10000 annotated sequences from all sequences, fungi-associated sequences and bacteria-associated sequences per site.
Figure 7: Average number of hits of cell death related functions per 10000 annotated sequences from all sequences, fungi-associated sequences and bacteria-associated sequences per site.
Figure 8: Average number of hits of desiccation response related functions per 10000 annotated sequences from all sequences, fungi-associated sequences and bacteria-associated sequences per site.
Figure 9: Conceptual model that might explain the higher ratio between fungi and bacteria in air relatively to airborne microbial cell sources such as soil and ocean water.
Table S1: Information on the sites.
Table S2: Standardized collected air volume and sampling starting date of each air sample that we collected for this study.
Table S3: Functional richness and evenness after rarefaction per site, based on the SEED functional classes.
Table S4: qPCR on the 16s rRNA gene and on the 18S rRNA gene on air and soil samples, and ratio between these qPCRs.
Table S5: Top 50 of the SEED functions observed in the air samples considering all the sequences (*i.e.* bacteria and fungi-associated sequences).
Table S6: SIMPER analysis results on groups identified on the PCoA analysis based on all sequences.

Figure S1: Surrounding landscapes of the sampling sites in a 50 km perimeter based on the land cover MODIS approach.
Figure S2: PCoA analysis of the Bray-Curtis dissimilarity matrix based on the functional potential structure of each air sample.
Figure S3: PCoA analysis of the Bray-Curtis dissimilarity matrix based on the functional potential structure of each site.
Figure S4: Average number of hits of hydrogen peroxide catabolic process related functions per 10000 annotated sequences from all sequences, fungi-associated sequences and bacteria-associated sequences per site.
Figure S5: Functional richness after rarefaction from all sequences, fungi-associated sequences and bacteria-associated sequences per site.
Figure S6: Average number of hits of methane monooxygenase process related functions per 10000 annotated sequences from all sequences per site.
Figure S7: Average number of hits of lipoate synthase and chromosome plasmid partitioning protein ParA per 10000 annotated sequences from all sequences per site.

Chapter 6
Figure 1: Overview of the role of different factors in controlling microbial communities of the planetary boundary layer.

List of peer-reviewed publications

**Publications on the thesis work*

Dommergue A., Amato P., **Tignat-Perrier R.**, Magand O., Thollot A., Joly M., Bouvier L., Sellegri K., Vogel T.M., Sonke J.E., Jaffrezo JL, Andrade M., Moreno I., Labuschagne C., Martin L., Zhang Q. and Larose C. Methods to investigate the global atmospheric microbiome. Front. Microbiol. 2019. ***Published in Frontiers in Microbiology***.

Tignat-Perrier R., Dommergue A., Thollot A., Keuschnig C., Magand O., Vogel T.M., Larose C. Global airborne microbial communities controlled by surrounding landscapes and wind conditions. ***Under review in Scientific Reports.***

Tignat-Perrier R., Dommergue A., Thollot A., Magand O., Vogel T.M., Larose C. Microbial functional signature in the planetary atmospheric boundary layer. ***In preparation for publication in Environmental Microbiology.***

Tignat-Perrier R., Dommergue A., Thollot A., Magand O., Amato P., Vogel T.M., Larose C. Seasonal changes of local landscapes drive airborne microbial community variation. ***In preparation for publication in FEMS.***

**Side publications*

Els N., Larose C., Baumann-Stanzer K., **Tignat-Perrier R.**, Keuschnig C., Vogel T.M., Sattler B. Microbial Composition in Seasonal Timeseries of Free-Tropospheric Air and Precipitation Reveals Community Separation. 2019. ***Published in Aerobiologia.***

Sanchez-Cid C., **Tignat-Perrier R.**, Franqueville L., Delaurière L., Schagat T., Vogel T.M. Sequencing depth rather than DNA extraction determines soil bacterial richness discovery. ***Submitted in Scientific Reports.***

Chapter 1: Bibliography - Airborne microbial communities of the troposphere

Introduction

The study of microorganisms in the atmosphere goes back to the 19th century when Pasteur cultivated microorganisms from the air. At that time, the principal issue was tracing the spread of diseases to the aerial dispersion of bacterial pathogens. Later, in the late 20th century, with the discovery of a rich airborne microbial world, attention broadened to the origin, transport dynamics, survival, as well as the airborne microorganism's role in air quality, meteorology and chemical cycles within the atmosphere.

The lowest part of the atmosphere that surrounds the Earth's surface is the troposphere. It is the densest and most dynamic layer in terms of chemistry and physics of aerosols (*i.e.* atmospheric particles). It represents a large volume of 6.6 billion cubic kilometers that is five times larger than that of oceans (1.3 billion cubic kilometers). The troposphere harbors complex chemical reactions and meteorological phenomena that are intricately linked and lead to the coexistence of gas, solid (*i.e.* particulate matter from a size of a few nanometers to millimeters like sand dust) and liquid phases (*i.e.* clouds, rain, fog).

These phases represent different potential biological niches (**Fig. 1**) that are all studied in aeromicrobiology (supplementary **Table S1**). Based on the literature[1–3], these niches harbor different microbial communities (**Fig. 2**) and different airborne microbial concentrations that might be due to significant differences in physico-chemical characteristics that constrain either microbial life within the niches or the destruction of specific members from their source. For example, liquid-phase associated microorganisms (*i.e.* cloud, rain and fog water) might be particularly different from microorganisms in the dry phase of the atmosphere (*i.e.* free microorganisms suspended in the gas phase and microorganisms attached to particulate matter such as sand dust) (**Fig. 2**).

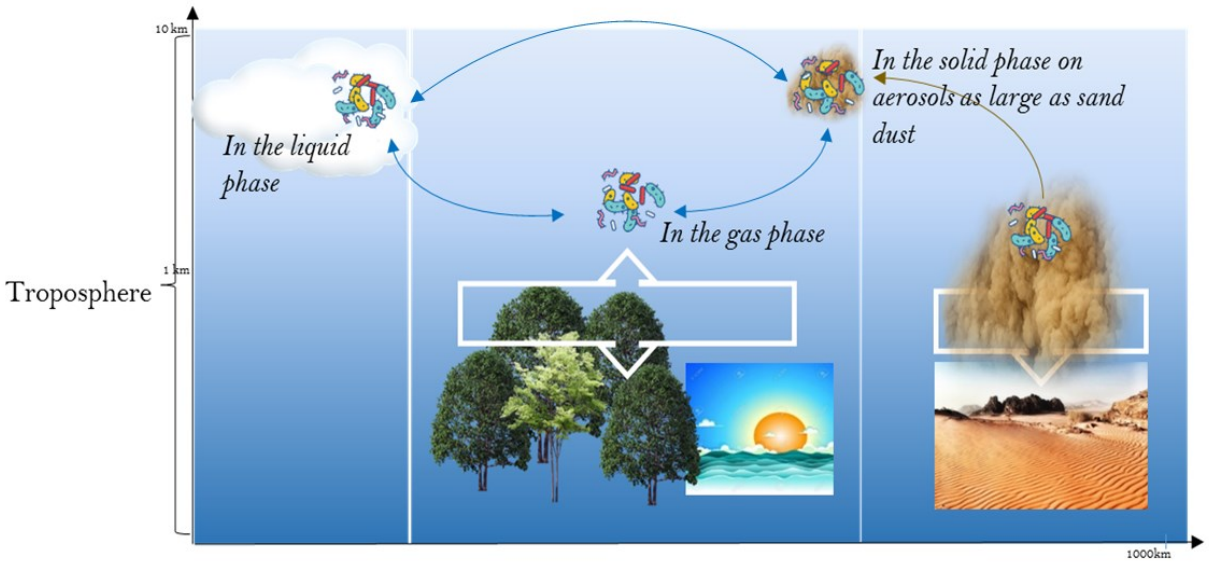


Figure 1: The different biological niches existing in the troposphere. Airborne microorganisms have been observed in the liquid phase (*i.e.* cloud, fog and rain water), gas phase (as free cells) and solid phase (as cells attached to aerosols from a microscopic to macroscopic size such as sand dust). These different niches might exchange microorganisms and represent different physico-chemical conditions for airborne microorganisms.

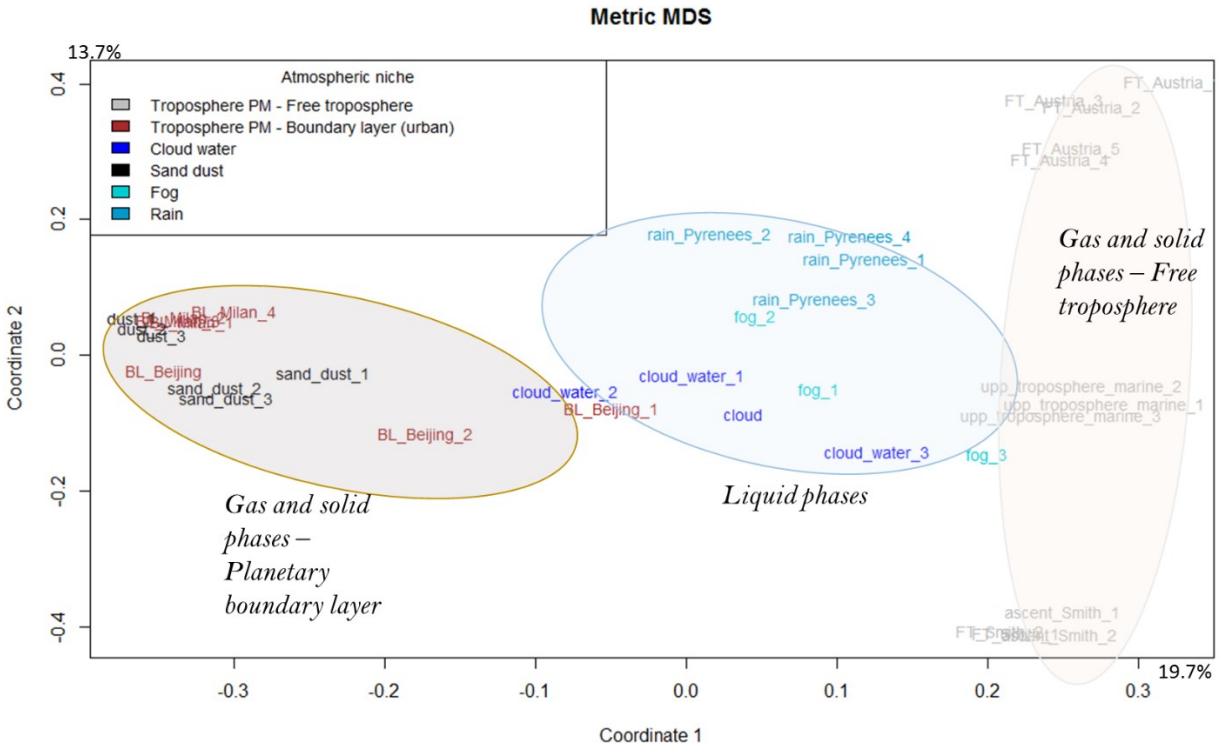


Figure 2: PCoA analysis of the Bray-Curtis dissimilarity matrix based on the bacterial community structure (amplicon sequencing) of samples coming from cloud water, rain, fog, planetary boundary layer and free troposphere particulate matter. The corresponding studies are listed in supplementary **Table S2**.

In addition to their aqueous nature, clouds are characterized by specific chemistry and physics compared to the dry troposphere. Clouds dissolve chemical species in water and support new and/or faster chemical reactions[4] that lead to an enhanced formation of secondary organic aerosols as well as strong oxidants like H_2O_2 and radicals[5,6]. Clouds are at the heart of current investigations on airborne microbial metabolism. In the laboratory, several culture-based and microcosm studies have shown that cloud-associated microorganisms can degrade the main carboxylic compounds found in cloud water (*i.e.* formate, acetate, formaldehyde[7]). Microbial communities were shown to influence the oxidative capacity of clouds through the reduction of oxidants like H_2O_2[6]. These chemical specificities are associated with physical processes like the ongoing presence of successive condensation-evaporation cycles and freeze-thaw cycles of the cloud water[8–12]. These processes create a highly specific biological niche in clouds that might constrain microbial life[13]. Clouds might themselves be formed on microbial cells originating from the dry troposphere as they offer a surface for the condensation of water vapor and act as cloud condensation nuclei (CCN). The role of microorganisms as CCN and ice nuclei in cloud formation has been intensively investigated[14–18].

Cloud-associated microbial communities and more generally liquid-phase associated microorganisms are not the center of interest of this PhD. Rather, we focused on free and particulate-matter associated microorganisms found in the dry phase of the troposphere (*i.e.* gas and solid phases). Moreover, we were specifically interested in the lower part of the troposphere (*i.e.* the planetary boundary layer – **Fig. 3**) representing the atmosphere that directly surrounds us. The way the planetary boundary layer interacts with humans, crops and diverse ecosystems like soils and oceans makes the understanding of its microbial composition of utmost importance. While planetary boundary layer microorganisms (from the dry phase) seem to vary significantly over space and time, the number of investigations remains limited (supplementary **Table S1**). Moreover, investigations on planetary boundary layer microorganisms are mainly of small spatial-scale (**Fig. 4**) and use different sampling strategies (supplementary **Table S1**) making them difficult to compare.

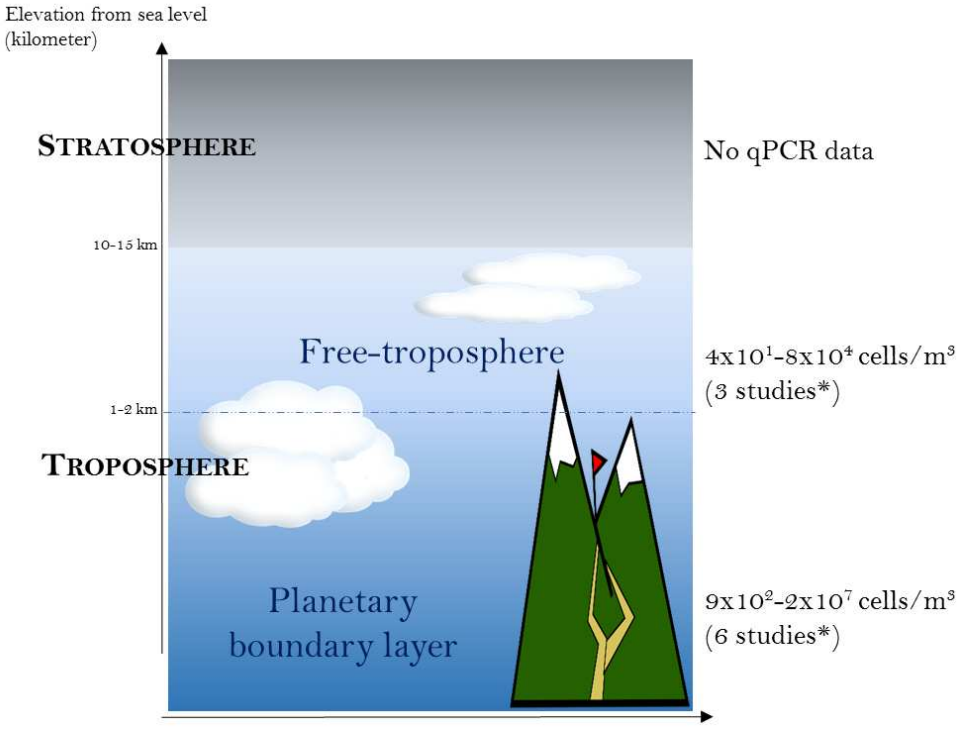


Figure 3: Vertical stratification of the first layers of the atmosphere and concentration range of the number of bacterial cells per cubic meter of air based on qPCR data in these layers.
**Free troposphere related studies:* Tanaka *et al.* (2019)[19], Zweifeil *et al.* (2012)[20], DeLeon-Rodriguez *et al.* (2013)[21]. Boundary layer related studies: Bertolini *et al.* (2013)[22], Zhen *et al.* (2017)[23], Genitsaris *et al.* (2017)[24], Gandolfi *et al.* (2015)[25], Cho *et al.* (2011)[26], Park *et al.* (2018)[27].

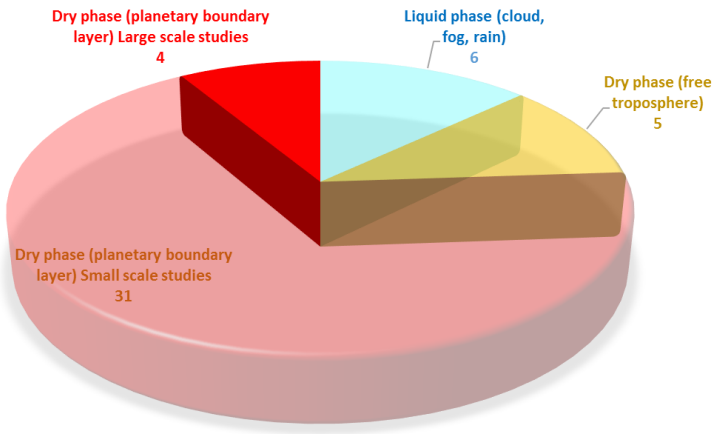


Figure 4: Grouping of the investigations on the airborne microbial community taxonomic structure using high through put sequencing (not cloning – see supplementary **Table S1**) depending on the atmospheric niche studied (*i.e.* liquid versus dry phase, planetary boundary layer versus free troposphere) and the scale of the study (*i.e.* small or large scale studies; only for the dry phase of the planetary boundary layer related studies). Small scale studies include studies at one or a few sites; large scale studies include studies at several sites (regional or continental scale). The number of studies per group is indicated under the group name.

This Chapter will review knowledge and knowledge gaps regarding major questions in microbial ecology of the troposphere such as "*Are airborne microorganisms undergoing a selective process during aerosolization?*", "*Is airborne microbial functional composition related to the physical and chemical conditions characterizing the atmosphere?*", "*Are airborne microorganisms metabolically active?*" and "*What factors structure airborne microbial communities at varying temporal and spatial scales?*" with a special interest for the dry troposphere. When necessary, we will present data and hypotheses obtained from investigations of airborne microbial communities from other atmospheric biological niches like clouds. We will discuss their relevance and their potential inference to airborne microbial communities of the dry troposphere.

We will start by describing the microbiological characteristics of the troposphere, especially in the gas and solid phases (*i.e.* the dry troposphere). We will then present potential interactions between airborne microbial communities and their physico-chemical environment as well as the potential contribution of airborne microorganisms to biogeochemical cycles and the role of the physico-chemical environment on constraining airborne microbial life. Finally, we will identify the main factors potentially responsible for the geographical distribution and temporal dynamics of troposphere microbial communities. The review will be followed by an outline of the objectives, hypotheses and relevance of the thesis work in this context.

Microbiological characteristics of the troposphere

Bioaerosols in the troposphere

The atmosphere consists of a layer of gases that is retained around the Earth by gravity. It is vertically stratified from the Earth's surface to the exosphere, which is the upper layer separating the Earth's atmosphere from space. The first layer that immediately surrounds the Earth is called the troposphere (**Fig. 3**). The troposphere is the densest layer in terms of aerosols and, consequently, the most dynamic layer in terms of aerosol physics and chemistry. Aerosols are tiny particles floating in the air and created either by condensation of gases or directly emitted by the Earth's surface. We will use either the term particulate matter (PM) or aerosols to refer to atmospheric particles, although aerosols consist of both the suspended particles (*i.e.* particulate matter) and their surrounding gas. The troposphere is also the layer

with the largest concentration of bioaerosols. Bioaerosols can represent around 30% of the atmospheric aerosol load (aerosols > 1 µm) in urban and rural air, and up to 80% in pristine rainforest air[28–32]. Bioaerosols are aerosols of biological origin and include plant debris, pollen, microorganisms (bacteria, fungal cells and spores, viruses as well as larger eukaryotic microorganisms like protozoans), and biological secretions. In the tropospheric layer, microbial cell concentrations can reach up to 10^6 cells per cubic meter of air in its lowest and densest part close to the ground (**Fig. 3**). The troposphere height extends up to ten to fifteen kilometers above sea level and the stratosphere is above it. The stratosphere is a layer of free atoms in which radical chemical reactions occur and is commonly called the ozone layer[4]. It is a layer with low air density and is composed of low particulate matter and ion concentrations and likely very low microbial concentrations[33]. Microorganisms have been isolated from as high as 77 km in the stratosphere[34], although it might represent the upper altitudinal limit. These microorganisms found in the stratosphere remained viable[35,36], even though the stratosphere combines challenging conditions, such as cold temperatures down to -70°C, exposure to destructive UV radiations, intense desiccation due to the extremely low relative humidity (down to 0%), high radical concentrations, hypobaric conditions, and low nutrient concentrations *etc*.

Tropospheric aerosol properties, mainly number and mass concentration, size distribution, chemical composition, over time, space and geography have been thoroughly studied[37]. Although investigations on the concentration and diversity of airborne microbial communities are becoming more numerous, their distribution on particulate matter remains unknown. Airborne microbial cells might exist mainly as aggregates or attached to particulate matter, while airborne fungi might exist mainly as single spores[38]. Particulate matter size range is broad, starting from less than one nanometer (example of secondary aerosols created by the condensation of gases) up to hundreds of micrometers like sand dust. Microbial cells entering freely in the atmosphere might attach to existing particulate matter or other microbial cells[39]. Conversely, particle-attached microbial cells might detach from their support in the air. Based on a compilation of data from more than one hundred investigations, Clauss *et al.* (2015)[38] determined that 15% of cultivable airborne bacterial cells were on particles < 2.1 µm (size) and 25% on particles > 7.2 µm, and that cultivable airborne fungal spores and cells were mainly distributed on particles between 1 and 3.2 µm (median-based values) on average in outdoor air. The distribution size was shown to depend on aerosolization processes occurring and

different meteorological conditions at the time of aerosolization such as air relative humidity[38]. Using a culture-independent concentration estimation approach, Sippula *et al.* (2013)[40] also observed that 77% of total bacteria were found in the size fraction > 2.4 µm.

Airborne microbial concentrations tend to show a vertical gradient with high concentrations near the Earth's surface and decreasing concentrations with altitude up through the troposphere and into the stratosphere[2,19] (**Fig. 3**). Simultaneous airplane samples collected at 690, 1000 and 3127 m altitude above sea level showed decreasing numbers of cultivable microorganisms per cubic meter of air with increasing altitude[41]. While airborne microbial concentrations have been repeatedly measured in the lower troposphere around our planet using culture-dependent and molecular analyses[23–27,42–44], cell concentrations in the upper troposphere and stratosphere remain unknown. To date, only one culture-independent investigation has been carried out in the stratosphere[33]. Transport of particles and microbial cells might be restricted between the upper troposphere and the stratosphere. While the upper troposphere supplies the stratosphere in microorganisms and chemical species involved in global ozone depletion, the stratosphere is a source of free radicals for the upper troposphere[4]. Although unknown, the stratosphere might support a low microbial diversity formed of microorganisms resistant to the extreme environmental conditions[33] (see next section). Current knowledge about stratospheric microorganisms is scarce and almost exclusively based on culture-based approaches, which limits investigations to a small fraction (< 1%) of the whole community. In-flight collection of stratospheric microorganisms remains expensive and an engineering challenge, which due to the very low concentration of cultivable cells in the stratosphere suffers from aircraft-associated contamination.

Planetary boundary layer versus free troposphere and vertical distribution of airborne microbial communities

The troposphere is itself divided in two layers, the planetary boundary layer and the free troposphere (**Fig. 3**). The planetary boundary layer interacts with the Earth's surface and represents the densest layer harboring the largest concentration of particulate matter and bioaerosols[4]. This PhD will focus specifically on the planetary boundary layer because it has the highest concentration of bioaerosols and it is the layer that surrounds most of the human population. Air mass dynamics within the planetary boundary layer are subject to mechanical and thermal convective turbulence that is controlled in part by the ground roughness and

Earth's surface heat[4]. As a consequence, the boundary layer's height changes according to location and even time of day throughout the year. In contrast, the free troposphere is driven horizontally by geostrophic wind and is vertically stable.

Aerosolized microbial cells generally enter the lowest layer of the atmosphere, the planetary boundary layer, from which an as yet unknown quantity might be transferred to the free troposphere. Some microbial inputs from the Earth's surface could directly reach the free troposphere during specific and violent meteorological events like dust storms or volcanic eruptions. The free troposphere can also be in direct contact with the Earth's surface. For example, surfaces of mountain peaks like Storm-Peak (+ 3220 m, Colorado, USA) and Pic-du-Midi (+ 2876 m, France) can at times be in the free troposphere during the night. Puy-de-Dôme site (+ 1465 m; France) is known to be influenced by free tropospheric air masses during certain meteorological conditions (more often in winter).

Once in the planetary boundary layer or the free troposphere, microbial cells will be transported over short or long distances depending on the meteorological conditions (windy, wet and dry precipitation), and if they travel freely or attached to a bigger particle. In the planetary boundary layer, airborne microorganisms have been shown to have a residence time of a few days before returning to the Earth's surface due to gravitation or precipitation if they behave like non biological aerosols[45]. In the free troposphere, their residence time might be several days during which they might be transported over thousands of kilometers[46]. Despite an obvious continuum of the troposphere and because of differences in aerosol dynamics, chemical composition and physical conditions, the planetary boundary layer and the free troposphere might support different microbial communities[2,19] (**Fig. 2**). The free troposphere is more stable than the planetary boundary layer in terms of physics and aerosol dynamics and might also be more stable in airborne microbial concentration and composition. Studies on the structure of oceanic microbial communities reported a biogeography and a vertical distribution of the microbial diversity[47–51]. Like the atmosphere, the ocean represents a continuous fluidic environment that is vertically stratified. Some microbial oceanographic studies showed that microbial community structure and abundance in the surface layers are less stable than those in the mesopelagic zone (the layer under the surface layer) due to abiotic and biotic perturbations that make the marine surface layer environment less stable[49,52]. Some studies on the vertical distribution of airborne microbial communities showed different communities in the planetary boundary layer and the free troposphere, and

suggested that some microbial taxa might be filtered out during vertical transport[2]. Due to their size, the largest and densest airborne microbial cells might be less prone to reaching the free troposphere than lighter cells. This hypothesis was supported by the observed increase in the ratio between bacteria and fungi at a remote mountain site in Austria (+ 3106 m)[2]. Another explanation could be that microbial cells floating in the free troposphere have more time to undergo selection and adaptation to the abiotic conditions as compared to those in the planetary boundary layer, so that only the microorganisms that are the most resistant to the harsh tropospheric conditions (UV radiation, cold temperature, radicals etc.) survive. Thermophilic strains with high resistance towards extreme conditions, which are often identified in heavy dust events, were shown to be ubiquitous and significantly increased in relative abundance in the free troposphere as compared to the planetary boundary layer at a remote mountain site in Austria (+ 3106 m)[2].

Interactions between physico-chemical characteristics and airborne microorganisms in the troposphere

Physico-chemical characteristics of the atmosphere that might constrain microbial life

A variety of chemical substances interact with tropospheric microorganisms and subsequently might have an effect on them. Solid and liquid aerosols can be composed of essential nutrients (carbon, oxygen, nitrogen, phosphorus, sulfur, hydrogen, trace minerals) that could be used as source of energy and matter for microbial metabolism. The troposphere also contains free radical species (the hydroxyl radical OH is the most common and reacts with nearly every chemical species) and compounds at potential toxic concentrations (heavy metals, persistent organic compounds, antibiotics) that may have negative effects on microbial development and survival. Temperature and photochemical processes constantly alter the aerosol properties as well as their chemical composition. The physical and chemical characteristics of the atmosphere that we think constitute the main constraints of microbial life in the dry troposphere are presented below.

UV radiation. UV radiation levels can be extremely high and destructive in the atmosphere. The highly energetic wavelengths (UV-C ~190-290 nm and UV-B ~290-320 nm) are responsible for direct DNA damage that could be lethal. Longer wavelengths (UV-A ~320-400 nm and visible light ~400-800 nm) contribute to intra-cellular reactive oxygen species (ROS)

production that can cause subsequent oxidative damage to DNA, RNA, lipids and proteins, altering microbial metabolism and survival[53,54]. Some microorganisms have developed a range of protection mechanisms against UV radiation. Cell aggregation, association with particles and production of carotenoid pigments to scavenge ROS are all mechanisms used by environmental microorganisms to reduce the effects of destructive UV radiation. The stratosphere supports by far the highest levels of UV radiation found on Earth as levels increase by around 11% with every 1000 m in altitude (WHO). Data on the impact of UV radiation on airborne microorganisms comes mainly from investigations using high UV levels such as those found in the upper troposphere or stratosphere layer. Smith *et al.* (2011)[36] showed that UV radiation was the most biocidal factor in the low stratosphere, and could kill up to 99.9% of *Bacillus subtilis* spores after 96 h. However, the authors pointed out that spore resistance might be dependent on the environment the cells germinated in[55–57]. Consequently, UV resistance might have been higher if the spores were directly isolated from the stratosphere and not germinated in culture media like was done in the study. Microbial strains isolated from the upper troposphere and lower stratosphere exhibited a higher resistance to UV radiation as compared to those from the troposphere at ground level[58]. Some *Deinococcus* and *Streptomyces* strains showed an extreme UV resistance and tended to form aggregates in culture medium. These aggregates were suggested to be a protection mechanism[58]. With the exception of sporulation and cell aggregation, no other protective mechanism against UV radiation has been observed in airborne microbial communities. UV radiation levels and consequently the need for UV protection mechanisms might depend on the height of the troposphere (*i.e.* planetary boundary layer or free troposphere height), geography (for example the tropics harbor higher UV levels) and surface conditioning (*i.e.* surface reflectance)[59].

Temperature shock and freeze-thaw cycles. At the same altitude, atmospheric temperature is highly dependent on the latitude and longitude of the site. It also decreases by 0.6 to 1°C for every 100 m increase of altitude and can reach -70°C in the upper stratosphere. Upward aerial transport of microorganisms with high-speed winds could occur rapidly and airborne microorganisms might suffer large temperature shocks. Airborne microorganisms present in an air parcel transported from the surface to 1 km altitude can undergo a temperature decrease of 5 to 10°C and a substantial increase in relative humidity[4]. Cold temperature and

freeze-thaw cycles generally occur at high latitudes, high altitudes and/or in clouds and so do the resulting impact. They slow down microbial metabolism, decrease membrane fluidity, and influence protein refolding. Freeze-thaw cycles could additionally lead to mechanical stress that might damage the cell membrane[60–62]. Freeze-thaw cycles were shown to alter the survival of microbial strains following UV radiation, H_20_2 exposure and osmotic shock when these factors were tested individually on strains isolated from clouds belonging to *Pseudomonas, Sphingomonas, Arthrobacter* and the yeast *Dioszegia*[8]. To date, no specific mechanism of protection against cold temperature and freeze-thaw cycles known to be used by environmental extremophile microorganisms[63] has been observed in airborne microbial communities.

Relative humidity and condensation/evaporation cycles. The troposphere harbors the whole range of relative humidity (RH) values, from values near 0% in the upper troposphere up to 100% above ground level. Investigations on the survival of aerosolized microorganisms under different RH showed different results depending on the species. While the survival of airborne *Flavobacterium* was not affected by RH ranging from 25 to 99% at 24°C[64], mid-range RH negatively impacted mycoplasma survival but not RH values outside of this range[65]. In the environment, desiccation resistance is generally associated to ionizing radiation resistance[66–70]. Yet, the mutual nature of the underlying mechanisms remains unknown. In the environment, the molecular mechanisms underlying desiccation resistance remain poorly defined and seem to involve wax ester biosynthesis[71] and DNA reparation mechanisms. Desiccation, like radiation, tends to induce DNA damage[70,72].

Evaporation/condensation cycles of water vapor occur in the troposphere, both in the dry troposphere and in clouds. In a water droplet, evaporation can concentrate metabolites in the near environment of the cells by up to 1000 times[8]. Evaporation/condensation cycles induce osmotic changes, leading to water fluxes between the intracellular and extracellular compartment of the cell to maintain osmolarity. These water fluxes can provoke cell damage, increase the concentration of metabolites in cells, and increase the concentration of compounds like radicals and metals around the cell[73,74]. Alsved *et al.* (2018)[75] showed that during evaporation, *Pseudomonas syringae* survival was enhanced when the relative humidity rapidly reached the level where salts become solid. Hence, small and salty liquid droplets were suggested as a more suitable environment when exposed to evaporation than large and

slightly salty liquid droplets[75]. Microbial cells could use compatible solutes that are osmoprotectants to control water fluxes. However, the effect of deleterious evaporation/condensation cycles on for airborne microbial communities and the mechanisms they use to protect themselves are unknown.

Concentration of radicals. The potential impact of the oxidizing nature of the atmosphere that is characterized by an enhanced presence of radicals (OH, O_2^-), nitrate radicals and OH precursors such as hydrogen peroxide (H_2O_2)[4,11] on airborne microorganisms has been mainly investigated in cloud water. Joly *et al.* (2015)[8] tested the effect of different concentrations of hydrogen peroxide on the survival of different microbial strains isolated from cloud water. They showed that the 50% lethal concentration of H_20_2 was different among the strains, and ten times higher than the typical concentration found in Puy-de-Dôme cloud water. Increases in ROS (reactive oxygen species) could occur during other environmental stresses, like UV radiation, as discussed above. They could be deleterious to DNA, RNA, proteins, lipids in cells and can lead to cell death. Anti-oxidant molecules such as vitamins, glutathione, carotenoid pigments and specific enzymes could help deal with an excess of radicals. Yet, the mechanisms involved in the resistance of airborne strains to high concentration of radicals remain unknown[8].

Physical selection versus microbial adaptation of airborne microbial communities

The question as to whether atmospheric chemistry and physics might be controlling factors in leading to the survival and/or development of microbial taxa with specific functions in the atmosphere remains open. On the one hand, the harsh physical and chemical conditions of the troposphere might cause the death of non-resistant cells, a process we consider as physical selection. Resistant cells might survive, and even develop if they are active and growing (discussed in a following section). On the other hand, microbial adaptation (*i.e.* genetic changes in the genome in response to the physical and chemical conditions) might also occur. This would increase adaptation to the tropospheric environment.

In the troposphere, microorganisms in the gas and solid phase face UV radiation, low temperature, low relative humidity and the presence of radicals which can all affect microbial survival or metabolism as discussed above. These conditions become more and more intense as altitude increases and are considered as extreme at the top of the troposphere[36]. UV

radiation levels can be both extremely high (stratosphere) and relatively low (planetary boundary layer) at different heights. While UV radiation might be a critical factor in shaping airborne microbial communities through the selection of resistant microorganisms over a certain height of the troposphere, this remains unknown for the lower troposphere. Survival studies have mainly been done under simulated cloud and stratospheric conditions, and on isolated cultivable microorganisms of an atmospheric origin. While Smith *et al.* (2011)[36] showed that UV radiation was the most biocidal factor in the low stratosphere, Joly *et al.* (2015)[8] suggested that freeze-thaw cycles and osmotic shock were the most damaging factors for microorganisms in clouds when these factors were tested individually on isolated strains.

Survival mechanisms such as dormancy, sporulation, aggregation between cells or with particulate matter, and microbial resistance to the extreme conditions encountered in the atmosphere are relatively common in the environment[8]. Fungal spores have evolved to survive and disseminate through the troposphere. They are known to be particularly resistant to atmospheric conditions, and especially to desiccation, UV radiation and oxidative stress[76]. Yet, their resistance might have been selected for on Earth surfaces before being aerosolized. Some might resist the physical selection but might not adapt while suspended in the air. While resistant microbial cells were observed in the air, the question about whether these resistant cells represent the majority of the airborne microbial cells remains. Little is known about the survival mechanisms of both airborne bacterial and fungal cells, and the ratio between resistant and sensitive cells in the air. We expect a higher abundance of resistant cells as conditions are more intense as altitude increases in the troposphere. Yang *et al.* (2008)[58] showed that microbial strains isolated from the upper troposphere exhibited a higher resistance to UV radiation as compared to strains from the atmosphere at ground level.

Microbial cells resistant to extreme conditions exist in the major sources of airborne microbial cells, (*e.g.,* in soil and water). Survival of airborne cells might be the result of an innate resistance (like fungal spores) or a resistance acquired while aerially transported. Genetic changes in airborne microbial genomes allowing a better survival and/or metabolism (and even development) in the atmosphere might be expected. However, microbial cells might face constantly changing conditions during aerial transport (*i.e.* changes in temperature, UV radiation, condensation/evaporation of water etc.), which could prevent their adaptation. In the ocean, a faster evolution of microorganisms than ocean currents can disperse them has been suggested in the Atlantic and Pacific oceans (oceanic surface current speed around 0.05

m/s;[77,78]). However, air currents could be 100 even 1000 times faster than surface oceanic currents. Inputs of new cells through aerosolization from Earth surfaces are large and continuous in the planetary boundary layer. The free troposphere might receive less cells than the layers close to the ground, and these cells might have initiated a selection process within the planetary boundary layer. We can thus expect to observe the effects of physical selection and microbial adaptation more in the free troposphere as compared to the planetary boundary layer[2].

Physical selection and microbial adaptation in the troposphere, if occurring, might lead to a functional differentiation of airborne microbial communities in response to atmospheric conditions as compared to their source environments. The impact of these processes on the functional potential of airborne microbial communities will be addressed in **Chapter 5** of this thesis.

Metabolic activity and growth of airborne microbial communities

Activity and growth. Airborne microbial cells might be a mix of dead and living cells. The atmosphere harbors carbonaceous sources and inorganic components essential for microbial metabolism, but the stressful conditions (*i.e.* UV radiation, radicals, desiccation, low temperature etc.) might affect the microbial metabolic potential of the living cells. UV radiation in particular has been shown to be a critical factor restraining microbial activity of the oceanic surface bacterioplankton[79–84]. It has been shown that irradiance affected bacterioplankton more in spring and summer[83] and that the activity could be suppressed up to 40% in the top five meters of the water column in near shore waters[80]. Metabolic activity in the atmosphere could be restricted to specific microbial cells resistant to the atmospheric conditions, as well as cells embedded in particulate matter and protected from the potential selective conditions (UV radiation, radicals, desiccation etc.)[85]. The first to date and still one of the rare functional metagenomic studies on airborne microorganisms from the dry troposphere was conducted in New York City and San Diego (USA), and revealed that airborne microorganisms carry a rich panel of putative functional genes[86]. Planetary boundary layer and cloud isolated microbial strains have been shown to metabolize the carbonaceous compounds found in the atmosphere[5,87,88]. rRNA-based studies identified the taxonomy of the potential active microbial taxa in the dry troposphere and cloud water[7,89,90]. Epiphytic, parasitic and endosymbiontes bacterial taxa (*i.e. Sphingomonas, Methylobacterium,*

Acidiphilium, Pseudomonas, Comamonas) have been suggested as the most active organisms due to their physiological properties (resistance to temperature and humidity shifts, high levels of UV radiation etc.) compatible with their maintenance in the dry troposphere and clouds[89,91]. The same was observed for fungi with plant pathogens and saprophytic taxa (Pleosporales, Magnaporthales, Xylariales, Conioscyphales etc.) potentially showing the highest activities[89,91]. In clouds, it has been suggested that bacteria might be more active than fungi based on a transcriptomic study[13].

Airborne microbial growth and reproduction have been suggested in cloud water. Half of the tested strains (11 out of 20) originating from Puy-de-Dôme (France) cloud water have been able to grow at 5°C, which is the average temperature of Puy-de-Dôme clouds[7]. Sattler *et al.* (2001)[92] suggested that bacterial production in cloud water might range from 3.6 to 19.5 days (production measurement at 0°C), which was comparable to those of phytoplankton in the ocean, *i.e.* about a week[93]. Temperature in the planetary boundary layer might be higher than 0°C; consequently, microbial replication time might be less than 4 days. Residence time in the air might be a critical factor for airborne microorganisms to reproduce, as microbial replication time might be of the same order as residence time.

***Role in atmospheric chemistry*.** Airborne microorganisms might transform atmospheric chemical compounds and play a role in biogeochemical cycles[17,87]. In clouds, microorganisms have been shown to use the main carboxylic compounds (monoacid and diacid compounds: formate, acetate, lactate, succinate, formaldehyde and methanol), organic nitrogen as well as the radical precursor H_20_2 after cultivation[5–7,12,88]. Using a liquid medium mimicking the composition of cloud water and a temperature at 5°C (average temperature of low-altitude clouds), biological activity was shown to drive the oxidation of carbonaceous compounds during the night (90 to 99%), while contributing 2 to 37% of the reactivity during the day alongside radical reactions mediated by photochemistry[5]. Studies on the metabolic activity of airborne microbial cells *in situ* present major technical issues. Most of the studies evaluating the metabolic potential of airborne microbial communities are based on cultivable microorganisms, and in this way are studies whose conditions (physics and chemistry) are far from those found in the atmosphere. Given the high taxonomic and functional microbial diversity found in the troposphere, we suppose that airborne microorganisms could have an impact on different biogeochemical cycles. The potentially significant contribution of chemical

transformations mediated by airborne microorganisms might explain the inconsistencies observed in some chemical cycles in the atmosphere, such as the cycle of secondary organic aerosols[94,95]. Reviews on the role of airborne microorganisms in atmospheric chemistry and more generally on the reactivity of bioaerosols in the air can be found in Ariya *et al.* (2002)[87], Delort *et al.* (2010)[14] and Estillore *et al.* (2016)[96].

Distribution factors of airborne microbial communities (geography and time) in the troposphere

Research on the temporal and spatial distribution of microbial communities of the troposphere (in both the liquid and dry phases) based on studies that took place mainly at ground level have revealed that local or regional surfaces[42,97], meteorology (or ground air circulation)[23,98–100], seasons[42,97,100,101] and global air circulation (that is, inputs from distant surfaces)[25,44,100,102,103] are partly responsible for the observed composition of airborne microbial communities. However, there is a lack of understanding of the relative contribution of these driving factors on the composition of the tropospheric microbial community.

Surface characteristics

Airborne microorganisms are emitted mainly by Earth surfaces, *i.e.* from natural (for example forests, oceans, deserts) and urbanized surfaces (agricultural fields, waste water treatment plants, cities). Burrows *et al.* (2009)[104,105] constrained a general atmospheric circulation model using data from the literature and estimated that 10^{24} bacteria are emitted into the atmosphere each year at a global scale. Microscopic and molecular biology analyses showed that bacterial cells are generally in higher concentration compared to fungal cells and spores in the troposphere[2,7,19,44,106,107]. Observations of the microbial diversity in the planetary boundary layer showed that airborne microorganisms from one atmospheric sample might come from many different ecosystems (plants, soil, ocean etc.) that might explain the observed large taxonomic diversity of airborne microbial communities. Aerosolization from Earth surfaces depends mainly on the landscapes (*i.e.* forest, grassland, ocean etc.), as well as the current meteorological conditions. Oceanic surfaces were shown to emit less than terrestrial surfaces[105]. Among terrestrial surfaces, grasslands might be the most effective emitters of microorganisms, while ice potentially emits 100 times less microbial cells[105]. Meteorological conditions would impact the aerosolization mechanism of the surfaces.

Different processes can lead to the uplifting of microbial cells and more generally particulate matter from surfaces to the air. Wind turbulence, blasting and splashing raindrops might also introduce microbial cells into the air. Strong wind speeds during storms can lift dust-associated microorganisms high in the troposphere[108,109]. Over oceanic surfaces specifically, sea spray and bubble-bursting in whitecaps (sea foam crest over the waves) and breaking waves are critical processes in emitting microbial cells in the atmosphere[110–112]. It is not clear in which conditions rain droplets contribute more to aerosolization or washout of microbial cells from the air[113] and seems to be highly dependent on the surface temperature, composition and relative humidity as well as rain intensity. During drier weather, wind direction (vertical direction) and speed might impact aerosolization fluxes. Alongside passive aerosolization mechanisms, active mechanisms used by many fungal species allow for a controlled aerosolization and dissemination of fungal spores[30,114].

Besides being deposited on the ground through precipitation (collision and combination of neighboring water droplets in clouds which create larger and larger water droplets that become too heavy to be suspended and fall), airborne microorganisms also settle back to the Earth's surface by dry deposition. Dry deposition or gravitational settling is size-correlated, with the bigger cells and particulate matter being preferentially deposited on the ground. Dry deposition might depend on meteorology, and specifically wind direction, wind speed and air relative humidity. The modeled residence time of airborne microbial cells defined as a round and free aerosol of 1 µm of diameter was estimated to be 3.4 days on average[104].

Physical selection of airborne microorganisms during aerosolization

Aerosolization might be a critical step that partly controls the observed composition of airborne microbial communities. Aerosolization might mediate the ratio between bacterial and fungal cells observed in the air as well as influence the ratio between the different populations within these kingdoms in acting as a filter. The ratio between airborne bacterial and fungal cells (hyphal fragments and spores) based on qPCR data and the number of 16S and 18S rRNA gene per bacterial and fungal cell, respectively, generally suggests that there are more bacteria in air (example of a ratio of 10 at a high-altitude site and 6.5 at a suburban site, if we consider that bacteria and fungi contain between 10 and 50 rRNA gene copies per cell, respectively[19]). This same ratio is much higher in soil (up to 100 times higher)[115] (and

personal data) and suggests that airborne microbial cells might be the reflection of selective processes that occur during aerosolization.

Small-sized cells, such as bacterial cells that are usually ten times smaller in size than fungal cells, might be preferentially aerosolized. Womack *et al.* (2015)[116] observed that the Amazonian forest air was more loaded in Ascomycota than Basidiomycota fungi. They suggested that Ascomycota fungi might be preferentially aerosolized because of their single-celled and filamentous vegetative growth forms that are much lighter than Basidiomycota spores. Low wind speed might be more effective in lifting light cells and light particulate matter while strong wind speed might also lift macroscopic dust and associated microorganisms. Aerosolization could also depend on the physiological properties of the cell membrane. Specific bacterial taxa (e.g., Actinobacteria and some Gammaproteobacteria) and lipid-enveloped viruses have been proposed to be preferentially aerosolized from oceans through bubble-bursting, as a result of hydrophobic properties of their cell envelope[117].

At a given site, aerosolization specificity might depend on the current meteorology and surface conditioning (soil moisture, presence of plants, season etc.) (see [118] for fungi). On-site studies showed that the ratio between fungal fragments (hyphae fragments) and spores from soil might be of one, and tightly dependent on meteorological conditions (wind speed and direction) and fungal species[119]. Aerosolization of spores from many fungi could also be an active mechanism that depends on season, relative humidity, temperature and/or the time of day (day/night)[116,118].

Only a few studies investigated the selective propriety of aerosolization mechanisms, yet aerosolization might be the first critical process controlling which microorganisms are present in the air[117,120]. The aerosolization of microorganisms from a surface might depend upon the surface conditioning (*i.e.* vegetation, snow cover etc.), meteorological conditions, cell size, cell membrane characteristics as well as the mechanism of aerosolization (active versus passive mechanisms); and the different mechanisms of aerosolization as stated above (wind turbulences, blasting and splashing raindrops, sea spray and bubble-bursting) might be specific each in their own way.

Newly aerosolized microorganisms might then be affected by other processes like death, rapid deposition or growth which will consequently change the ratios between the different microbial species obtained just after aerosolization. Aerosolization might be stressful and

even lethal for microorganisms. Laboratory studies on *E. coli* showed that up to half, and even 99% depending on the aerosolization mechanism and intensity, of the aerosolized bacterial cells suffered from cell membrane damage after 10 min of aerosolization[121], which might lead to subsequent death. Another study showed that the viability of *Pseudomonas fluorescens* bacteria decreased by over 50% after 90 min of continuous aerosolization[122]. The sublethal damages occurring during aerosolization were associated to a differential gene expression of respiratory, cold-shock, metabolism and more generally stress-response genes[121–124]. Alsved *et al.* (2018)[75] showed that a faster drying (with complete removal of liquid water) of small and salty liquid droplets at low relative humidity increased the survival rate of *Pseudomonas syringae.* They suggested that, because these kind of liquid droplets are mainly formed from liquid environments like oceans, aerosolization from oceans might promote the survival of cells as compared to dry environments like soil.

Long-range transport of airborne microorganisms and contribution of local versus distant sources

It is mainly within the planetary boundary layer that gases and particulate matter are emitted, dispersed, transformed and deposited since it in direct interaction with the Earth's surface. Gas and particulate components are continuously exchanged with vegetation, oceans and biological organisms. Short-lived chemical species (for example methyl iodide CH3I: 3.5 days or dimethylsulfide: 1 day) will mainly residence in the planetary boundary layer, while some long-lived substances might eventually pass into the free troposphere and enter global circulation.

Large and visible particles such as sand dust have provided evidence for long-range transport of aerosols (*i.e.* transport over hundreds of kilometers), including bioaerosols. Aerial long-range transport of microorganisms was particularly studied during dust storm events originating from Asian and African deserts[46]. Powerful natural events like dust storms, volcano eruptions and forest fires can lift microorganisms up into the high troposphere and stratosphere, and lead to the aerial transport of microorganisms far away from their source environments. Asian dust storms can take from seven to nine days to cross the Pacific Ocean; African dust storms can take from three to five days to reach the Caribbean and the USA[125]. Interactions between dust and microorganisms of the troposphere during transport remain unknown. Some studies observed a complete change of the tropospheric microbial

community abundance and structure of the downwind sites[3,27,106,126–129]. Dust-associated microbial communities seem to be taxonomically different from the ones of the troposphere (culture-based approach) and in higher concentration (up to ten times higher[27,106,128,129]). Culture-independent studies showed that dust-associated microbial diversity was high and that microorganisms might have both the original desert soil and the troposphere as origins. The capacity of microorganisms to be transported over long distances through the air has raised concern about the role airborne microorganisms might have on both the ecology of downwind ecosystems and public health with the potential dissemination of plant and human pathogens as well as allergens and genetic material[3,46,108,125,130]. Aerial dissemination of human and plant pathogens between regions separated by thousands of kilometers has been suggested in several disease outbreaks[131].

At a given sampling location, airborne microbial communities that are collected might be a mix of microbes originating from local sources and distant sources. Spatial-scale studies showed that the surrounding surface conditioning might be a critical factor[42,97,132]. Other studies, using backward trajectories and microbial bioindicators of specific ecosystems (for example, marine-associated microorganisms), showed the importance of the global air circulation and long-range inputs in the overall composition of the airborne microbial communities at specific sites[25,44,100,102,103].

The relative contribution of the local or regional and distant sources in controlling airborne microbial communities is poorly understood. Their relative contribution might be dependent on the geographical site (altitude from sea level, surface conditioning) and global meteorological characteristics (influence of the free-troposphere). Investigations are technically difficult and may need large-scale investigations of airborne microorganisms (**Chapter 3**).

Meteorology

Of the meteorological factors, wind, temperature and relative humidity have been shown to be partly responsible for troposphere microbial community composition[23,133]. Local meteorology can have an impact on the existing microbial communities by forcing the deposition of airborne cells on the Earth's surface during precipitation events or strong wind. They can also impact aerosolization and thus inputs of microbial cells from the different surfaces, with the change of wind direction for example. Meteorology is responsible for the

transformation of the surface conditioning (*i.e*. vegetation, snow cover etc.) throughout the year and especially over the seasons. Thus, meteorological conditions might be critical in the temporal distribution of microbial communities in the troposphere. A change in the surface conditioning might lead to a change in the Earth's surface microbial communities and thus a change in the diversity of the aerosolized microbial cells. Many studies showed a strong seasonal shift in airborne bacterial populations[42,97,100,132]. The wind speed might be able to act on the distant sources, by increasing the inputs from the distant sources in high-wind conditions. Separating the effect of meteorology from the effect of the surface conditions is difficult, as meteorological conditions correlate with surface conditioning (*i.e.* vegetation, snow cover etc.) over the year.

The relative contribution of the surrounding landscapes, meteorology and inputs from long-range transport (*i.e.* sources distant from hundreds of kilometers) could be different depending on the characteristics of the geographical site, and could change over the year. While the latitudinal and longitudinal position of the site could lead to changes in meteorological activity and diversity of the local sources, the height, latitude and remoteness of the site could have an impact on the influence of the free-troposphere and inputs from distant sources. Using a worldwide-scale approach, part of this thesis addressed geographic and temporal distribution of airborne microbial communities and identified the main controlling factors (**Chapter 3** and **4**). Better understanding the controlling factors of airborne microbial communities is critical for anticipating the effect of global warming and land use changes in the distribution of microbial communities of the troposphere.

PhD objectives, approach and working hypotheses

The objective of my PhD was to increase our understanding of some of the major processes in the microbial ecology of the dry troposphere. Especially, we wanted to provide elements of a response to questions such as "*What drives airborne microbial community composition?*" and "*Can we predict airborne microbial composition?*".

A limited number of studies investigated the structuring factors of microbial communities in the dry troposphere using a multidisciplinary approach (*e.g.,* investigation of meteorological conditions[23,25,100,132] and/or atmospheric chemistry[23,25,100,134] and/or the landscapes[42,132] in addition to airborne microbial communities). Moreover, the majority of these investigations were carried out on individual or few sampling sites[23,25,42,100,132]. However, the use of a

multidisciplinary approach works better if associated with a global-scale approach in order to get a large range of the microbial, chemical, meteorological and/or landscape conditions using the same methodology. To our knowledge, no currently published studies used this global-scale approach.

Our multidisciplinary and global-scale approach is presented in **Chapter 2**. Both airborne bacterial and fungal communities were investigated at different meteorological stations around the globe over two months or more to gain insight on the temporal variability of the microbial composition at each site. Sites were chosen based on their different latitudinal positions, altitude from sea level, local meteorology and surrounding landscapes (marine, terrestrial, polar, urban etc.).

Our main interest was to investigate the distribution of troposphere microbial communities and identify the main sources and factors responsible for the observed composition of airborne microbial communities both in terms of taxonomy and functional potential using this global-scale approach (**Chapter 2**). By improving knowledge on the distribution of airborne microbial communities, we wanted to study the relative importance of the local and distant sources of airborne microorganisms and the factors that structure airborne microbial communities (**Chapter 3**), what drives the seasonal shift in airborne microbial community composition (**Chapter 4**) and the potential functional selection and/or adaptation of airborne microbial communities in response to atmospheric abiotic conditions (**Chapter 5**).

We suggest that local sources and local environmental factors like local meteorology would have a greater contribution to the distribution of microbial communities of the dry troposphere of planetary boundary layer than distant sources. Thus, we expect that the surrounding landscapes (marine, terrestrial, polar etc.) would drive both the taxonomic structure and functional potential of planetary boundary layer microbial communities at a global scale. We suggest that airborne microbial communities are a subsample of the surface microbial communities that underwent a selective process during both aerosolization and aerial transport. We suggest that microorganisms harboring specific physical characteristic (*i.e.* spore, membrane characteristics, aggregation of cells etc.) and/or genetic and enzymatic microbial properties (*i.e.* efficient DNA repair mechanisms etc.) might be more likely to survive the adverse environmental conditions (UV radiation, desiccation, etc.) of the troposphere. The set of our assumptions are summarized in **Figure 5**.

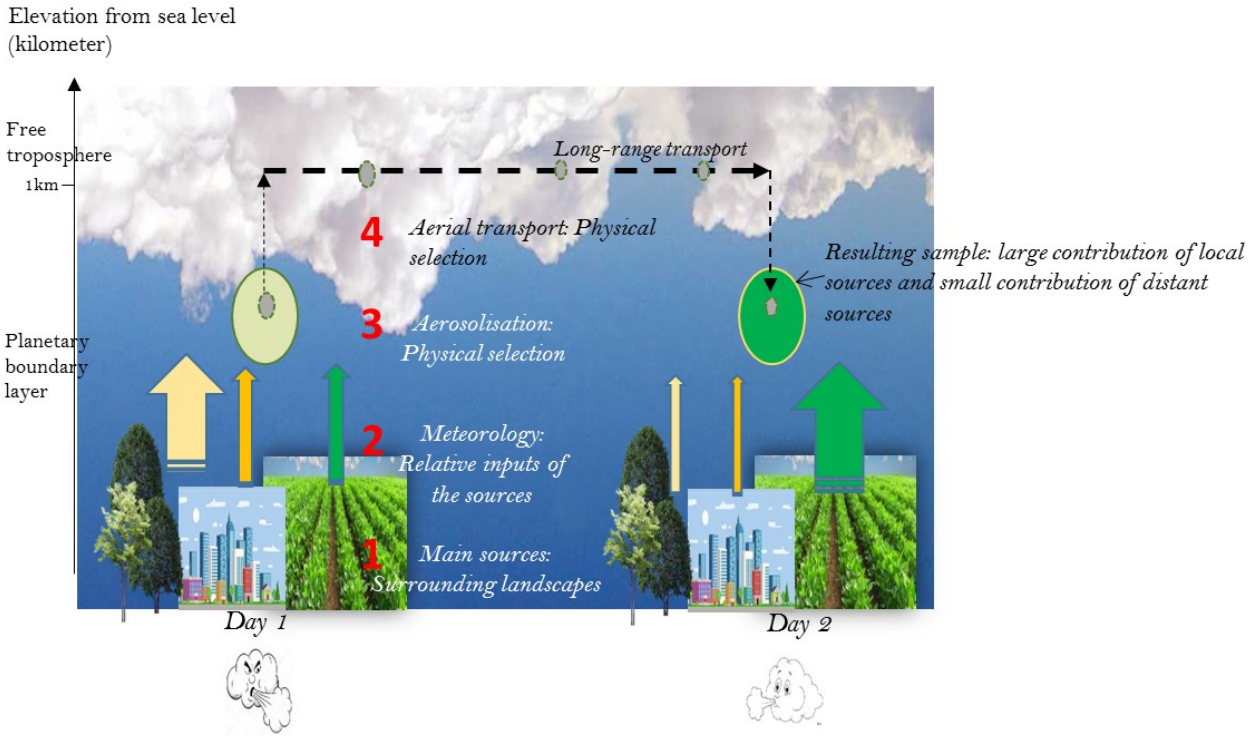


Figure 5: Sources and processes driving the composition of airborne microbial communities based on the PhD's hypotheses, in relative order of expected importance ("1" is the most important to "4" the less important).

References

1. Triadó-Margarit, X., Caliz, J., Reche, I. & Casamayor, E. O. High similarity in bacterial bioaerosol compositions between the free troposphere and atmospheric depositions collected at high-elevation mountains. *Atmos. Environ.* **203**, 79–86 (2019).
2. Els, N., Baumann-Stanzer, K., Larose, C., Vogel, T. M. & Sattler, B. Beyond the planetary boundary layer: Bacterial and fungal vertical biogeography at Mount Sonnblick, Austria. *Geo Geogr. Environ.* **6**, e00069 (2019).
3. Katra, I. *et al.* Richness and Diversity in Dust Stormborne Biomes at the Southeast Mediterranean. *Sci. Rep.* **4**, 5265 (2014).
4. Seinfeld, J. H. & Pandis, S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, 3rd Edition | Environmental Chemistry | Chemistry | Subjects | Wiley*. (Wiley-Interscience, 1998).
5. Vaïtilingom, M. *et al.* Contribution of Microbial Activity to Carbon Chemistry in Clouds. *Appl. Environ. Microbiol.* **76**, 23–29 (2010).
6. Vaïtilingom, M. *et al.* Potential impact of microbial activity on the oxidant capacity and organic carbon budget in clouds. *Proc. Natl. Acad. Sci.* **110**, 559–564 (2013).
7. Amato, P. *et al.* A fate for organic acids, formaldehyde and methanol in cloud water: their biotransformation by micro-organisms. *Atmospheric Chem. Phys.* **7**, 4159–4169 (2007).
8. Joly, M. *et al.* Survival of microbial isolates from clouds toward simulated atmospheric stress factors. *Atmos. Environ.* **117**, 92–98 (2015).
9. Pruppacher, H. R. & Jaenicke, R. The processing of water vapor and aerosols by atmospheric clouds, a global estimate. *Atmospheric Res.* **38**, 283–295 (1995).

10. Marinoni, A. *et al.* Hydrogen peroxide in natural cloud water: Sources and photoreactivity. *Atmospheric Res.* **101**, 256–263 (2011).
11. Deguillaume, L. *et al.* Classification of clouds sampled at the puy de Dôme (France) based on 10 yr of monitoring of their physicochemical properties. *Atmospheric Chem. Phys.* **14**, 1485–1506 (2014).
12. Hill, K. A. *et al.* Processing of atmospheric nitrogen by clouds above a forest environment. *J. Geophys. Res. Atmospheres* **112**, (2007).
13. Amato, P. *et al.* Metatranscriptomic exploration of microbial functioning in clouds. *Sci. Rep.* **9**, 1–12 (2019).
14. Delort, A.-M. *et al.* A short overview of the microbial population in clouds: Potential roles in atmospheric chemistry and nucleation processes. *Atmospheric Res.* **98**, 249–260 (2010).
15. Failor, K. C., Schmale, D. G., Vinatzer, B. A. & Monteil, C. L. Ice nucleation active bacteria in precipitation are genetically diverse and nucleate ice by employing different mechanisms. *ISME J.* **11**, 2740–2753 (2017).
16. Ariya, P. *et al.* Physical and chemical characterization of bioaerosols--Implications for nucleation processes. *Int. Rev. Phys. Chem.* **28**, 1–32 (2009).
17. Ariya, P. A. & Amyot, M. New Directions: The role of bioaerosols in atmospheric chemistry and physics. *Atmos. Environ.* **38**, 1231–1232 (2004).
18. Haga, D. I. *et al.* Ice nucleation by fungal spores from the classes *Agaricomycetes*, *Ustilaginomycetes*, and *Eurotiomycetes*, and the effect on the atmospheric transport of these spores. *Atmospheric Chem. Phys.* **14**, 8611–8630 (2014).
19. Tanaka, D. *et al.* Airborne Microbial Communities at High-Altitude and Suburban Sites in Toyama, Japan Suggest a New Perspective for Bioprospecting. *Front. Bioeng. Biotechnol.* **7**, (2019).
20. Zweifel, U. L. *et al.* High bacterial 16S rRNA gene diversity above the atmospheric boundary layer. *Aerobiologia* **28**, 481–498 (2012).
21. DeLeon-Rodriguez, N. Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. (2013). Available at: http://www.pnas.org/content/110/7/2575.full. (Accessed: 25th July 2017)
22. Bertolini, V. *et al.* Temporal variability and effect of environmental variables on airborne bacterial communities in an urban area of Northern Italy. *Appl. Microbiol. Biotechnol.* **97**, 6561–6570 (2013).
23. Zhen, Q. *et al.* Meteorological factors had more impact on airborne bacterial communities than air pollutants. *Sci. Total Environ.* **601–602**, 703–712 (2017).
24. Genitsaris, S. *et al.* Variability of airborne bacteria in an urban Mediterranean area (Thessaloniki, Greece). *Atmos. Environ.* **157**, 101–110 (2017).
25. Gandolfi, I. *et al.* 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 (2015).
26. Cho, B. C. & Hwang, C. Y. Prokaryotic abundance and 16S rRNA gene sequences detected in marine aerosols on the East Sea (Korea). *FEMS Microbiol. Ecol.* **76**, 327–341 (2011).
27. Park, J., Li, P.-F., Ichijo, T., Nasu, M. & Yamaguchi, N. Effects of Asian dust events on atmospheric bacterial communities at different distances downwind of the source region. *J. Environ. Sci.* **72**, 133–139 (2018).

28. Després, V. *et al.* Primary biological aerosol particles in the atmosphere: a review. *Tellus B Chem. Phys. Meteorol.* **64**, 15598 (2012).
29. Huffman, J. A. *et al.* Size distributions and temporal variations of biological aerosol particles in the Amazon rainforest characterized by microscopy and real-time UV-APS fluorescence techniques during AMAZE-08. *Atmospheric Chem. Phys.* **12**, 11997–12019 (2012).
30. Sesartic, A., Lohmann, U. & Storelvmo, T. Bacteria in the ECHAM5-HAM global climate model. *Atmospheric Chem. Phys.* **12**, 8645–8661 (2012).
31. Graham, B. *et al.* Composition and diurnal variability of the natural Amazonian aerosol. *J. Geophys. Res. Atmospheres* **108**, (2003).
32. Pöschl, U. Atmospheric aerosols: composition, transformation, climate and health effects. *Angew. Chem. Int. Ed Engl.* **44**, 7520–7540 (2005).
33. Smith, D. J. *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.* **9**, 1752 (2018).
34. Imshenetsky, A. A., Lysenko, S. V. & Kazakov, G. A. Upper boundary of the biosphere. *Appl. Environ. Microbiol.* **35**, 1–5 (1978).
35. DasSarma, P. & DasSarma, S. Survival of microbes in Earth's stratosphere. *Curr. Opin. Microbiol.* **43**, 24–30 (2018).
36. Smith, D. J., Griffin, D. W., McPeters, R. D., Ward, P. D. & Schuerger, A. C. Microbial survival in the stratosphere and implications for global dispersal. *Aerobiologia* **27**, 319–332 (2011).
37. Fuzzi, S. *et al.* Particulate matter, air quality and climate: lessons learned and future needs. *Atmospheric Chem. Phys.* **15**, 8217–8299 (2015).
38. Clauss, M. Particle size distribution of airborne micro-organisms in the environment-A review. *Landbauforsch. Volkenrode* **65**, 77–100 (2015).
39. Amato, P. *et al.* Survival and ice nucleation activity of bacteria as aerosols in a cloud simulation chamber. *Atmospheric Chem. Phys.* **15**, 6455–6465 (2015).
40. Sippula, O. *et al.* Characterization of Chemical and Microbial Species from Size-Segregated Indoor and Outdoor Particulate Samples. *Aerosol Air Qual. Res.* **13**, 1212–1230 (2013).
41. Fulton, J. D. Microorganisms of the Upper Atmosphere: III. Relationship between Altitude and Micropopulation. *Appl. Microbiol.* **14**, 237–240 (1966).
42. Bowers, R. M., McCubbin, I. B., Hallar, A. G. & Fierer, N. Seasonal variability in airborne bacterial communities at a high-elevation site. *Atmos. Environ.* **50**, 41–49 (2012).
43. Bowers, R. M. *et al.* Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. *Environ. Sci. Technol.* **47**, 12097–12106 (2013).
44. Mayol, E. *et al.* Long-range transport of airborne microbes over the global tropical and subtropical ocean. *Nat. Commun.* **8**, 201 (2017).
45. Jaenicke, R. Atmospheric aerosols and global climate. *J. Aerosol Sci.* **11**, 577–588 (1980).
46. Griffin, D. W., Gonzalez-Martin, C., Hoose, C. & Smith, D. J. Global-Scale Atmospheric Dispersion of Microorganisms. in *Microbiology of Aerosols* 155–194 (John Wiley & Sons, Ltd, 2017). doi:10.1002/9781119132318.ch2c
47. Zhang, H. *et al.* Stress resistance, motility and biofilm formation mediated by a 25kb plasmid pLMSZ08 in Listeria monocytogenes. *Food Control* **94**, 345–352 (2018).

48. Ma, Y., Zeng, Y., Jiao, N., Shi, Y. & Hong, N. Vertical distribution and phylogenetic composition of bacteria in the Eastern Tropical North Pacific Ocean. *Microbiol. Res.* **164**, 624–633 (2009).
49. Ghiglione, J.-F. *et al.* Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17633–17638 (2012).
50. Shi, Y., Tyson, G. W., Eppley, J. M. & DeLong, E. F. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J.* **5**, 999–1013 (2011).
51. DeLong, E. F. *et al.* Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**, 496–503 (2006).
52. Yu, Z., Yang, J., Liu, L., Zhang, W. & Amalfitano, S. Bacterioplankton community shifts associated with epipelagic and mesopelagic waters in the Southern Ocean. *Sci. Rep.* **5**, 12897 (2015).
53. Moan, J. & Peak, M. J. Effects of UV radiation of cells. *J. Photochem. Photobiol. B* **4**, 21–34 (1989).
54. Santos, A. L. *et al.* Wavelength dependence of biological damage induced by UV radiation on bacteria. *Arch. Microbiol.* **195**, 63–74 (2013).
55. Rangel, D. E. N. *et al.* Stress tolerance and virulence of insect-pathogenic fungi are determined by environmental conditions during conidial formation. *Curr. Genet.* **61**, 383–404 (2015).
56. Rangel, D. E. N., Anderson, A. J. & Roberts, D. W. Evaluating physical and nutritional stress during mycelial growth as inducers of tolerance to heat and UV-B radiation in Metarhizium anisopliae conidia. *Mycol. Res.* **112**, 1362–1372 (2008).
57. Hagiwara, D. *et al.* Temperature during conidiation affects stress tolerance, pigmentation, and trypacidin accumulation in the conidia of the airborne pathogen Aspergillus fumigatus. *PLoS ONE* **12**, (2017).
58. Yang, Y., Yokobori, S. & Yamagishi, A. UV-resistant bacteria isolated from upper troposphere and lower stratosphere. *Biol.Sci.Space* **22**, (2008).
59. Vorob'eva, L. I. Stressors, Stress Reactions, and Survival of Bacteria: A Review. *Appl. Biochem. Microbiol.* **40**, 217–224 (2004).
60. D'Amico, S., Collins, T., Marx, J.-C., Feller, G. & Gerday, C. Psychrophilic microorganisms: challenges for life. *EMBO Rep.* **7**, 385–389 (2006).
61. Nedwell, D. B. Effect of low temperature on microbial growth: lowered affinity for substrates limits growth at low temperature. *FEMS Microbiol. Ecol.* **30**, 101–111 (1999).
62. Berry, E. D. & Foegeding, P. M. Cold Temperature Adaptation and Growth of Microorganisms. *J. Food Prot.* **60**, 1583–1594 (1997).
63. Satyanarayana, T., Raghukumar, C. & Sisinthy, S. Extremophilic microbes: Diversity and perspectives. *Curr. Sci.* **89**, 78–90 (2005).
64. Ehrlich, R., Miller, S. & Walker, R. L. Effects of Atmospheric Humidity and Temperature on the Survival of Airborne Flavobacterium. *Appl. Microbiol.* **20**, 884–887 (1970).
65. Wright, D. N., Bailey, G. D. & Hatch, M. T. Survival of airborne Mycoplasma as affected by relative humidity. *J. Bacteriol.* **95**, 251–252 (1968).
66. Fredrickson, J. K. *et al.* Protein oxidation: key to bacterial desiccation resistance? *ISME J.* **2**, 393–403 (2008).
67. Fredrickson, J. K. *et al.* Geomicrobiology of High-Level Nuclear Waste-Contaminated Vadose Sediments at the Hanford Site, Washington State. *Appl. Environ. Microbiol.* **70**, 4230–4241 (2004).

68. Sanders, S. W. & Maxcy, R. B. Isolation of radiation-resistant bacteria without exposure to irradiation. *Appl. Environ. Microbiol.* (1979).
69. Rainey, F. A. *et al.* Extensive Diversity of Ionizing-Radiation-Resistant Bacteria Recovered from Sonoran Desert Soil and Description of Nine New Species of the Genus Deinococcus Obtained from a Single Soil Sample. *Appl. Environ. Microbiol.* **71**, 5225–5235 (2005).
70. Mattimore, V. & Battista, J. R. Radioresistance of Deinococcus radiodurans: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. *J. Bacteriol.* **178**, 633–637 (1996).
71. Finkelstein, D., Brassell, S. & Pratt, L. Microbial biosynthesis of wax esters during desiccation: Adaptation for colonization of the earliest terrestrial environments? *Geology* **38**, 247–250 (2010).
72. Dose, K., Bieger-Dose, A., Labusch, M. & Gill, M. Survival in extreme dryness and DNA-single-strand breaks. *Adv. Space Res.* **12**, 221–229 (1992).
73. Sleator, R. D. & Hill, C. Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiol. Rev.* **26**, 49–71 (2002).
74. Wood, J. M. Bacterial responses to osmotic challenges. *J. Gen. Physiol.* **145**, 381–388 (2015).
75. Alsved, M. *et al.* Effect of Aerosolization and Drying on the Viability of Pseudomonas syringae Cells. *Front. Microbiol.* **9**, 3086 (2018).
76. Dijksterhuis, J. & Samson, R. A. *Food Mycology: A Multifaceted Approach to Fungi and Food*. (CRC Press, 2007).
77. Martiny, A. C., Tai, A. P. K., Veneziano, D., Primeau, F. & Chisholm, S. W. Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ. Microbiol.* **11**, 823–832 (2009).
78. Hellweger, F. L., van Sebille, E. & Fredrick, N. D. Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. *Science* **345**, 1346–1349 (2014).
79. Hernández, K. L., Quiñones, R. A., Daneri, G., Farias, M. E. & Helbling, E. W. Solar UV radiation modulates daily production and DNA damage of marine bacterioplankton from a productive upwelling zone (36°S), Chile. *J. Exp. Mar. Biol. Ecol.* **343**, 82–95 (2007).
80. Herndl, G. J., Müller-Niklas, G. & Frick, J. Major role of ultraviolet-B in controlling bacterioplankton growth in the surface layer of the ocean. *Nature* **361**, 717–719 (1993).
81. Winter, C., Moeseneder, M. M. & Herndl, G. J. Impact of UV Radiation on Bacterioplankton Community Composition. *Appl. Environ. Microbiol.* **67**, 665–672 (2001).
82. Alonso-Sáez, L., Gasol, J. M., Lefort, T., Hofer, J. & Sommaruga, R. Effect of natural sunlight on bacterial activity and differential sensitivity of natural bacterioplankton groups in northwestern Mediterranean coastal waters. *Appl. Environ. Microbiol.* **72**, 5806–5813 (2006).
83. Ruiz-González, C. *et al.* Seasonal patterns in the sunlight sensitivity of bacterioplankton from Mediterranean surface coastal waters. *FEMS Microbiol. Ecol.* **79**, 661–674 (2012).
84. Ruiz Gonzalez, C., Simó, R., Sommaruga, R. & Gasol, J. M. Away from darkness: a review on the effects of solar radiation on heterotrophic bacterioplankton activity. *Front. Microbiol.* **4**, (2013).
85. Smith, D. J., Griffin, D. W. & Jaffe, D. A. The high life: Transport of microbes in the atmosphere. *Eos Trans. Am. Geophys. Union* **92**, 249–250 (2011).

86. Yooseph, S. *et al.* A Metagenomic Framework for the Study of Airborne Microbial Communities. *PLOS ONE* **8**, e81862 (2013).
87. Ariya, P. A., Nepotchatykh, O., Ignatova, O. & Amyot, M. Microbiological degradation of atmospheric organic compounds. *Geophys. Res. Lett.* **29**, 34-1-34–4 (2002).
88. Amato, P. *et al.* Microbial population in cloud water at the Puy de Dôme: Implications for the chemistry of clouds. *Atmos. Environ.* **39**, 4143–4153 (2005).
89. Klein, A. M., Bohannan, B. J. M., Jaffe, D. A., Levin, D. A. & Green, J. L. Molecular Evidence for Metabolically Active Bacteria in the Atmosphere. *Front. Microbiol.* **7**, (2016).
90. Womack, A. M. UV-resistant bacteria isolated from upper troposphere and lower stratosphere (PDF Download Available). *ResearchGate* (2010). doi:http://dx.doi.org/10.2187/bss.22.18
91. Amato, P. *et al.* Active microorganisms thrive among extremely diverse communities in cloud water. *PLOS ONE* **12**, e0182869 (2017).
92. Sattler, B., Puxbaum, H. & Psenner, R. Bacterial growth in supercooled cloud droplets. *Geophys. Res. Lett.* **28**, 239–242 (2001).
93. Falkowski, P., Barber, R. & Smetacek, V. Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science* **281**, 200–7 (1998).
94. Yang, W., Li, J., Wang, M., Sun, Y. & Wang, Z. A Case Study of Investigating Secondary Organic Aerosol Formation Pathways in Beijing using an Observation-based SOA Box Model. *Aerosol Air Qual. Res.* **18**, (2018).
95. Khan, M. a. H. *et al.* A modeling study of secondary organic aerosol formation from sesquiterpenes using the STOCHEM global chemistry and transport model. *J. Geophys. Res. Atmospheres* **122**, 4426–4439 (2017).
96. Estillore, A. D., Trueblood, J. V. & Grassian, V. H. Atmospheric chemistry of bioaerosols: heterogeneous and multiphase reactions with atmospheric oxidants and other trace gases. *Chem. Sci.* **7**, 6604–6616 (2016).
97. Bowers, R. M., McLetchie, S., Knight, R. & Fierer, N. 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 (2011).
98. Šantl-Temkiv, T., Gosewinkel, U., Starnawski, P., Lever, M. & Finster, K. Aeolian dispersal of bacteria in southwest Greenland: their sources, abundance, diversity and physiological states. *FEMS Microbiol. Ecol.* **94**, (2018).
99. Barberán, A. *et al.* Continental-scale distributions of dust-associated bacteria and fungi. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 5756–5761 (2015).
100. Innocente, E. *et al.* Influence of seasonality, air mass origin and particulate matter chemical composition on airborne bacterial community structure in the Po Valley, Italy. *Sci. Total Environ.* **593–594**, 677–687 (2017).
101. Yamamoto, N. *et al.* Particle-size distributions and seasonal diversity of allergenic and pathogenic fungi in outdoor air. *ISME J.* **6**, 1801–1811 (2012).
102. Cáliz, J., Triadó-Margarit, X., Camarero, L. & Casamayor, E. O. A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 12229–12234 (2018).
103. Fröhlich-Nowoisky, J. *et al.* Biogeography in the air: fungal diversity over land and oceans. *Biogeosciences* **9**, 1125–1136 (2012).

104. Burrows, S. M., Elbert, W., Lawrence, M. G. & Pöschl, U. Bacteria in the global atmosphere – Part 1: Review and synthesis of literature data for different ecosystems. *Atmos Chem Phys* **9**, 9263–9280 (2009).
105. Burrows, S. M. *et al.* Bacteria in the global atmosphere – Part 2: Modeling of emissions and transport between different ecosystems. *Atmospheric Chem. Phys.* **9**, 9281–9297 (2009).
106. Li, Y., Lu, R., Li, W., Xie, Z. & Song, Y. Concentrations and size distributions of viable bioaerosols under various weather conditions in a typical semi-arid city of Northwest China. *J. Aerosol Sci.* **106**, 83–92 (2017).
107. Alghamdi, M. A. *et al.* Microorganisms associated particulate matter: A preliminary study. *Sci. Total Environ.* **479–480**, 109–116 (2014).
108. Kellogg, C. A. & Griffin, D. W. Aerobiology and the global transport of desert dust. *Trends Ecol. Evol.* **21**, 638–644 (2006).
109. Smith, D. J. *et al.* Intercontinental Dispersal of Bacteria and Archaea by Transpacific Winds. *Appl. Environ. Microbiol.* **79**, 1134–1139 (2013).
110. Mayol, E., Jiménez, M. A., Herndl, G. J., Duarte, C. M. & Arrieta, J. M. Resolving the abundance and air-sea fluxes of airborne microorganisms in the North Atlantic Ocean. *Front. Microbiol.* **5**, (2014).
111. Aller, J. Y., Kuznetsova, M. R., Jahns, C. J. & Kemp, P. F. The sea surface microlayer as a source of viral and bacterial enrichment in marine aerosols. *J. Aerosol Sci.* **36**, 801–812 (2005).
112. Blanchard, D. C. & Syzdek, L. D. Water-to-Air Transfer and Enrichment of Bacteria in Drops from Bursting Bubbles. *Appl. Environ. Microbiol.* **43**, 1001–1005 (1982).
113. Joung, Y. S., Ge, Z. & Buie, C. R. Bioaerosol generation by raindrops on soil. *Nat. Commun.* **8**, 14668 (2017).
114. Elbert, W., Taylor, P. E., Andreae, M. O. & Pöschl, U. Contribution of fungi to primary biogenic aerosols in the atmosphere: wet and dry discharged spores, carbohydrates, and inorganic ions. *Atmospheric Chem. Phys.* **7**, 4569–4588 (2007).
115. Malik, A. A. *et al.* Soil Fungal:Bacterial Ratios Are Linked to Altered Carbon Cycling. *Front. Microbiol.* **7**, (2016).
116. Womack, A. M. *et al.* Characterization of active and total fungal communities in the atmosphere over the Amazon rainforest. *Biogeosciences* **12**, 6337–6349 (2015).
117. Michaud, J. M. *et al.* Taxon-specific aerosolization of bacteria and viruses in an experimental ocean-atmosphere mesocosm. *Nat. Commun.* **9**, 2017 (2018).
118. Crandall, S. G. & Gilbert, G. S. Meteorological factors associated with abundance of airborne fungal spores over natural vegetation. *Atmos. Environ.* **162**, 87–99 (2017).
119. Górny, R. & Lawniczek-Walczyk, A. Effect of two aerosolization methods on the release of fungal propagules from a contaminated agar surface. *Ann. Agric. Environ. Med. AAEM* **19**, 279–84 (2012).
120. Fahlgren, C. *et al.* Seawater mesocosm experiments in the Arctic uncover differential transfer of marine bacteria to aerosols. *Environ. Microbiol. Rep.* **7**, 460–470 (2015).
121. Thomas, R. J. *et al.* The Cell Membrane as a Major Site of Damage during Aerosolization of Escherichia coli. *Appl. Environ. Microbiol.* **77**, 920–925 (2011).
122. Zhen, H., Han, T., Fennell, D. E. & Mainelis, G. Release of Free DNA by Membrane-Impaired Bacterial Aerosols Due to Aerosolization and Air Sampling. *Appl. Environ. Microbiol.* **79**, 7780–7789 (2013).

123. Ng, T. W., Chan, W. L. & Lai, K. M. Importance of stress-response genes to the survival of airborne Escherichia coli under different levels of relative humidity. *AMB Express* **7**, (2017).
124. Ng, T. W. *et al.* Differential gene expression in Escherichia coli during aerosolization from liquid suspension. *Appl. Microbiol. Biotechnol.* **102**, 6257–6267 (2018).
125. Griffin, D. W. Atmospheric Movement of Microorganisms in Clouds of Desert Dust and Implications for Human Health. *Clin. Microbiol. Rev.* **20**, 459–477 (2007).
126. Weil, T. *et al.* Legal immigrants: invasion of alien microbial communities during winter occurring desert dust storms. *Microbiome* **5**, (2017).
127. Yamaguchi, N. *et al.* Abundance and Community Structure of Bacteria on Asian Dust Particles Collected in Beijing, China, during the Asian Dust Season. *Biol. Pharm. Bull.* **39**, 68–77 (2016).
128. Dong, L. *et al.* Concentration and size distribution of total airborne microbes in hazy and foggy weather. *Sci. Total Environ.* **541**, 1011–1018 (2016).
129. Griffin, D. W. *et al.* Atmospheric microbiology in the northern Caribbean during African dust events. *Aerobiologia* **19**, 143–157 (2003).
130. Aylor, D. E. Spread of Plant Disease on a Continental Scale: Role of Aerial Dispersal of Pathogens. *Ecology* **84**, 1989–1997 (2003).
131. Brown, J. K. M. & Hovmøller, M. S. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* **297**, 537–541 (2002).
132. Uetake, J. *et al.* Seasonal changes of airborne bacterial communities over Tokyo and influence of local meteorology. *bioRxiv* 542001 (2019). doi:10.1101/542001
133. Zhai, Y. *et al.* A review on airborne microorganisms in particulate matters: Composition, characteristics and influence factors. *Environ. Int.* **113**, 74–90 (2018).
134. Liu, H. *et al.* The distribution variance of airborne microorganisms in urban and rural environments. *Environ. Pollut.* **247**, 898–906 (2019).

Chapter 1: Bibliography - Airborne microbial communities of the troposphere

Supplementary Information

Table S1: List of investigations on the airborne microbial community taxonomic structure using high through put sequencing (not cloning) to date to the best of our knowledge. The type of samples, sampling strategy and sequencing technology are presented. We tried to be as exhaustive as possible.

Studies using amplicon sequencing (high throughput, not cloning)				
Study	**Year**	**Type of samples (atmospheric niche) and location**	**Sampling strategy**	**Sequencing technology**
Triado-Margarit et al., 2019[1]	2019	troposphere PM* Sierra Nevada, USA	passive automatic sampler	454
Liu et al., 2019[2]	2019	troposphere PM Hangzhou, China	glass fiber filter, impactor	Ion
Maki et al., 2019[3]	2019	troposphere PM over deserts, China	polycarbonate filter, air pump	MiSeq
Uetake et al., 2019[4]	2019	troposphere PM Tokyo, Japan	quartz filter, impactor	MiSeq
Evans et al., 2019[5]	2019	fog Namibia and Maine	passive fog collector	454
Els et al., 2019[6]	2019	troposphere PM Mount Sonnblick, Austria	liquid impingement	MiSeq
Archer et al., 2019[7]	2019	troposphere PM Antarctica	liquid impingement	MiSeq
Tanaka et al., 2019[8]	2019	troposphere PM Japan	vacuum pump, polycarbonate filter	MiSeq
Jiaxian et al., 2019[9]	2019	cloud water Southern China	cloud water collector	MiSeq
Cáliz et al., 2018[10]	2018	rain and snow Central Pyrenees, Spain	wet and dry passive sampler	Miseq
Park et al., 2018[11]	2018	dust Beijing, China	glass fiber filter, impactor	Ion
Zhu et al., 2018[12]	2018	cloud water Mount Tai, China	cloud water collector	MiSeq
Yan et al., 2018[13]	2018	troposphere PM Beijing, China	impingement	MiSeq
Smith et al., 2018[14]	2018	troposphere PM high troposphere, USA	aircraft bioaerosol collector, gelatinous filter membrane	MiSeq
Du et al., 2018[15]	2018	troposphere PM Beijing, China	quartz filter, impactor	MiSeq
Woo et al., 2018[16]	2018	troposphere PM Seoul, South Korea	anderson sampler, glass fiber substrate	MiSeq
Fang et al., 2018[17]	2018	troposphere PM urban ecosystems Hangzhou, China	liquid impingement	MiSeq
Amato et al., 2017[18]	2017	cloud water Puy-de-Dôme, France	cloud water sampler	MiSeq
Zhen et al., 2017[19]	2017	troposphere PM Beijing, China	glass fiber filter, impactor	HiSeq
Maki et al., 2017[20]	2017	dust Japan	vacuum pump, polycarbonate filter	MiSeq
Gat et al., 2017[21]	2017	dust Eastern Mediterranean	quartz filter, impactor	MiSeq
Innocente et al., 2017[22]	2017	troposphere PM Milan and Venice, Italy	quartz filter, impactor	HiSeq
Genitsaris et al., 2017[23]	2017	troposphere PM Urban Mediterranean area	vacuum filtration, liquid medium	454
Mayol et al., 2017[24]	2017	troposphere PM over oceans	liquid impingement	454
Cuthbertson et al., 2017[25]	2017	troposphere PM arctic	different techniques	MiSeq
Mhuireach et al., 2016[26]	2016	troposphere PM Oregon, USA	vacuum pump, cellulose ester filter	MiSeq
Gandolfi et al., 2015[27]	2015	troposphere PM Milan and Venice, Italy	quartz filter, impactor	MiSeq
Barberan et al., 2015[28]	2015	dust on USA houses	dust collection	Miseq and HiSeq
Be et al., 2015[29]	2015	troposphere PM urban areas, USA	sampling unit filter	GAIIx and HiSeq
Seifried et al., 2015[30]	2015	troposphere PM over ocean	liquid impingement	454
Womack et al., 2015[31]	2015	troposphere PM Amazon forest	liquid impingement	MiSeq
Maki et al., 2015[32]	2015	dust Noto Peninsula, Japan	vacuum pump, polycarbonate filter	454
Barberan et al., 2014[33]	2014	dust Pyrenees, Spain	passive collector GF/F filter	454
Dannemiller et al., 2014[34]	2014	troposphere PM Israel	quartz filter, impactor	454
Katra et al., 2014[35]	2014	dust Israel	dust collector, quartz filters	454

DeLeon-Rodriguez et al., 2013[36]	2013	troposphere PM	Mediterranean Sea	vacuum pump, cellulose nitrate membrane	454
Bertolini et al., 2013[37]	2013	troposphere PM	Milan, Italy	quartz filter, impactor	Illumina GA-IIx
Bowers et al., 2013[38]	2013	troposphere PM	Denver and Greeley, USA	quartz filter, impactor	454
Bowers et al., 2012[39]	2012	troposphere PM	Storm-Peak, USA	vacuum pump, cellulose nitrate filter	454
Franzetti et al., 2011[40]	2011	troposphere PM	Milan, Italy	quartz filter, impactor	454

*PM=particulate matter

Table S2: List of investigations whose datasets have been used to do the multivariate analysis PCoA in Figure 2.

Name on the PCoA	Study	Atmospheric niche and location	Sampling strategy	Sequencing technology
fog_1 to 3	Evans *et al.*, 2019[5]	fog Namibia and Maine	passive fog collector	454
FT_Austria_1 to 5	Els *et al.*, 2019[6]	PM Mount Sonnblick, Austria	liquid impingement	MiSeq
rain_Pyrenees_1 to 4	Cáliz *et al.*, 2018[10]	rain and snow Central Pyrenees, Spain	wet and dry passive sampler	Miseq
sand_dust_1 to 3	Park *et al.*, 2018[11]	dust Beijing, China	glass fiber filter, impactor	Ion
cloud	Zhu *et al.*, 2018[12]	cloud water Mount Tai, China	cloud water collector	MiSeq
BL_Beijing	Yan *et al.*, 2018[13]	PM Beijing, China	impingement	MiSeq
FT or ascent_Smith_1 to 2	Smith *et al.*, 2018[14]	PM high troposphere, USA	aircraft bioaerosol collector, gelatinous filter membrane	MiSeq
cloud_water_1 to 3	Amato *et al.*, 2017[18]	cloud water Puy-de-Dôme, France	cloud water sampler	MiSeq
BL_Beijing_1 to 2	Zhen *et al.*, 2017[19]	PM Beijing, China	glass fiber filter, impactor	HiSeq
dust_1 to 3	Gat *et al.*, 2017[21]	dust (PM!) Eastern Mediterranean	quartz filter, impactor	MiSeq
BL_Milan_1 to 3	Gandolfi *et al.*, 2015[27]	PM Milan and Venice, Italy	quartz filter, impactor	MiSeq
upp_troposphere_marine_1 to 3	DeLeon-Rodriguez *et al.*, 2013[36]	PM Mediterranean Sea	vacuum pump, cellulose nitrate membrane	454

References

1. Triadó-Margarit, X., Caliz, J., Reche, I. & Casamayor, E. O. High similarity in bacterial bioaerosol compositions between the free troposphere and atmospheric depositions collected at high-elevation mountains. *Atmospheric Environment* **203**, 79–86 (2019).
2. Liu, H. *et al.* The distribution variance of airborne microorganisms in urban and rural environments. *Environmental Pollution* **247**, 898–906 (2019).
3. Maki, T. *et al.* Aeolian Dispersal of Bacteria Associated With Desert Dust and Anthropogenic Particles Over Continental and Oceanic Surfaces. *Journal of Geophysical Research: Atmospheres* **124**, 5579–5588 (2019).
4. Uetake, J. *et al.* Seasonal changes of airborne bacterial communities over Tokyo and influence of local meteorology. *bioRxiv* 542001 (2019). doi:10.1101/542001
5. Evans, S. E., Dueker, M. E., Logan, J. R. & Weathers, K. C. The biology of fog: results from coastal Maine and Namib Desert reveal common drivers of fog microbial composition. *Science of The Total Environment* **647**, 1547–1556 (2019).
6. Els, N., Baumann-Stanzer, K., Larose, C., Vogel, T. M. & Sattler, B. Beyond the planetary boundary layer: Bacterial and fungal vertical biogeography at Mount Sonnblick, Austria. *Geo: Geography and Environment* **6**, e00069 (2019).
7. Archer, S. D. J. *et al.* Microbial dispersal limitation to isolated soil habitats in the McMurdo Dry Valleys of Antarctica. *bioRxiv* 493411 (2018). doi:10.1101/493411
8. Tanaka, D. *et al.* Airborne Microbial Communities at High-Altitude and Suburban Sites in Toyama, Japan Suggest a New Perspective for Bioprospecting. *Front. Bioeng. Biotechnol.* **7**, (2019).
9. Jiaxian, P. *et al.* Diversity of bacteria in cloud water collected at a National Atmospheric Monitoring Station in Southern China. *Atmospheric Research* **218**, 176–182 (2019).

10. Cáliz, J., Triadó-Margarit, X., Camarero, L. & Casamayor, E. O. A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12229–12234 (2018).
11. Park, J., Li, P.-F., Ichijo, T., Nasu, M. & Yamaguchi, N. Effects of Asian dust events on atmospheric bacterial communities at different distances downwind of the source region. *Journal of Environmental Sciences* **72**, 133–139 (2018).
12. Zhu, C. *et al.* Chemical Composition and Bacterial Community in Size-Resolved Cloud Water at the Summit of Mt. Tai, China. *Aerosol and Air Quality Research* **18**, (2017).
13. Yan, D. *et al.* Structural Variation in the Bacterial Community Associated with Airborne Particulate Matter in Beijing, China, during Hazy and Nonhazy Days. *Appl. Environ. Microbiol.* **84**, (2018).
14. Smith, D. J. *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* **9**, 1752 (2018).
15. Du, P., Du, R., Lu, Z., Ren, W. & Fu, P. Variation of Bacterial and Fungal Community Structures in PM2.5 Collected during the 2014 APEC Summit Periods. *Aerosol and Air Quality Research* **18**, (2017).
16. Woo, C., An, C., Xu, S., Yi, S.-M. & Yamamoto, N. Taxonomic diversity of fungi deposited from the atmosphere. *ISME J* **12**, 2051–2060 (2018).
17. Fang, Z., Guo, W., Zhang, J. & Lou, X. Influence of Heat Events on the Composition of Airborne Bacterial Communities in Urban Ecosystems. *Int J Environ Res Public Health* **15**, (2018).
18. Amato, P. *et al.* Active microorganisms thrive among extremely diverse communities in cloud water. *PLOS ONE* **12**, e0182869 (2017).
19. Zhen, Q. *et al.* Meteorological factors had more impact on airborne bacterial communities than air pollutants. *Sci. Total Environ.* **601–602**, 703–712 (2017).
20. Maki, T. *et al.* Variations in airborne bacterial communities at high altitudes over the Noto Peninsula (Japan) in response to Asian dust events. *Atmospheric Chemistry and Physics* **17**, 11877–11897 (2017).
21. Gat, D., Mazar, Y., Cytryn, E. & Rudich, Y. Origin-Dependent Variations in the Atmospheric Microbiome Community in Eastern Mediterranean Dust Storms. *Environ. Sci. Technol.* **51**, 6709–6718 (2017).
22. Innocente, E. *et al.* Influence of seasonality, air mass origin and particulate matter chemical composition on airborne bacterial community structure in the Po Valley, Italy. *Sci. Total Environ.* **593–594**, 677–687 (2017).
23. Genitsaris, S. *et al.* Variability of airborne bacteria in an urban Mediterranean area (Thessaloniki, Greece). *Atmospheric Environment* **157**, 101–110 (2017).
24. Mayol, E. *et al.* Long-range transport of airborne microbes over the global tropical and subtropical ocean. *Nature Communications* **8**, 201 (2017).
25. Cuthbertson, L. *et al.* Characterisation of Arctic Bacterial Communities in the Air above Svalbard. *Biology (Basel)* **6**, (2017).
26. Mhuireach, G. Á., Betancourt-Román, C. M., Green, J. L. & Johnson, B. R. Spatiotemporal Controls on the Urban Aerobiome. *Front. Ecol. Evol.* **7**, (2019).
27. Gandolfi, I. *et al.* 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 (2015).

28. Barberán, A. *et al.* Continental-scale distributions of dust-associated bacteria and fungi. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5756–5761 (2015).
29. Be, N. A. *et al.* Metagenomic analysis of the airborne environment in urban spaces. *Microb. Ecol.* **69**, 346–355 (2015).
30. Seifried, J. S., Wichels, A. & Gerdts, G. Spatial distribution of marine airborne bacterial communities. *Microbiologyopen* **4**, 475–490 (2015).
31. Womack, A. M. *et al.* Characterization of active and total fungal communities in the atmosphere over the Amazon rainforest. *Biogeosciences* **12**, 6337–6349 (2015).
32. Maki, T. *et al.* Vertical distribution of airborne bacterial communities in an Asian-dust downwind area, Noto Peninsula. *Atmospheric Environment* **119**, 282–293 (2015).
33. Barberán, A., Henley, J., Fierer, N. & Casamayor, E. O. Structure, inter-annual recurrence, and global-scale connectivity of airborne microbial communities. *Science of The Total Environment* **487**, 187–195 (2014).
34. Dannemiller, K. C., Lang-Yona, N., Yamamoto, N., Rudich, Y. & Peccia, J. Combining real-time PCR and next-generation DNA sequencing to provide quantitative comparisons of fungal aerosol populations. *Atmospheric Environment* **84**, 113–121 (2014).
35. Katra, I. *et al.* Richness and Diversity in Dust Stormborne Biomes at the Southeast Mediterranean. *Scientific Reports* **4**, 5265 (2014).
36. DeLeon-Rodriguez, N. Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. (2013). Available at: http://www.pnas.org/content/110/7/2575.full. (Accessed: 25th July 2017)
37. Bertolini, V. *et al.* Temporal variability and effect of environmental variables on airborne bacterial communities in an urban area of Northern Italy. *Appl. Microbiol. Biotechnol.* **97**, 6561–6570 (2013).
38. Bowers, R. M. *et al.* Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. *Environ. Sci. Technol.* **47**, 12097–12106 (2013).
39. Bowers, R. M., McCubbin, I. B., Hallar, A. G. & Fierer, N. Seasonal variability in airborne bacterial communities at a high-elevation site. *Atmospheric Environment* **50**, 41–49 (2012).
40. Franzetti, A., Gandolfi, I., Gaspari, E., Ambrosini, R. & Bestetti, G. Seasonal variability of bacteria in fine and coarse urban air particulate matter. *Appl. Microbiol. Biotechnol.* **90**, 745–753 (2011).

Chapter 2: Methods to investigate the global atmospheric microbiome

Section 1: Protocol optimization and quality control

Published in Frontiers in Microbiology

Dommergue A[1], Amato P[3], Tignat-Perrier R[1,2], Magand O[1], Thollot A[1,2], Joly Muriel[3], Bouvier Laetitia[4], Sellegri K.[4], Vogel T[2], Sonke JE[5], Jaffrezo J-L[1], Marcos Andrade[6], Isabel Moreno[6], Labuschagne C, Martin L, Zhang Q[8] and Larose C[2]

[1]Institut des Géosciences de l'Environnement, Univ. Grenoble Alpes/Centre National de la Recherche Scientifique/Institut de Recherche pour le Développement, G-INP, Grenoble, France

[2]CNRS UMR 5005, Environmental Microbial Genomics, Laboratoire Ampère, École Centrale de Lyon, Université de Lyon, France

[3]Institut de Chimie de Clermont-Ferrand (ICCF), UMR6096 CNRS--Université Clermont Auvergne-Sigma, 63000 Clermont-Ferrand, France

[4]Laboratory for Meteorological Physics (LaMP), Université Clermont Auvergne, 63000 Clermont-Ferrand, France

[5]Géosciences Environnement Toulouse, Centre National de la Recherche Scientifique/Institut de Recherche pour le Développement/Université de Toulouse, Toulouse, France

[6]Laboratorio de Fisica de la Atmosfera, Instituto de Investigaciones Fısicas, Universidad Mayor de San Andres, La Paz, Bolivia

[7]South African Weather Service (SAWS), Stellenbosch, South Africa

[8]Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences (CAS), Beijing 100101, China

Abstract

The interplay between microbes and atmospheric physical and chemical conditions is an open field of research that can only be fully addressed using multidisciplinary approaches. The lack of coordinated efforts to gather data at representative temporal and spatial scales limits aerobiology to help understand large scale patterns of global microbial biodiversity and its causal relationships with the environmental context. This paper presents the sampling strategy and analytical protocols developed in order to integrate different fields of research such as microbiology, –omics biology, atmospheric chemistry, physics and meteorology to characterize atmospheric microbial life. These include control of chemical and microbial contaminations from sampling to analysis and identification of experimental procedures for characterizing airborne microbial biodiversity and its functioning from the atmospheric samples collected at remote sites from low cell density environments. We used high-volume sampling strategy to address both chemical and microbial composition of the atmosphere, because it can help overcome low aerosol and microbial cell concentrations. To account for contaminations, exposed and unexposed control filters were processed along with the samples. We present a method that allows for the extraction of chemical and biological data from the same quartz filters. We tested different sampling times, extraction kits and methods to optimize DNA yield from filters. Based on our results, we recommend supplementary sterilization steps to reduce filter contamination induced by handling and transport. These include manipulation under laminar flow hoods and UV sterilization. In terms of DNA extraction, we recommend a vortex step and a heating step to reduce binding to the quartz fibers of the filters. These steps have led to a 10-fold increase in DNA yield, allowing for downstream omics analysis of air samples. Based on our results, our method can be integrated into pre-existing long-term monitoring field protocols for the atmosphere both in terms of atmospheric chemistry and biology. We recommend using standardized air volumes and to develop standard operating protocols for field users to better control the operational quality.

Introduction

Biological particles are known to represent a significant fraction (~ 20-70%) of the total number of aerosols > 0.2 µm, with large spatial and temporal variations[1–4]. Among these, microorganisms are of particular interest in fields as diverse as epidemiology, including

phytopathology[5,6], bioterrorism, forensic science and public health[7], and environmental sciences, like microbial ecology[8–10], meteorology and climatology[11,12]. More precisely concerning the latter, airborne microorganisms contribute to the pool of particles nucleating the condensation and crystallization of water and they are thus potentially involved in cloud formation and in the triggering of precipitation[13,14]. Additionally, viable microbial cells act as chemical catalyzers interfering with atmospheric chemistry[15]. The constant flux of bacteria from the atmosphere to the Earth's surface due to precipitation and dry deposition can also affect global biodiversity, but they are rarely taken into account when conducting ecological surveys[16–19]. As stressed by these studies attempting to decipher and understand the spread of microbes over the planet[13,20,21], concerted data are needed for documenting the abundance and distribution of airborne microorganisms, including at remote and altitudes sites.

Airborne bacteria are emitted by most Earth surfaces (plants, oceans, land and urban areas) to the atmosphere via a variety of mechanical processes such as aeolian soil erosion, sea spray production, or mechanical disturbances including anthropogenic activities[9,22]. Due to their relatively small size (the median aerodynamic diameter of bacteria-containing particles is around 2-4 µm[23], these can then be transported upward by turbulent fluxes[24] and carried by wind to long distances. As a consequence, bacteria are present in the air up to at least the lower stratosphere[25–27]. Given that the atmosphere is a large conveyor belt that moves air over thousands of kilometers, microorganisms are disseminated globally[28–30]. Airborne transport of microbes is therefore likely pervasive at the global scale, yet there have been only a limited number of studies that have looked at the spatial distribution of microbes across different geographical regions[30,31]. One of the main difficulties is linked with the low microbial biomass associated with a high diversity existing in the atmosphere outdoor (~ 10^2-10^5 cells/m^3 [21,32,33]), thus requiring reliable sampling procedures and controls. Furthermore, the site location and its environmental specificities have to be accounted for to some extent by considering chemical and meteorological variables[34].

While these studies have led to novel findings regarding the link that may exist between airborne bacteria and their source and receptacle environments, the lack of uniform sampling and analysis methodology weaken the conclusions that can be drawn from independent studies. Here we aimed to provide sample collection and preparation methods intended to generate reproducible data, applicable to most sampling locations. This should allow the

investigation of long-range transport, surface ecosystem interconnectivity and distribution of microorganisms in relation to meteorological and chemical contexts. Here, we present a method that allows simultaneous -sampling for chemical and microbiological characterization of aerosols, which can be deployed for long term monitoring at atmospheric observation sites throughout the planet. This has only been carried out previously in urban areas and the methods were not developed or optimized for non-urban environments[35] or for subsequent chemical analysis[36,37]. The main objectives were to: **1)** define appropriate sampling methods and duration; **2)** set up quality controls in order to improve the detection limit for various chemical species; **3)** improve DNA extraction methods from low biomass samples; **4)** ensure data intercomparability and **5)** develop simplified experimental workflows that can easily be carried out by non-specialist onsite technical staff. These protocols were tested at 10 distinct sites covering various geographic regions of the globe. The need for a coordinated network for global monitoring of aerobiology has recently been identified in a number of recently published studies[28,38,39], however, a standardized sampling method has yet to be proposed. Based on our results, our method can be integrated into pre-existing long-term monitoring field protocols for the atmosphere both in terms of atmospheric chemistry and biology, and could be included in future projects.

Material and Methods

Experimental strategy

A variety of methodologies for bioaerosol sampling, including passive sampling, filtration and impaction techniques exist[40], however these have yet to be harmonized for concerted studies. In order to deal with the equipment available at most international monitoring stations, a sampling protocol that could be carried out by non-specialized personnel using on-site sampling equipment needed to be designed; these conditions constrained the choice of our bioaerosol sampling strategy toward high-volume samplers (high air flow-rate) on large diameter/size quartz fiber filters. Sampling time as well as DNA extraction protocols were improved for ensuring obtaining sufficient biological material for analyses from these types of filters, while maintaining the sampling time as short as possible for allowing detecting variations in connection with environmental variables. The sampling strategy and protocol development are outlined in the following section.

Filter selection and development of sterilization protocols

The analysis for chemical compounds and elements (*e.g.*, elemental carbon (EC) and organic carbon (OC)) requires the combustion of a quartz fiber filter, which constrained the choice of sampling material. Depending on the sampler model, two filters sizes were used (5.9" round filter and 8"x 10" rectangular types) with filtration surface areas of 163 and 526 cm^2, respectively. Several sterilization methods were tested to improve the biological quality of the filter without altering the detection limits of the chemical parameters. A standard method for in atmospheric chemistry protocols is to dry heat the quartz filters at 500°C for at least a few hours[41]. We tested additional heating and sterilization steps in laminar flow hoods or UV-exposure (10 minutes) on both filters and storage material. By adding these supplementary steps, we were able to significantly reduce the background OC concentrations of our blank filters as compared to the standard method (standard method OC concentration=0.93 ± 0.35 µg/cm² of filter, new method OC concentration=0.55 ± 0.26 µg/cm², p = 0.0009, Student T test).

Based on our results, the following protocol was defined: filters were heated to 500°C for 8 hours in order to remove traces of organic carbon including DNA. Filters were then handled within a laminar flow hood (UV sterilized, 254 nm, PSM – ESI FLUFRANCE BIOCYT 120, 10 minutes on each filter side) and individually stored in a folded aluminum foil and a thermally-sealed plastic (PE) bag or a zip-lock bag. All the material including foils, plastic bags, tweezers that would be in contact with the filters was UV-sterilized (2 J/cm2 for 2 minutes, 254 nm, CrossLinker, Bio-Link BLX). The filter holders were also UV-sterilized and stored individually in sterile bags. At the sampling sites, the field operators were trained and were given a detailed protocol (see SI for detailed protocol) in order to replace and handle the filters properly at defined sampling times. After collection, filters were sealed in a folded sterile aluminum foil and plastic bags and stored at -20°C. At most sites, no microbiological safety cabinet was available; thus clean benches were made using pre-UV-sterilized plastic sheets in order to minimize contamination. All the sterile material was provided in sufficient quantity to the field operators. After sampling, filters (exposed and controls) were shipped to France for analysis from each sampling site at below zero temperature.

Optimization of DNA extraction

While quartz filters have been used for microbial studies in the air[36,42,43], limitations exist regarding the integrity of the samples collected (DNA degradation, cell mortality[37]). In addition, classic DNA extraction methods need to be improved for investigations in remote sites with low biomass. Since we were constrained by the choice of quartz filters for chemical analysis, we tested different extraction protocols in order to optimize DNA extraction yield from these filters. We tested different DNA extraction kits developed for environmental samples, (*e.g.*, DNeasy PowerWater, DNeasy PowerSoil and DNeasy Blood & Tissue kits from Qiagen). To do so, we set up a size selective high volume air sampling instrument (DIGITEL) equipped with a PM10 size-selective inlet in order to collect airborne particulate matter smaller than 10 µm (cut-off aerodynamic diameter) from the roof of the laboratory in Grenoble (France) to mimic field conditions. An atmospheric sample was collected for 24 hours on a large filter. Nine sub-samples were collected from this filter and DNA was extracted according to the manufacturer's instructions with the following modification: after the lysis step, the lysate was centrifuged in a syringe for four minutes at 1000 rcf to drain filter debris that tended to absorb the lysis solution. This additional centrifugation step increased lysate volume recovery by more than five-fold, potentially increasing DNA recovery. DNA concentrations were compared following quantification with a fluorometric method (Qubit dsDNA HS Assay Kit from Thermo Fisher Scientific, manufacturer's instructions on 10 µL of sample) and 16S rRNA gene copy numbers were compared following quantification using qPCR. Briefly, the V3 region of the 16S rRNA gene was amplified using the SensiFast SYBR No-Rox kit (Bioline) and the following primers sequences: Eub 338f 5'-ACTCCTACGGGAGGCAGCAG-3' as the forward primer and Eub 518r 5'-ATTACCGCGGCTGCTGG-3' as the reverse primer[44] on a Rotorgene 3000 machine (Qiagen). The reaction mixture of 20 µL contained 10 µL of SYBR master mix, 2µL of DNA and RNAse-free water to complete the final 20 µL volume. The 2-step qPCR program consisted of an initial step at 95°C for 2min for enzyme activation, then 35 cycles of 5 sec at 95°C and 20 sec at 60°C for hybridization and elongation, respectively. A final step was added to obtain a denaturation from 55°C to 95°C with increments of 1°C. The amplicon length was around 200 bp. PCR products obtained from DNA from a pure culture of *E.coli* were cloned in a plasmid (pCR™2.1-

TOPO® vector, Invitrogen) and used as standard after quantification with the Broad-Range Qubit Fluorometric Quantification (Thermo Fisher Scientific).

The DNA concentrations measured were not significantly different among the three DNA extraction kits (Soil 0.008 ± 0.005 ng/µL, Water 0.010 ± 0.005 ng/µL, Tissue 0.012 ± 0.002 ng/µL, P>0.22, One-way Anova, Tukey tests). However, the number of 16S rRNA gene copies per cubic meter of air was on average ten times higher with DNeasy PowerWater and PowerSoil kits than it was with DNeasy Blood & Tissue kit. Differences in extraction efficiency have been previously observed for several metagenomic and taxonomic studies and it is suggested that a variety of methods be tested to optimize results before studying new environments[45,46]. Based on these results and for practical reasons, we chose to use the DNeasy PowerWater kit. The DNeasy PowerBead tubes of the DNeasy PowerWater kit are 5 mL tubes (compared to the 2 mL PowerBead tubes of the DNeasy PowerSoil kit) allowing extraction of a larger filter surface.

The second series of extraction tests were carried out to determine whether the filter handling protocol had an effect on extraction efficiency. We tested 4 different treatments in triplicate: blank filters (processed and brought to the field but unexposed), filters that were passively exposed to the atmosphere for 5 minutes, filters that collected atmospheric samples during 24 hours without heat treatment and filters that collected atmospheric samples during 24 hours with heat treatment. DNA was extracted from filters using the protocol outlined above and quantified using Qubit. Based on the results of our test, heating the quartz filters at 500°C for 8 hours before sampling has a significant impact on DNA extraction efficiency and reduced yield by up to 10-fold (24 hours no heating 0.33 ± 0.16 ng/µL, n=3, 24 hours heating 0.038 ± 0.005 ng/µL, n=3, P<0.05, Student t-test), but this is a critical step to ensure that trace carbon is removed from the filters prior to sampling. Therefore, we needed to further optimize our method to increase DNA extraction efficiency from heat sterilized quartz filters. Different options were tested, such as modifying the pH by addition of 1 M CaCO3 and shaking, as described in Bertolini *et al.* (2013)[35], as well as adding a sonication step[37], Neither of these methods increased yield significantly and we preferred to avoid adding a solution to our samples. Considering that high temperatures help desorbing DNA from the silica in the quartz filter[47], we also tested the efficiency of a 1 hour thermal treatment at 65°C during the lysis step[36,37] on DNA yield using eleven samples. We obtained four to five times higher DNA

concentrations and 100 to 1000 times higher 16s rRNA gene copies per cubic meter of air using this optimized lysis technique. The final protocol is summarized in **Figure 1** and was applied to all the filters collected during the sampling campaign.

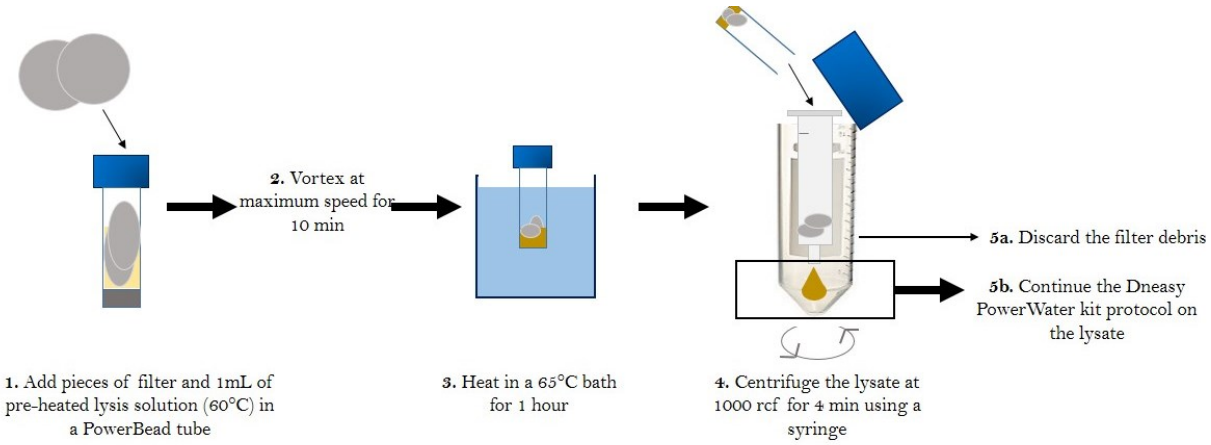


Figure 1: Summary of the modified DNA extraction protocol developed for quartz filters.

Optimization of sampling time for DNA analysis

Once the filter treatment and DNA extraction protocols were validated, we carried out different tests to optimize sample collection duration. We used the same sampling set-up on the roof of the laboratory in Grenoble to collect atmosphere samples. Several different sampling times were considered: 5 minutes, 1 hour, 24 hours and 72 hours. DNA was extracted from filters using the protocol outlined above and quantified using Qubit. Filters analyzed in triplicate at 5 minutes and 1 hour where found to be below the detection limit of the apparatus (0.01 ng/µL). Filters collected at 24 hours showed a significantly lower yield than those collected at 72 hours (24 hours=0.02 ± 0.01 ng/µL, 72 hours=0.08 ± 0.03 ng/µL, n=3, $P<0.05$, student t-test), which suggests that sampling duration impacts DNA yield. Based on the results of this test and considering the remoteness of the sites, we decided to sample continuously for 7 days.

Protocol testing and deployment

Ten sites were chosen based on latitudinal positions, known chemical characteristics, historic atmospheric data and logistical support. The characteristics for each site are described in **Table 1** and the geographic distribution can be seen in **Figure 2**. Sites included Arctic and Antarctic stations as well as mid-latitude stations. In order to access information on long range transport of aerosols, dusts and airborne microorganisms, three sites that are frequently in the free troposphere were selected. Temporal variability is an important but poorly understood factor of microbial community diversity[35], so we completed the dataset with continuous sampling for more than one entire year at a 1-week resolution at a single reference site: Puy-de-Dôme, France. To avoid snap-shot sampling, each site was sampled for a minimum of 2 months, so manned research sites were selected. At all sampling sites, meteorological parameters, such as wind speed and direction, rainfall, temperature, humidity, air pressure, and solar radiation were systematically recorded. Depending on the site, continuous measurements providing additional information for data interpretation were also collected. These included aerosol properties (size, concentration) and gas concentrations (ambient O3, nitrogen species, CO, CO2, CH4, H2O, O2, and volatile organic compounds (VOCs). Finally, total gaseous mercury (TGM) concentration was also continuously monitored by our collaborators at Cape Point Station, Nam Co, Amsterdam Island and Villum Research Station.

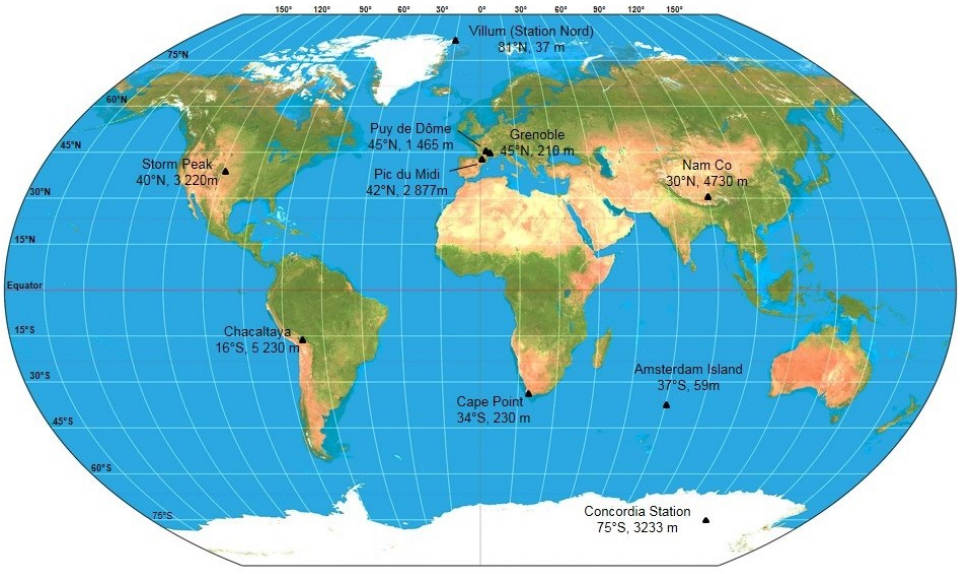


Figure 2: Global distribution of the sampling sites and their respective elevation above sea level.

Table 1: Summary of sampling sites characteristics. Column 4 describes the high volume sampler as well as the quartz fiber filter size and type. Column 8 gives the number of filters (samples + blanks) collected at each sampling site. Volume are expressed at standard ambient temperature and pressure (SATP).

Sampling site code	Sampling site location	Coordinates and altitude	Site characteristics	Sampler model	Collection period (dd/mm/yyyy)		Sampling duration	Mean air volume per filter (SATP approx.)	Number of samples
					Start	Stop			
AMS	Amsterdam Island, France	37°47'82"S 77°33'04''E 59 m asl	remote and natural background of marine environments[48]	Customma de PM10 5.9" ∅ size	07/09/2016	10/11/2016	1 week	5000 m3	9
CAP	Cape Point Station, South Africa	34°21'26"S 18°29'51''E 230 m asl	South Atlantic Ocean, greater Cape Town and other continental sources[49]	Digital DA77 PM10 5.9" ∅ size	11/10/2016	05/12/2016	1 week	4600 m3	7
CHC	Chacaltaya, Bolivia	16°20'47"S 68°07'44'' W 5380 m asl	tropical free troposphere local urban pollution[50,51]	Customma de PM10 5.9" ∅ size	27/06/2016	11/11/2016	1 week *	2000 m3	16
DMC	Concordia Station, Antarctica	75°06'00"S 123°19'58'' E 3233 m asl	Cold environment, polar boundary layer[52]	TISCH TSP 8"x10" size	19/12/2014	31/01/2015	2 weeks	16000m3	3
GRE	Grenoble, France	45°11'38"N 05°45'44''E 210 m asl	Urban air, European air masses	Digital DA77 PM10 5.9" ∅ size	30/06/2017	14/09/2017	1 week	4700 m3	10
NCO	NamCo, China	30°46'44"N 90°59'31''E 4730 m asl	Tibetan Plateau, no major anthropogenic sources, cold and dry conditions, intense solar radiation[53]	Chinese HV PM10 8"x10" size	16/05/2017	28/07/2017	1 week	5300 m3	9
PDD	Puy-de-Dôme, France	45°46'20"N 02°57'57''E 1465 m asl	Urban, oceanic, continental air masses, free troposphere at times[54–57]	Customma de PM10 5.9" ∅ size	23/06/2016	23/08/2017	1-2 week(s)	10000 m3	53
PDM	Pic du Midi, France	42°56'11"N 00°08'34''E 2876 m asl	Mountain top, North Atlantic Ocean, European continent and anthropogenic air masses[58–60]	TISCH PM10 8"x10" size	20/06/2016	04/10/2016	1 week	8000 m3	13
STN	Villum research Station, Station Nord, Greenland	81°34'24"N 16°38'24''E 37 m asl	emission from sea-ice, form the Arctic Ocean and from long-range transport from northern Eurasia[61]	Digital DA77 PM10 5.9" ∅ size	20/03/2017	29/06/2017	1 week	5200 m3	13
STP	Storm Peak Laboratory, USA	40°27'18"N 106°44'38'' E 3220 m asl	Free troposphere, westerly winds, urban pollution during the day[62,63]	TISCH PM10 8"x10" size	11/07/2017	04/09/2017	1 week	5700 m3	7

*Nighttime only

Several brands of size selective high volume air sampling instruments (TISCH, DIGITEL, home-made) were used in the present study. All these samplers are based on the same physical principle. Air was drawn into the sampler and through a large quartz fiber filter by means of a powerful pump, so that particulate material impacted the filter surface. All the samplers, but

one (at Concordia station in Antarctica -DMC), were equipped with a PM10 size-selective inlet in order to collect particulate matter smaller than 10 µm (cut-off aerodynamic diameter). The use of the PM10 inlet was an important aspect in order to guarantee that both chemical and microbial data could be compared among the sites and to prevent rain and hydrometeors from reaching the filter and modifying its porosity and air flow rate properties. At DMC, total suspended particles were collected (median aerodynamic diameter of 20 µm approx). Sampling air flow rates were between 30 and 70 m^3/h (± 2%) and collected volumes ranged from 2000 m^3 to 10000 m^3 over a one-week period (except at DMC, where the samples were collected over a two-week period). Flow regulation of the pump was controlled with a flow-meter that was regularly checked and calibrated. The total volume of air was subsequently corrected to standard ambient temperature and pressure (SATP, 298K, 101.325 kPa) to standardize air collection at all the sites (**Table 1**). For some mountain sites (PDM, CHC, STP), we generally sampled at nights in order to limit the sampling of the planetary boundary layer air.

Quality control

In addition to the 140 samples, we also collected 38 blank filters named "transportation blanks" (TB, 18 filters) and "field blanks" (FB, 20 filters) in order to monitor and check the quality of the sampling protocol (**Fig. 3**). The transportation blanks were filters shipped back and forth to the sampling sites but without any manipulation. The field blanks were exposed for 24 to 72 hours without switching on the high-volume sampler and then processed and stored similarly as the actual samples. The field blanks had significantly higher OC concentrations as compared to the transportation blanks, but these were in the same range as the values obtained for blank filters using the standard sterilization technique (without subsequent sterilization steps).

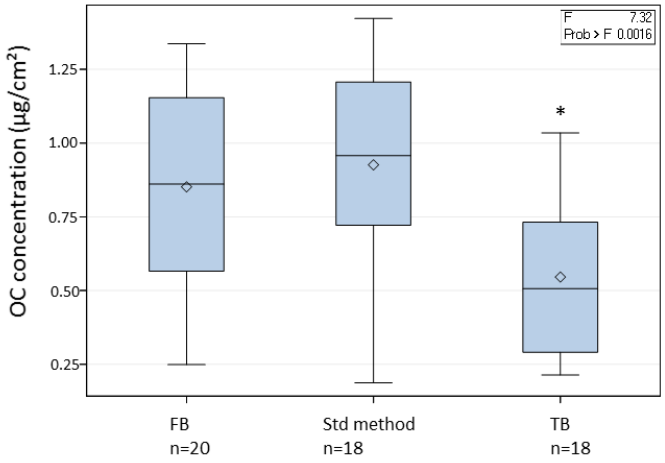


Figure 3: Organic concentrations on different types of filters: field blanks (FB), transport blanks (TB) and filters treated using the standard method (Std method). Significance was tested using One-way Anova and Tukey tests.

Chemical analyses

Elemental carbon (EC), organic carbon (OC), sugar anhydrides and alcohols, major soluble anions and cations as described by Waked *et al.* (2014)[64] as well as total mercury were systematically analyzed in all the quartz fiber filters (including blank filters). EC and OC were analyzed from a 1.5 cm2 punch sample using a thermo optical transmission method on a Sunset Lab analyzer[65]. Punches of 38 mm-diameter from each samples were extracted using ultrapure water under mechanical agitation for a period of 30 minutes. The extracts were then filtered with 0.22 µm Nucleopore filters before injection in the instruments[66]. The extracts were used for quantification of sugar anhydrides and alcohols (levoglucosan, mannosan, galactosan, inositol, glycerol, erythriol, xylitol, arabitol, sorbitol, mannitol, trehalose, rhamnose, glucose, fructose and sucrose) by HPLC-PAD using a set of Methrom columns (MetroSep A Supp 15 and Metrosep Carb1) in a Thermo Scientific™ Dionex™ ICS-5000+ Capillary HPIC™ system. Soluble anions (MSA, $SO4^{2-}$, $NO3^{-}$, Cl^{-}, Ox) and cations (Na^{+}, $NH4^{+}$, K^{+}, Mg^{2+}, Ca^{2+}) were analyzed by ion chromatography (IC, Dionex ICS3000) on the same extracts. AS/AG 11HC and CS/CG 12A columns were used for anions and cations analyses, respectively. Finally, twenty-five low molecular weight organic acids (C3-C9) were analyzed from the same extracts by LC-MS (DX500 – LCQ Fleet with an inverse phase C18 column). On some of the filters, more than twenty-five organic components were detected (glycolic acid, glyoxylic acid, tartaric acid, malic acid, lactic acid, malonic acid, succinic acid, hydroxybutyric acid, methylmalonic acid, fumaric acid,

ketobutyric acid, maleic acid, glutaric acid, oxoheptanedioic acid, citraconic acid, methlysuccinic acid, methylglutaric acid, adipic acid, pimelic acid, phtalic acid, pinic acid, isophtalic acid, suberic acid, benzoic acid, azelaic acid, and sebacic acid). Total mercury measurements from filter samples were performed with a DMA-80 (Milestone) analytical system based on the principles of sample thermal decomposition, mercury amalgamation and atomic absorption detection. Additionally, filters collected at AMS, CAP, PDD and PDM sampling sites were analyzed by LC-MS technique. Except for total mercury measurements (which were performed at GET, Toulouse), all analyses were performed at the AirOSol chemical analytical platform facility at IGE, Grenoble, France.

Results

For most of the sites, the TB were in the range of 10^1-10^2 16S rRNA copies per mm² (**Table 2**). A few high outliers remained and they could be attributed to the cleaning and packing procedures, and to the DNA extraction (including possible cross contamination during the subsampling phase). FB consisting of filters exposed to the atmosphere for up to one week but with no air forced to pass through, had, as expected, 16S rRNA gene concentrations one- to five-fold higher than the TB. These combined the passive contribution of the atmospheric environment and the DNA contamination occurring during the different phases of filter handling in the field. Except for the polar sites and CHC, the concentration of 16S rRNA gene copies in blank samples were < 0.3 % that in the corresponding atmospheric samples. The blanks at CHC were up to 7% of the average number of copies in the atmospheric samples, due to the low concentrations of DNA sampled from air at this high altitude site. At both polar sites (DMC and Villum) the 16S rRNA gene concentrations were similar to controls, indicating very low biomass.

Table 2: Number of 16S rRNA gene copies per mm^2. TB=Transportation blanks, FB 72h=Field blanks exposed for 72hours. NA indicates non available measurements.

Site code	Sampling site location	Type of sample	n	Minimum value	Maximum value	Mean	SD
AMS	Amsterdam Island, France	TB	2	29	41	35	8
		FB 72 h	2	85	115	100	21
		Samples	9	2941	82353	45686	26521
CAP	Cape Point Station, South Africa	TB	2	29	65	47	25
		FB 72 h	2	41	47	44	4
		Samples	7	4118	94118	42689	39281
CHC	Chacaltaya, Bolivia	TB	2	79	274	176	137
		FB 72 h ; FB 168 h	3	588	1088	838	354
		Samples	16	353	32353	11928	10088
DMC	Concordia Station, Antarctica	FB	2	41	76	59	25
		Samples	3	29	71	53	21
GRE	Grenoble, France	TB	NA	NA	NA	NA	NA
		FB 72 h	2	31	126	79	68
		Samples	10	12647	705882	326242	270626
NAM	NamCo, China	TB	2	941	1471	1206	374
		FB 24 h	1	-	-	233	-
		Samples	9	21765	882353	355686	304699
PDD	Puy-de-Dôme, France	TB	8	21	1765	477	639
		FB 48 h ; FB 144 h	2	143	882	401	417
		Samples	63	94	58823529	1418588	7510189
PDM	Pic-du-Midi, France	TB	3	40	94	58	31
		FB 168h	1	-	-	107	-
		Samples	14	1176	64706	21443	18917
STN	Villum research Station, Station Nord, Greenland	TB	2	74	235	154	114
		FB 72 h	2	59	529	294	333
		Samples	13	29	882	222	299
STP	Storm-Peak Laboratory, USA	TB	2	44	88	66	31
		FB 72 h	2	176	471	324	208
		Samples	7	706	441176	230353	177472

Organic carbon (OC) results (**Table 3**) also confirmed that the sampling protocols were adequately designed with a mean OC value of 0.55 ± 0.26 µg/cm^2 for all the transportation blanks, and of 0.85 ± 0.32 µg/cm² for the field blanks. In the case of the PDD samples, TB or FB represented less than 1.5% of the OC content in a sample. At remote sites with a very low OC concentration such as at AMS, the FB fraction reached up to 30% of the value in samples, but was thus still clearly distinct. Based on our experience in atmospheric chemistry field programs[49–51], these are low concentrations for blank series. As shown in **Figure 3**, our protocols significantly lower the transportation blank regarding OC. Inevitably, the handling of the filter clearly induces some contamination. This contamination can be significantly reduced when using a laminar flow hood, as in Grenoble (field blank of 0.49 µg/cm²) and PDD.

Table 3: Organic carbon concentrations expressed in µg per cm^2. TB=Transportation blanks, FB 72h=Field blanks exposed for 72hours. NA refers to non available data due to the absence of analytical data for this particular filter.

Site	Type of sample	n	Minimum value	Maximum value	Mean	SD
AMS	TB	2	0.45	0.51	0.48	0.04
	FB 72 h	2	0.54	1.10	0.82	0.39
	Samples	9	2.13	3.89	2.83	0.63
CAP	TB	2	0.69	0.72	0.71	0.01
	FB 72 h	2	0.87	1.29	1.08	0.29
	Samples	7	5.77	10.69	7.93	2.20
CHC	TB	2	0.21	0.23	0.22	0.01
	FB 72 h ; FB 168 h	2	0.49	0.59	0.54	0.05
	Samples	16	4.53	12.09	8.10	2.34
DC	FB	2	NA	NA	NA	NA
	Samples	3	NA	NA	NA	NA
GRE	TB	2	0.27	0.28	0.28	0.01
	FB 72 h	2	0.37	0.61	0.49	0.17
	Samples	10	73.50	145.65	105.10	21.87
NAM	TB	0	NA	NA		NA
	FB 24 h	1	NA	NA	0.99	NA
	Samples	9	3.97	12.09	8.50	2.87
PDD	TB	6	0.45	1.02	0.67	0.21
	FB 48 h ; FB 144 h	5	0.53	1.28	0.90	0.25
	Samples	63	13.63	166.50	68.53	35.91
PDM	TB	2	0.74	0.84	0.80	0.05
	FB 168h	1	-	-	1.22	-
	Samples	14	4.47	32.47	19.02	7.02
STN	TB	2	0.64	1.04	0.83	0.10
	FB 72 h	2	1.20	1.34	1.27	0.28
	Samples	13	1.87	9.66	5.01	2.41
STP	TB	2	0.29	0.32	0.31	0.01
	FB 72 h	2	0.25	0.75	0.50	0.25
	Samples	7	3.18	108.75	70.10	36.82

Figure 4 reports the range of OC and 16S rRNA gene concentrations measured at the different sampling sites. As expected, the urban site (Grenoble) had the highest OC values (3.63 ± 0.78 µg/m^3) and 16S rRNA gene concentrations, of around 10^6 copies/m^3 of air (1.1 ± 0.9x10^6). Comparable gene copies concentrations were observed at Storm-Peak (1.6 ± 1.1x10^6 copies/m^3), NamCo (3.6 ± 3.1x10^6 copies/m^3), but lower OC values comparatively to the Grenoble site.

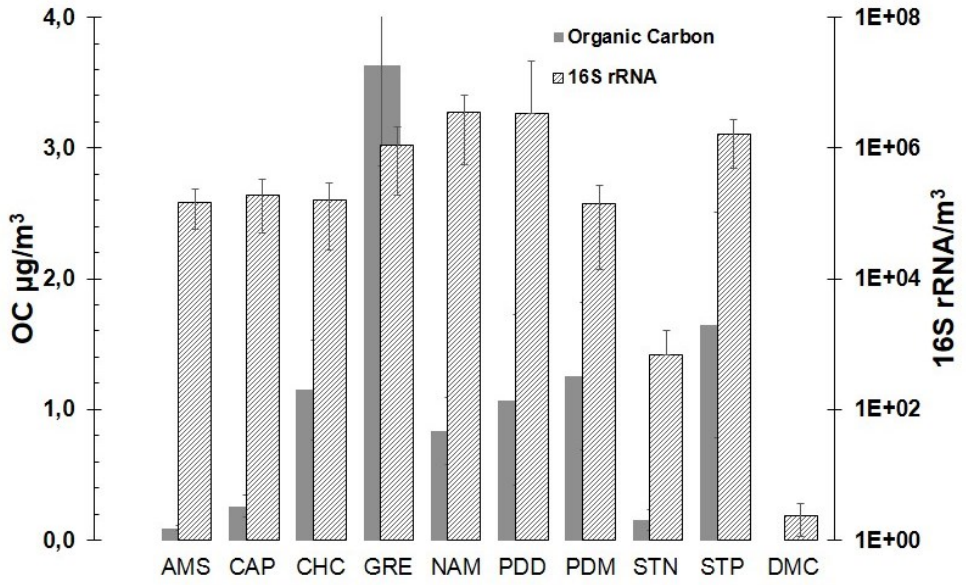


Figure 4: OC and 16S rRNA gene concentrations (mean ± SD) measured at the different sampling sites.

This indicates that DNA material has, at least in part, sources distinct from OC. Puy-de-Dôme station, where 53 samples were collected over one year, showed a range of gene copies from 10^4 to 10^8 copies/m^3 (3.4 ± 18x10^6 copies/m^3 in average (± standard deviation) illustrating the great temporal variability of airborne biological material in the air a single site. This is to be related with the wide diversity of air masses and meteorological conditions that can occur at a given site over a year (see for example[34]). More remote sites had all 16S gene concentrations one order of magnitude lower than at Puy-de-Dôme, such as Chacaltaya (1.6 ± 1.3x10^5 copies/m^3 - night samples only), Cape-Point (1.9 ± 1.4x10^5 copies/m^3), Pic-du-Midi (1.4 ± 1.3x10^5 copies/m^3) and Amsterdam-Island (1.5 ± 1.0x10^5 copies/m^3). In turn, the OC concentrations were clearly distinct between the sites, with Amsterdam Island showing the lowest OC levels of this study 0.09 ± 0.02 µg/m^3. The arctic site had a very low 16S rRNA gene concentration of 10^2/m^3 for corresponding average OC concentrations of 1.6 ± 0.9 µg/m^3, while the Antarctic site on the plateau showed gene copies similar to the blank.

Conclusion

We developed suitable easy-to-use and standardized protocols that generated consistent microbial and chemical datasets from the same samples at concentrations high enough for confident analysis. The validity of these protocols was then tested by deploying them on 10

sites over the globe and collect samples to explore atmospheric bacteria, fungi, or viruses over large spatial scales. For large scale studies coupling chemical measurements to biological measurements, we were able to demonstrate that quartz filters can be used, but that the extraction protocols must be optimized to maximize DNA yield. In addition to using the protocol outlined here, we also recommend the following:

- carefully prepare the filter and all the material using a combination of heating (500°C) and UV treatment (254 nm)
- provide detailed protocols intended to limit contamination (SOP, see SI for an example) to field users
- include enough control samples (including transportation and field work blanks) to monitor the quality of the sampling procedure
- carefully design atmospheric sampling at peak stations in order to take into account vertical turbulent mixing, and night-time hours (in general) should be preferred to avoid the influence of local sources of aerosols.
- Correct sampling times for the remoteness of the sites and for the measurements to be carried out. For example, a one week sampling with a volume of around 5000 (normalized) m^3 is sufficient for amplicon and metagenomic sequencing for most remote sites, except Antarctica, where the biomass is too low for DNA investigations even from total filtered volumes of 16000 m^3.
- Collected volumes should be normalized using STP or SATP standards.

One of the main disadvantages of a weekly sampling is the loss of information regarding rapid atmospheric chemistry processes and rapid changes in terms of aerosol sources. In turn, it has the advantage to smooth the data and avoid the stochastic-like behavior of biological content in the air often observed. With the development of better extraction protocols and more sensitive sequencing techniques, this limitation could be overcome, allowing for daily sampling in the future.

Author Contributions

AD, CL, TMV and PA conceived the study. OM implemented the technical phase and protocols together with AT and RTP. RTP performed the biomolecular work. JLJ supervised the analytical

work. MJ, LB, KS, JS, MA, IM, CL, LM, QZ helped and participated to the implementation of the field experiments

Funding

This program is funded by ANR-15-CE01-0002 – INHALE. We thank the Région Auvergne - Rhône Alpes for the doctoral grant for R. Tignat-Perrier. Part of the analytical equipment was supported at IGE by the AirOSol platform within Labex OSUG@2020 (ANR10 LABX56). AD and QZ thank the program CAMPUS France program XU GUANGQI (38718ZC INHALE_NAMCO).

Acknowlegdments

We thank IPEV GMOStral 1028 program. We are very grateful to G. Hallar and I. McCubbin at Storm Peak, J. Savarino, Nicolas Caillon, R. Jacob for DMC samples, Benny and Jesper for the sampling at Villum Research Station, and to B. Jensen and H. Skov for their support, I. Jouvie at Amsterdam Island. We thank R. Edwards, at Curtin University and J. Schauer and C. Worley at University of Wisconsin-Madison for lending us a high-volume sampler. We thank A. Nicosia, M. Ribeiro (OPGC/LaMP), L. Besaury (FRE), and the UMS 831 Pic du Midi Observatory team for help with sample collection. A. Poulain is acknowledged for his ideas and thoughts on the experimental design and strategy. We also acknowledge V. Lucaire, L. Jullien and A. Vella for the lab work in the AirOsol platform.

References

1. Graham, B. *et al.* Composition and diurnal variability of the natural Amazonian aerosol. *Journal of Geophysical Research: Atmospheres* **108**, (2003).
2. Huffman, J. A. *et al.* Size distributions and temporal variations of biological aerosol particles in the Amazon rainforest characterized by microscopy and real-time UV-APS fluorescence techniques during AMAZE-08. *Atmospheric Chemistry and Physics* **12**, 11997–12019 (2012).
3. Jaenicke, R. Abundance of cellular material and proteins in the atmosphere. *Science* **308**, 73 (2005).
4. Matthias-Maser, S. & Jaenicke, R. The size distribution of primary biological aerosol particles with radii > 0.2 µm in an urban/rural influenced region. *Atmospheric Research* **39**, 279–286 (1995).
5. Morris, C. E., Barny, M.-A., Berge, O., Kinkel, L. L. & Lacroix, C. Frontiers for research on the ecology of plant-pathogenic bacteria: fundamentals for sustainability. *Molecular Plant Pathology* **18**, 308–319 (2017).
6. Aylor, D. E. Spread of Plant Disease on a Continental Scale: Role of Aerial Dispersal of Pathogens. *Ecology* **84**, 1989–1997 (2003).
7. Soldevilla, C. G. *et al.* SPANISH AEROBIOLOGY NETWORK (REA): MANAGEMENT AND QUALITY MANUAL. 36

8. Mayol, E. *et al.* Long-range transport of airborne microbes over the global tropical and subtropical ocean. *Nature Communications* **8**, 201 (2017).
9. Michaud, J. M. *et al.* Taxon-specific aerosolization of bacteria and viruses in an experimental ocean-atmosphere mesocosm. *Nat Commun* **9**, 2017 (2018).
10. Monteil, C. L., Bardin, M. & Morris, C. E. Features of air masses associated with the deposition of Pseudomonas syringae and Botrytis cinerea by rain and snowfall. *ISME J* **8**, 2290–2304 (2014).
11. Sesartic, A., Lohmann, U. & Storelvmo, T. Bacteria in the ECHAM5-HAM global climate model. *Atmospheric Chemistry and Physics* **12**, 8645–8661 (2012).
12. Pouzet, G. *et al.* Atmospheric Processing and Variability of Biological Ice Nucleating Particles in Precipitation at Opme, France. *Atmosphere* **8**, 229 (2017).
13. Morris, C. E. *et al.* Bioprecipitation: a feedback cycle linking earth history, ecosystem dynamics and land use through biological ice nucleators in the atmosphere. *Glob Chang Biol* **20**, 341–351 (2014).
14. Fröhlich-Nowoisky, J. *et al.* Bioaerosols in the Earth system: Climate, health, and ecosystem interactions. *Atmospheric Research* **182**, 346–376 (2016).
15. Vaïtilingom, M. *et al.* Potential impact of microbial activity on the oxidant capacity and organic carbon budget in clouds. *PNAS* **110**, 559–564 (2013).
16. Hughes, K. A. & Convey, P. The protection of Antarctic terrestrial ecosystems from inter- and intra-continental transfer of non-indigenous species by human activities: A review of current systems and practices. *Global Environmental Change* **20**, 96–112 (2010).
17. Bar-On, Y., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proceedings of the National Academy of Sciences* **115**, 201711842 (2018).
18. Leyronas, C., Morris, C. E., Choufany, M. & Soubeyrand, S. Assessing the Aerial Interconnectivity of Distant Reservoirs of Sclerotinia sclerotiorum. *Front Microbiol* **9**, 2257 (2018).
19. Reche, I., D'Orta, G., Mladenov, N., Winget, D. M. & Suttle, C. A. Deposition rates of viruses and bacteria above the atmospheric boundary layer. *ISME J* **12**, 1154–1162 (2018).
20. Burrows, S. M., Elbert, W., Lawrence, M. G. & Pöschl, U. Bacteria in the global atmosphere – Part 1: Review and synthesis of literature data for different ecosystems. *Atmos. Chem. Phys.* **9**, 9263–9280 (2009).
21. Bowers, R. M. *et al.* Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. *Environ. Sci. Technol.* **47**, 12097–12106 (2013).
22. Joung, Y. S., Ge, Z. & Buie, C. R. Bioaerosol generation by raindrops on soil. *Nat Commun* **8**, 14668 (2017).
23. Després, V. *et al.* Primary biological aerosol particles in the atmosphere: a review. *Tellus B: Chemical and Physical Meteorology* **64**, 15598 (2012).
24. Carotenuto, F. *et al.* Measurements and modeling of surface–atmosphere exchange of microorganisms in Mediterranean grassland. *Atmospheric Chemistry and Physics* **17**, 14919–14936 (2017).
25. Wainwright, M., Wickramasinghe, N. C., Narlikar, J. V. & Rajaratnam, P. Microorganisms cultured from stratospheric air samples obtained at 41 km. *FEMS Microbiol. Lett.* **218**, 161–165 (2003).
26. DeLeon-Rodriguez, N. Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. (2013). Available at: http://www.pnas.org/content/110/7/2575.full. (Accessed: 25th July 2017)
27. Smith, D. J. *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* **9**, 1752 (2018).
28. Smith, D. J. Aeroplankton and the Need for a Global Monitoring Network. *BioScience* **63**, 515–516 (2013).
29. Smith, D. J. *et al.* Intercontinental Dispersal of Bacteria and Archaea by Transpacific Winds. *Appl Environ Microbiol* **79**, 1134–1139 (2013).

30. Griffin, D. W., Gonzalez-Martin, C., Hoose, C. & Smith, D. J. Global-Scale Atmospheric Dispersion of Microorganisms. in *Microbiology of Aerosols* 155–194 (John Wiley & Sons, Ltd, 2017). doi:10.1002/9781119132318.ch2c
31. Barberán, A. *et al.* Continental-scale distributions of dust-associated bacteria and fungi. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5756–5761 (2015).
32. Burrows, S. M. *et al.* Bacteria in the global atmosphere – Part 2: Modeling of emissions and transport between different ecosystems. *Atmospheric Chemistry and Physics* **9**, 9281–9297 (2009).
33. Amato, P. *et al.* Active microorganisms thrive among extremely diverse communities in cloud water. *PLOS ONE* **12**, e0182869 (2017).
34. Deguillaume, L. *et al.* Classification of clouds sampled at the puy de Dôme (France) based on 10 yr of monitoring of their physicochemical properties. *Atmospheric Chemistry and Physics* **14**, 1485–1506 (2014).
35. Bertolini, V. *et al.* Temporal variability and effect of environmental variables on airborne bacterial communities in an urban area of Northern Italy. *Appl. Microbiol. Biotechnol.* **97**, 6561–6570 (2013).
36. Jiang, W. *et al.* Optimized DNA extraction and metagenomic sequencing of airborne microbial communities. *Nat Protoc* **10**, 768–779 (2015).
37. Luhung, I. *et al.* Protocol Improvements for Low Concentration DNA-Based Bioaerosol Sampling and Analysis. *PLOS ONE* **10**, e0141158 (2015).
38. Pearce, D. A. *et al.* Aerobiology Over Antarctica - A New Initiative for Atmospheric Ecology. *Front Microbiol* **7**, 16 (2016).
39. Cáliz, J., Triadó-Margarit, X., Camarero, L. & Casamayor, E. O. A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12229–12234 (2018).
40. Haddrell, A. E. & Thomas, R. J. Aerobiology: Experimental Considerations, Observations, and Future Tools. *Appl. Environ. Microbiol.* **83**, e00809-17 (2017).
41. Jaffrezo, J.-L., Aymoz, G. & Cozic, J. Size distribution of EC and OC in the aerosol of Alpine valleys during summer and winter. *Atmospheric Chemistry and Physics* **5**, 2915–2925 (2005).
42. Després, V. R. *et al.* 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 (2007).
43. Smith, D. J. *et al.* Free Tropospheric Transport of Microorganisms from Asia to North America. *Microb Ecol* **64**, 973–985 (2012).
44. Fierer, N., Jackson, J. A., Vilgalys, R. & Jackson, R. B. Assessment of Soil Microbial Community Structure by Use of Taxon-Specific Quantitative PCR Assays. *Appl. Environ. Microbiol.* **71**, 4117–4120 (2005).
45. Delmont, T. O. *et al.* Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J* **6**, 1677–1687 (2012).
46. Zielińska, S. *et al.* The choice of the DNA extraction method may influence the outcome of the soil microbial community structure analysis. *Microbiologyopen* **6**, (2017).
47. Vandeventer, P. E., Mejia, J., Nadim, A., Johal, M. S. & Niemz, A. DNA adsorption to and elution from silica surfaces: influence of amino acid buffers. *J Phys Chem B* **117**, 10742–10749 (2013).
48. Sciare, J. *et al.* Long-term observations of carbonaceous aerosols in the Austral Ocean atmosphere: Evidence of a biogenic marine organic source. *Journal of Geophysical Research: Atmospheres* **114**, (2009).
49. Brunke, E.-G., Labuschagne, C., Ebinghaus, R., Kock, H. H. & Slemr, F. Gaseous elemental mercury depletion events observed at Cape Point during 2007–2008. *Atmospheric Chemistry and Physics* **10**, 1121–1131 (2010).
50. Andrade, M. *et al.* Puesta en marcha de una nueva estación de monitoreo climático en los andes centrales de Bolivia: la estación Gaw/Chacaltaya. **26**, 06–15 (2015).

51. Rose, C. *et al.* Frequent nucleation events at the high altitude station of Chacaltaya (5240 m a.s.l.), Bolivia. *Atmospheric Environment* **102**, 18–29 (2015).
52. Angot, H. *et al.* New insights into the atmospheric mercury cycling in central Antarctica and implications on a continental scale. *Atmospheric Chemistry and Physics* **16**, 8249–8264 (2016).
53. Yin, X. *et al.* Surface ozone at Nam Co in the inland Tibetan Plateau: variation, synthesis comparison and regional representativeness. *Atmospheric Chemistry and Physics* **17**, 11293–11311 (2017).
54. Sellegri, K. *et al.* Contribution of gaseous and particulate species to droplet solute composition at the Puy de Dôme, France. *Atmospheric Chemistry and Physics* **3**, 1509–1522 (2003).
55. Vaïtilingom, M. *et al.* Long-term features of cloud microbiology at the puy de Dôme (France). *Atmospheric Environment* **56**, 88–100 (2012).
56. Gabey, A. M. *et al.* Observations of fluorescent and biological aerosol at a high-altitude site in central France. *Atmospheric Chemistry and Physics* **13**, 7415–7428 (2013).
57. Farah, A. *et al.* Seasonal Variation of Aerosol Size Distribution Data at the Puy de Dôme Station with Emphasis on the Boundary Layer/Free Troposphere Segregation. *Atmosphere* **9**, 244 (2018).
58. Gheusi, F. *et al.* Pic 2005, a field campaign to investigate low-tropospheric ozone variability in the Pyrenees. *Atmospheric Research* **101**, 640–665 (2011).
59. Fu, X., Heimbürger, L.-E. & Sonke, J. E. Collection of atmospheric gaseous mercury for stable isotope analysis using iodine- and chlorine-impregnated activated carbon traps. *J. Anal. At. Spectrom.* **29**, 841–852 (2014).
60. Fu, X., Marusczak, N., Wang, X., Gheusi, F. & Sonke, J. E. Isotopic Composition of Gaseous Elemental Mercury in the Free Troposphere of the Pic du Midi Observatory, France. *Environ. Sci. Technol.* **50**, 5641–5650 (2016).
61. Fenger, M. *et al.* Sources of anions in aerosols in northeast Greenland during late winter. *Atmospheric Chemistry and Physics* **13**, 1569–1578 (2013).
62. Borys, R. D. & Wetzel, M. A. Storm Peak Laboratory: A Research, Teaching, and Service Facility for the Atmospheric Sciences. *Bull. Amer. Meteor. Soc.* **78**, 2115–2124 (1997).
63. Obrist, D., Hallar, A. G., McCubbin, I., Stephens, B. B. & Rahn, T. Atmospheric mercury concentrations at Storm Peak Laboratory in the Rocky Mountains: Evidence for long-range transport from Asia, boundary layer contributions, and plant mercury uptake. *Atmospheric Environment* **42**, 7579–7589 (2008).
64. Waked, A. *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. *Atmospheric Chemistry and Physics* **14**, 3325–3346 (2014).
65. Birch, M. E. & Cary, R. A. Elemental Carbon-Based Method for Monitoring Occupational Exposures to Particulate Diesel Exhaust. *Aerosol Science and Technology* **25**, 221–241 (1996).
66. Piot, C. *et al.* Quantification of levoglucosan and its isomers by High Performance Liquid Chromatography – Electrospray Ionization tandem Mass Spectrometry and its applications to atmospheric and soil samples. *Atmospheric Measurement Techniques* **5**, 141–148 (2012).
67. Sprovieri, F. *et al.* Atmospheric mercury concentrations observed at ground-based monitoring sites globally distributed in the framework of the GMOS network. *Atmospheric Chemistry and Physics* **16**, 11915–11935 (2016).
68. Daellenbach, K. R. *et al.* Long-term chemical analysis and organic aerosol source apportionment at nine sites in central Europe: source identification and uncertainty assessment. *Atmospheric Chemistry and Physics* **17**, 13265–13282 (2017).
69. Pandolfi, M. *et al.* A European aerosol phenomenology – 6: scattering properties of atmospheric aerosol particles from 28 ACTRIS sites. *Atmospheric Chemistry and Physics* **18**, 7877–7911 (2018).

Chapter 2: Methods to investigate the global atmospheric microbiome

Section 2: Molecular biology analyses

We proposed a sampling strategy and experimental protocol to study the atmospheric microbiome and atmospheric particulate matter chemistry at a global scale. We discussed our choices regarding the sampling method on quartz filters. We explained the different steps that have led us to develop an optimal DNA extraction method for airborne microorganisms collected on quartz filters.

In the second part of Chapter 2, we will present the other molecular biology analyses used to estimate bacterial and fungal cell concentrations as well as the taxonomic diversity and functional potential of airborne microbial communities.

Real-time quantitative PCR method (qPCR)

To estimate bacterial and fungal cell concentration, we used real time qPCR analysis on the total DNA extracted from quartz filters. Concentrations were reported in number of cells per cubic meter of air.

16S rRNA gene qPCR. The bacterial cell concentration was approximated by the number of 16S rRNA gene copies per cubic meter of air, which is a gene present in every prokaryotic cell. The V3 region of the 16S rRNA gene was amplified using the SensiFast SYBR No-Rox kit (Bioline) on a Rotorgene 3000 machine (Qiagen). The primer pair Eub 338f/Eub 518r were chosen for their sensitivity and specificity[1]. Still, in silico ncbi PCR analysis revealed that archaeal sequences might amplify using these primers (without mismatch).

18S rRNA gene qPCR. The fungal cell concentration was approximated by the number of 18S rRNA gene copies per cubic meter of air, which is a gene present in every eukaryotic cell. The region located at the end of the SSU 18S rRNA gene, near the ITS1 region, was quantified using the primer pairFR1/FF390 which is specific to fungi[2]. In silico analysis revealed that 18.2% of the sequences matched with non-fungal eukaryotic sequences (considering no mismatch)[2].

Amplicon and metagenomic MiSeq Illumina sequencing

Investigation on the relative importance of sequencing depth and DNA extraction method on the observed microbial richness (Sanchez-Cid et al. (submitted), in Side Publications)

While culture-based methods access roughly 1% of the diversity of a sample, culture-independent methods access to a larger part of the diversity, although still far from the actual diversity[3,4]. We used high throughput DNA sequencing to investigate airborne microbial

diversity. Bacterial and fungal amplicon sequencing was based on the sequencing of a region of the 16S rRNA gene and ITS region, respectively. The produced reads were then annotated to specific taxa based on their nucleotide sequence and public protein databases (see Material and Methods section in Chapter 3 for details). The taxonomic diversity is dependent on the size and accuracy of the databases. Metagenomic sequencing was based on sequencing of genomic DNA, then annotation of the produced reads to putative functional genes using databases (see Material and Methods section in Chapter 5 for details).

Due to the bias inherent to the DNA extraction method, sequencing and read annotation[5–7], the taxonomic and functional microbial diversity identified using high throughput sequencing is still partial. It is also of an unknown accuracy from the actual diversity, both in terms of the number of taxa or functions observed (richness) and their relative abundance (evenness). The DNA extraction method has been shown to have a critical impact on the observed microbial community diversity of a sample[5,7]. Depending on the method employed (physical versus chemical cell lysis, bead-beating step or not etc.), the lysis step would be more or less efficient and specific. More violent lysis techniques have been suggested to extract DNA from cells difficult to lyse. The sequencing technology (*i.e.* read length, error rate, sequencing depth) has also been shown to affect the observed microbial community diversity. Furthermore, within a sequencing technology, the sequencing depth has been correlated to the observed richness of a sample[6]. The larger is the sequencing depth, the higher the number of different taxa or functions are detected[6]. In *Sanchez-Cid et al.* (see the paper in SI), we evaluated the relative contribution of the sequencing depth and the DNA extraction method on the taxonomic and functional richness observed after amplicon and metagenomic sequencing, respectively. We used the MiSeq Illumina technology.

The different DNA extraction methods used (QIAGEN DNeasyPowerSoil kit, ZymoBIOMICS DNA Mini kit, Phenol/Chloroform extraction, Maxwell PureFood 1 and 2) did not all work on our INHALE air samples (quartz filter samples). To get enough DNA to do the subsequent molecular biology analyses from the air samples, we needed to set up an adapted DNA extraction method, as presented above in this Chapter[8]. Thus, to do our investigation on the relative importance of the sequencing depth and DNA extraction method on the observed taxonomic and functional richness of a sample, we used two soil samples (Scottish agricultural and Côte Saint-André soils). We showed that sequencing depth had a greater influence than the DNA extraction method on bacterial richness discovery at both taxonomical and functional levels. Furthermore, at an equal sequencing depth, the differences observed between methods in both soils were more likely a product of random subsampling than that of the DNA extraction itself.

This study allowed me to develop an expertise in MiSeq Illumina sequencing technology, from library preparation to the use of the machine and analysis of data. I leveraged this expertise to the metagenomic investigation and interpretation of airborne microbial community diversity.

Amplicon sequencing

Library preparation. The V3-V4 region of the 16S rRNA gene was amplified using the Platinum Taq Polymerase (ThermoFisher Scientific) and the primer pair recommended by Illumina library preparation protocol ("16S Metagenomic Sequencing Library Preparation"). To test the reproducibility of the molecular biology analyses, we divided by two (technical duplicates) two samples and did the different analyses from the DNA extraction to the 16s rRNA gene sequencing separately. We validated that the structure of the two duplicated samples was much more similar between each other based on the Bray-Curtis dissimilarity value than with any other sample (see Chapter 2). The ITS2 region was amplified using the primer pair ILL_5.8S_Fun/ILL_ITS4_Fun[9].

The other steps of the library preparation (amplicon PCR clean-up, index PCR, index PCR clean-up, normalization and pooling) were performed following the Illumina library preparation protocol. The amplicons were sequenced by a paired-end MiSeq sequencing using the technology V3 (16S rRNA gene) and V2 (ITS) of Illumina with 2 x 250 cycles.

Reads quality filtering and taxonomic annotation. The base quality of the reads 1 and reads 2 was controlled using the FASTX-Toolkit software. PANDAseq[10] was used to assemble the read 1 and the read 2 for each sample. The resulting sequences were annotated at the genus or species level by RDP Classifier[11] using the RDP 16srrna and fungallsu databases for *16S rRNA* gene and ITS sequencing, respectively. The RDP 16srrna database includes archaeal sequences which represented maximum <0.1% (average under 0.05%) of the annotated sequences in our dataset.

Metagenomic sequencing

Metagenomic library preparation. Metagenomic libraries were prepared from 1 ng of DNA using the Nextera XT Library Prep Kit and Indexes, as detailed in Illumina's "Nextera XT DNA Library Prep Kit" reference guide with some modifications for samples whose DNA concentration was < 1 ng as follows. The tagmented DNA was amplified over 13 PCR cycles instead of 12 PCR cycles, and the libraries (after indexing) were resuspended in 30 µL of RBS buffer instead of 52.5 µL. DNA sequencing was performed using the technology V2 and MiSeq system of Illumina with 2 x 250 cycles.

Reads quality filtering and taxonomic and functional annotation of the metagenomic reads. The read 1 and read 2 for each sample were not assemble but merge in a common file before filtering them based on read quality using the software FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Filtered reads were functionally annotated using Diamond and the NR (non-redundant) database. Functions were then grouped by the SEED last level functional class (around 7000 classes) using MEGAN6[12] (see Chapter 3 for details). We also annotated the filtered reads using the eggNOG-Mapper software[13] (using Diamond and specific databases; see Chapter 3 for details).

References

1. Fierer, N., Jackson, J. A., Vilgalys, R. & Jackson, R. B. Assessment of Soil Microbial Community Structure by Use of Taxon-Specific Quantitative PCR Assays. *Appl. Environ. Microbiol.* **71**, 4117–4120 (2005).
2. Chemidlin Prévost-Bouré, N. *et al.* Validation and application of a PCR primer set to quantify fungal communities in the soil environment by real-time quantitative PCR. *PLoS ONE* **6**, e24166 (2011).
3. Caro-Quintero, A. & Ochman, H. Assessing the Unseen Bacterial Diversity in Microbial Communities. *Genome Biol Evol* **7**, 3416–3425 (2015).
4. Rappé, M. S. & Giovannoni, S. J. The uncultured microbial majority. *Annu. Rev. Microbiol.* **57**, 369–394 (2003).
5. Brooks, J. P. *et al.* The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiology* **15**, 66 (2015).
6. Lemos, L. N., Fulthorpe, R. R., Triplett, E. W. & Roesch, L. F. W. Rethinking microbial diversity analysis in the high throughput sequencing era. *Journal of Microbiological Methods* **86**, 42–51 (2011).
7. Cruaud, P. *et al.* Influence of DNA extraction method, 16S rRNA targeted hypervariable regions, and sample origin on microbial diversity detected by 454 pyrosequencing in marine chemosynthetic ecosystems. *Appl. Environ. Microbiol.* **80**, 4626–4639 (2014).
8. Dommergue, A. *et al.* Methods to investigate the global atmospheric microbiome. *Front. Microbiol.* **10**, (2019).
9. Taylor, D. L. *et al.* Accurate Estimation of Fungal Diversity and Abundance through Improved Lineage-Specific Primers Optimized for Illumina Amplicon Sequencing. *Appl. Environ. Microbiol.* **82**, 7217–7226 (2016).
10. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**, 31 (2012).
11. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology* **73**, 5261–5267 (2007).
12. Huson, D. H., Richter, D. C., Mitra, S., Auch, A. F. & Schuster, S. C. Methods for comparative metagenomics. *BMC Bioinformatics* **10 Suppl 1**, S12 (2009).
13. Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).

Chapter 3: Global airborne microbial communities controlled by surrounding landscapes and wind conditions

Under review in Scientific Reports

Romie Tignat-Perrier[1,2], Aurélien Dommergue[1], Alban Thollot[1], Christoph Keuschnig[2], Olivier Magand[1], Timothy M. Vogel[2], Catherine Larose[2]

[1]Institut des Géosciences de l'Environnement, Université Grenoble Alpes, CNRS, IRD, Grenoble INP, Grenoble, France

[2]Environmental Microbial Genomics, Laboratoire Ampère, Ecole Centrale de Lyon, Université de Lyon, Ecully, France

Abstract

The atmosphere is an important route for transporting and disseminating microorganisms over short and long distances. Understanding how microorganisms are distributed in the atmosphere is critical due to their role in public health, meteorology and atmospheric chemistry. In order to determine the dominant processes that structure airborne microbial communities, we investigated the diversity and abundance of both bacteria and fungi from the PM10 particle size (particulate matter of 10 micrometers or less in diameter) as well as particulate matter chemistry and local meteorological characteristics over time at nine different meteorological stations around the world. The bacterial genera *Bacillus* and *Sphingomonas* as well as the fungal species *Pseudotaeniolina globaosa* and *Cladophialophora proteae* were the most abundant taxa of the dataset, although their relative abundances varied greatly based on sampling site. Bacterial and fungal concentration was the highest at the high-altitude and semi-arid plateau of Namco (China; $3.56x10^6 \pm 3.01x10^6$ cells/m^3) and at the high-altitude and vegetated mountain peak Storm-Peak (Colorado, USA; $8.78x10^4 \pm 6.49x10^4$ cells/m^3), respectively. Surrounding ecosystems, especially within a 50 km perimeter of our sampling stations, were the main contributors to the composition of airborne microbial communities. Temporal stability in the composition of airborne microbial communities was mainly explained by the diversity and evenness of the surrounding landscapes and the wind direction variability over time. Airborne microbial communities appear to be the result of large inputs from nearby sources with possible low and diluted inputs from distant sources.

Introduction

Microbial transport in the atmosphere is critical for understanding the role microorganisms play in meteorology, atmospheric chemistry and public health. Recently, studies have shown that up to 10^6 microbial cells can be found in one cubic meter of air[1] and that they might be metabolically active[2,3]. Different processes, including aerosolization and transport, might be important in selecting which microorganisms exist in the atmosphere. For example, specific bacterial taxa (*e.g.*, *Actinobacteria* and some *Gammaproteobacteria*) have been proposed to be preferentially aerosolized from oceans[4]. Once aerosolized, microbial cells enter the planetary boundary layer, defined as the air layer near the ground, directly influenced by the planetary surface, from which they might eventually be transported upwards by air currents into the free troposphere (air layer above the planetary boundary layer) or even higher into the stratosphere[5–8]. Microorganisms might undergo a selection process during their way up into the troposphere and the stratosphere[9]. Studies from a limited number of sites have investigated possible processes implicated in the observed microbial community distribution in the atmosphere, such as meteorology[1,2,10–12], seasons[11,13–16], surface conditions[12–14,16] and global air circulation[11,17–20]. Most research has described the airborne microbial communities at one specific site per study. Airborne fungal communities (not only fungal spores) have frequently been overlooked, despite constituting a significant health concern for crops[21,22] and in allergic diseases[23]. A few have initiated the investigation of microbial geographic distribution by examining regional and even continental patterns[10,16,19], although in most cases at a only few time points. Some long-range transport has been reported between regions separated by thousands of kilometers[5,6]. Several factors, such as the local landscapes, local meteorological conditions, and inputs from long-range transport have all been cited as partially responsible for the composition of airborne microbial communities[11,14,17,24–26], but their relative contribution remains unclear. Probabilistically, proximity should have an effect, and therefore, local Earth sources of microorganisms should contribute significantly to atmospheric microbial communities especially in the planetary boundary layer. In addition, meteorological conditions (*e.g.*, wind speed and direction) might lead to different temporal variability of airborne microbial communities by mediating the relative inputs of microbial populations from the different surrounding landscapes. Our goal was to evaluate the relative importance of environmental processes on the geographical and temporal variations in

airborne microbial communities. This was carried out by following changes in community structure and abundance of both airborne bacteria and fungi as well as particulate matter chemistry and local meteorological characteristics over time at nine sites around the world.

Material and Methods

Air samples (seven to sixteen per site) were collected in 2016 and 2017 at nine sites from different latitudes (from the Arctic to the sub-Antarctica) and elevations from sea level (from 59 m to 5230 m; **Fig. 1** and supplementary **Table S1**). We collected particulate matter smaller than 10 µm (PM10) on pre-treated quartz fiber filters using high volume air samplers (TISCH, DIGITEL, home-made) equipped with a PM10 size-selective inlet. Quartz fiber filters were heated to 500°C for 8 hours to remove traces of organic carbon including DNA. All the material including the filter holders, aluminium foils and plastic bags in which the filters were transported were UV-sterilized as detailed in Dommergue *et al.* (2019)[27] (Chapter 2). A series of field and transportation blank filters were done to monitor and check the quality of the sampling protocol as presented in Dommergue *et al.* (2019)[27] (Chapter 2).

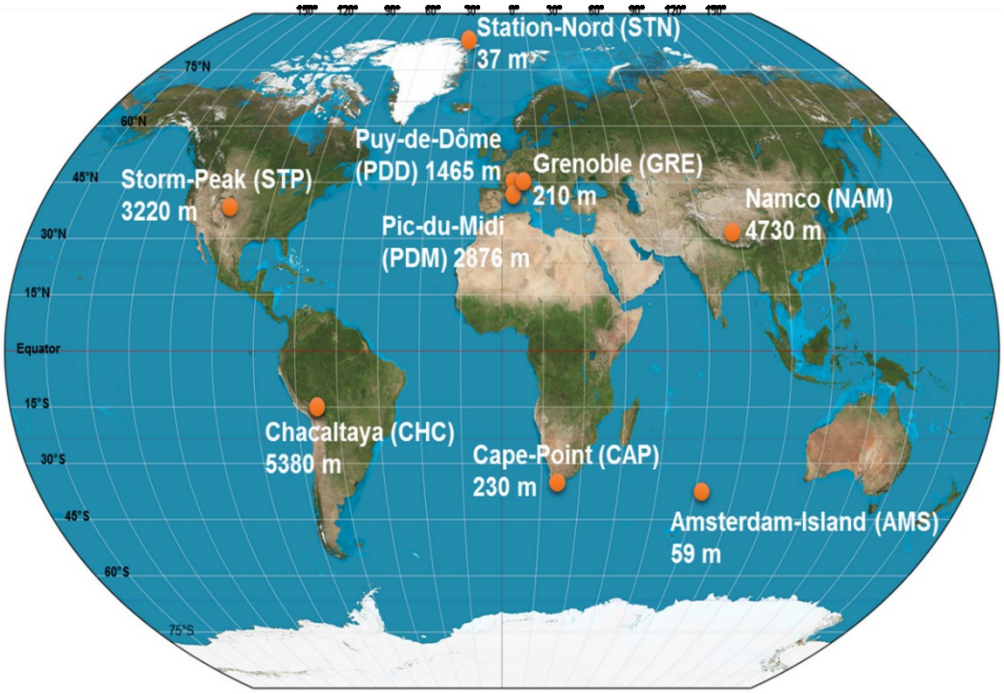


Figure 1: Map showing the geographical location and elevation from sea level of the nine sampling sites.

The collection time per sample lasted one week. Depending on the site, the collected volumes ranged from 2000 m^3 to 10000 m^3 after standardization using SATP standards (Standard Ambient Pressure and Temperature) (supplementary **Table S1**). The environment type varied from marine (Amsterdam-Island) to coastal (Cape-Point), polar (Station-Nord) and terrestrial (Grenoble, Chacaltaya, Puy-de-Dôme, Pic-du-Midi, Storm-Peak and Namco) (**Table 1**). For mountain peaks, we sampled at night to minimize sampling in the planetary boundary layer (the filter was left in the sampler during day time). Detailed sampling protocols are presented in *Dommergue et al.* (2019)[27]. Elemental carbon (EC), organic carbon (OC), sugar anhydrides and alcohols (levoglucosan, mannosan, galactosan, inositol, glycerol, erythriol, xylitol, arabitol, sorbitol, mannitol, trehalose, rhamnose, glucose), major soluble anions (methylsulfonic acid [MSA], SO_4^{2-}, NO_3^-, Cl^-, oxalate) and cations (Na^+, NH_4^+, K^+, Mg^{2+}, Ca^{2+}) were analyzed[27].

At all sampling sites, meteorological parameters (*i.e.* wind speed and direction, temperature and relative humidity) were collected every hour. Meteorological data were used to produce wind roses using the openair R package[28]. For each sample, backward trajectories of the air masses were calculated over 3 days (maximum height from sea level: 1 km) using HYSPLIT[29] and plotted on geographical maps using the openair R package.

We extracted DNA from 3 circular pieces (punches) from the quartz fiber filters (diameter of one punch: 38 mm) using the DNeasy PowerWater kit with the following modifications as detailed in *Dommergue et al.* (2019)[27]. During the cell lysis, we heated at 65°c the PowerBead tube containing the 3 punches and the pre-heated lysis solution during one hour after a 10-min vortex treatment at maximum speed. We then centrifuged the mixture at 1000 rcf during 4 min to separate the filter debris from the lysate using a syringe. From this step, we continued the extraction following the DNeasy PowerWater protocol. DNA concentration was measured using the High Sensitive Qubit Fluorometric Quantification (Thermo Fisher Scientific) then DNA was stored at -20°C. Real-time qPCR analyses on the 16S rRNA (data previously published in Dommergue *et al.* (2019)[27] or Chapter 2) and 18S rRNA genes were carried out (regions, primers[30,31] and protocols in Supplementary Information) to approximate the concentration of bacterial and fungal cells per cubic meter of air. Although bacteria and fungi might have more than one copy of 16S rRNA and 18S rRNA gene per genome, respectively, attempts to correct for metagenomics datasets are unproductive[32]. Microbial community structure was obtained using MiSeq Illumina amplicon sequencing of the bacterial V3-V4 region of the 16S

rRNA gene and the fungal ITS2 region. The library preparation protocol, read quality filtering and taxonomic annotation of every sequence using RDP Classifier[33] are detailed in Supplementary Information. RDP classifier was used in part to avoid errors due to sequence clustering. The raw read number per sample and the percentage of sequences annotated using RDP Classifier for both the 16S rRNA gene and ITS sequencings are presented in supplementary **Table S1**.

All graphical and multivariate statistical analyses were carried out in the R environment, using the vegan[34], pvclust[35], ade4[36] and Hmisc[37] R packages. The raw abundances of the bacterial genera or fungal species were transformed in relative abundances to counter the heterogeneity in the number of sequences per sample and then standardized using Hellinger's transformation. Chao1 estimations of the richness for both bacterial and fungal communities were calculated and averaged for each site. MODIS (Moderate resolution imaging spectroradiometer) land cover approach (5'x5' resolution)[38,39] was used to quantify landscapes in a diameter range of the sampling sites (50, 100 and 300 km). The perimeter of 50 km was chosen because the landscapes were best correlated to airborne microbial community structures. The different MODIS land covers are described in supplementary **Table S2**. We weighted these relative surfaces by their associated bacterial cell concentration reported by Burrows *et al.* (2009)[40,41] (supplementary **Table S3**) to predict the relative contribution of each landscape to the aerial emission of bacterial cells.

Hierarchical clustering analyses (average method) were carried out on either the Bray-Curtis dissimilarity matrix (bacterial and fungal community structure) or the Euclidean distance matrix (PM10 chemistry, landscapes and the relative contributions of the landscapes). A Mantel test was used to evaluate the similarities in the distribution of the samples in the different data sets. Distance-based redundancy analyses (RDA) were carried out to evaluate the part of the variance between the samples based on the microbial community structure explained by chemistry. Prior to the cluster analysis based of the chemical dataset, chemical concentrations were log10-transformed to approach a Gaussian distribution. SIMPER analysis was used to identify the major bacterial or fungal contributors to the difference between the detected groups.

Spearman correlations were calculated to test the correlation between microbial abundance and richness and quantitative environmental factors like chemical concentrations or meteorological parameters. ANOVAs were used to test the influence of qualitative factors

such as localization on both bacterial and fungal abundance and richness and TukeyHSD tests to identify which group had a significantly different mean. To compare the temporal variability in microbial communities at each site, we interpreted both the Bray-Curtis dissimilarity value averaged per site and the associated standard deviation. We calculated a new statistic, *i.e.* a similarity index, as follows. We averaged the values of dissimilarity obtained from the Bray-Curtis matrix for each pair of samples from the same site. Then we subtracted these values from 1 to get similarity values and finally we divided the similarity values by the standard deviation ($\frac{1 - Bray\ Curtis\ dissimilarity\ value\ averaged\ per\ site}{Bray\ Curtis\ standard\ deviation}$). The higher is the similarity in microbial community structure between the samples, the higher is the similarity index. A multiple linear regression model was used to test the influence of wind and surrounding landscape characteristics (number of different landscapes and landscape evenness) of the sites on the temporal variability of both the fungal and bacterial community structure (evaluated by the similarity index). Pielou's evenness was used to determine how similar the different relative surfaces of each landscape surrounding the sites were. The temporal variability of meteorological parameters was calculated both within weekly samples (average variance of hourly data for one week) and between weekly samples of the same site (standard deviation of the above average variance over weeks).

Results

Geographical distribution of airborne microbial communities

Airborne microbial concentrations varied between 9.2×10^{1} to 1.3×10^{8} cells per cubic meter of air for bacteria and from not detectable ($< 2 \times 10^{0}$) to 1.9×10^{5} cells per cubic meter of air for fungi. A high correlation was observed between the bacterial and the fungal concentrations from all the sites and all sampling times ($N = 95$, $R = 0.85$, $P = 2.2 \times 10^{-16}$). The average bacterial and fungal concentrations per site varied from 7.3×10^{2} to 3.6×10^{6} and from 5.2×10^{0} to 8.8×10^{4} cells per cubic meter of air, respectively, and were different between the sites ($P = 2.2 \times 10^{-11}$ and $P = 5.7 \times 10^{-15}$ for bacteria and fungi, respectively; **Table 1**). The polar site Station-Nord and the high altitude plateau site of Namco had the lowest and highest average bacterial concentration, respectively. The highest average fungal concentration was observed in atmospheric samples from the urban site of Grenoble and the mountain site of Storm-Peak, while the lowest was from Station-Nord (**Table 1**).

Table 1: Summary of bacterial and fungal abundances and bacterial (genus level) and fungal (species level) Chao1 richness estimations averaged per site and associated to a standard deviation. Reference letters indicate the group membership based on Tukey's HSD post hoc tests. The environment type, coordinates, elevation from sea level, collection start and end per site and the number of samples collected per site are shown. The 16S rRNA gene qPCR data was previously published in Dommergue *et al.*, 2019 (Chapter 2).

Site	Name	Environment type	Coordinates and elevation from sea level	Collection start and end	Number of samples	16S rRNA gene copies/m^3	18S rRNA gene copies/m^3	Bacterial Chao1 richness estimation	Fungal Chao1 richness estimation
AMS	Amsterdam-Island, France	Marine, remote	37°47'82"S 77°33'04''E 59 m asl	07/09/2016 - 10/11/2016	9	$1.49x10^{5} \pm 9.17x10^{4}$ ^{a}	$7.51x10^{3} \pm 6.96x10^{3}$ $^{a'd'e'}$	$7.15x10^{2} \pm 1.33x10^{2}$ ^{a}	$2.25x10^{2} \pm 4.52x10^{1}$ ^{a},
CAP	Cape-Point Station, South Africa	Coastal	34°21'26"S 18°29'51''E 230 m asl	11/10/2016 - 05/12/2016	7	$1.89x10^{5} \pm 1.39x10^{5}$ ^{a}	$1.74x10^{3} \pm 1.21x10^{3}$ $^{a'e'}$	$7.64x10^{2} \pm 4.39x10^{2}$ ^{a}	$4.40x10^{2} \pm 1.52x10^{2}$ $^{b' d'}$
STN	Station-Nord, Greenland	Polar	81°34'24"N 16°38'24''E 37 m asl	20/03/2017 - 29/06/2017	13	$7.34x10^{2} \pm 9.22x10^{2}$ ^{b}	$5.24x10^{0} \pm 1.11x10^{1}$ $^{b'}$	$2.17x10^{2} \pm 7.66x10^{1}$ ^{b}	$1.04x10^{2} \pm 5.15x10^{1}$ $^{d'}$
GRE	Grenoble, France	Terrestrial, urban	45°11'38"N 05°45'44''E 210 m asl	30/06/2017 - 14/09/2017	10	$1.20x10^{6} \pm 9.38x10^{5}$ ^{ac}	$5.28x10^{4} \pm 3.61x10^{4}$ $^{d'}$	$7.54x10^{2} \pm 1.05x10^{2}$ ^{a}	$7.44x10^{2} \pm 1.06x10^{2}$ $^{c'}$
PDD	Puy-de-Dôme, France	Terrestrial, continental, mountain peak	45°46'20"N 02°57'57''E 1465 m asl	23/06/2016 - 21/09/2016	12	$3.04x10^{5} \pm 4.65x10^{5}$ ^{a}	$4.82x10^{3} \pm 1.03x10^{4}$ $^{a'e'}$	$5.77x10^{2} \pm 1.41x10^{2}$ ^{a}	$3.70x10^{2} \pm 1.21x10^{2}$ ^{a,b},
PDM	Pic-du-Midi, France	Terrestrial, high-altitude mountain peak	42°56'11"N 00°08'34''E 2876 m asl	20/06/2016 - 04/10/2016	13	$1.51x10^{5} \pm 1.27x10^{5}$ ^{a}	$6.40x10^{3} \pm 5.86x10^{3}$ $^{a'd'e'}$	$5.78x10^{2} \pm 1.48x10^{2}$ ^{a}	$5.14x10^{2} \pm 1.49x10^{2}$ $^{b'}$
CHC	Chacaltaya, Bolivia	Terrestrial, high-altitude mountain peak	16°20'47"S 68°07'44''W 5380 m asl	27/06/2016 - 11/11/2016	16	$1.62x10^{5} \pm 1.35x10^{5}$ ^{a}	$1.05x10^{3} \pm 1.02x10^{3}$ $^{e'}$	$6.63x10^{2} \pm 2.05x10^{2}$ ^{a}	$3.59x10^{2} \pm 7.98x10^{1}$ $^{a, b'}$
NAM	Namco, China	Terrestrial, high-altitude plateau	30°46'44"N 90°59'31''E 4730 m asl	16/05/2017 - 14/09/2017	9	$3.56x10^{6} \pm 3.01x10^{6}$ ^{c}	$4.97x10^{3} \pm 3.44x10^{3}$ $^{a'd'e'}$	$6.67x10^{2} \pm 7.48x10^{1}$ ^{a}	$4.00x10^{2} \pm 9.08x10^{1}$ $^{a, b'}$
STP	Storm-Peak Laboratory, USA	Terrestrial, high-altitude mountain peak	40°27'18"N 106°44'38''E 3220 m asl	11/07/2017 - 04/09/2017	7	$1.63x10^{6} \pm 1.15x10^{6}$ ^{ac}	$8.78x10^{4} \pm 6.49x10^{4}$ $^{a'd'}$	$6.62x10^{2} \pm 1.18x10^{2}$ ^{a}	$3.52x10^{2} \pm 2.21x10^{2}$ $^{a, b'}$

The most abundant bacterial genera overall for all the samples were *Bacillus* (8.23%), *Sphingomonas* (5.62%), *Hymenobacter* (4.32%), *Romboutsia* (2.77%), *Methylobacterium* (2.63%), and *Clostridium* (2.18%) (a list of the highest 50 genera is shown in supplementary **Table S4**), although their relative contribution to each site varied. A heatmap of the relative abundances of the fifty most abundant bacterial genera in each site is represented in **Figure 2**. For example, *Bacillus* averaged 15.63% relative abundance in Puy-de-Dôme samples, but only 0.03% in the marine Amsterdam-Island samples. Average bacterial Chao1 richness estimations varied between 577 +- 141 (Puy-de-Dôme) and 217 +- 76 (Station-Nord). They did not differ significantly ($P > 0.05$) between sites with the exception of Station-Nord which showed the lowest Chao1 value ($P = 4.9 \times 10^{-11}$) (**Table 1**). The most abundant fungal species in the whole dataset were *Pseudotaeniolina globosa* (5.44%), *Cladophialophora proteae* (3.67%), *Ustilago bullata* (3.22%), *Alternaria* sp (2.60%), and *Botryotinia fuckeliana* (*Botrytis cinerea*) (2.47%) (a list of the highest 50 species is shown in supplementary **Table S5**), although their relative contribution to each site varied. A heatmap of the relative abundances of the fifty most abundant fungal species in each site is represented in **Figure 2**. For example, *Pseudotaeniolina globosa* averaged 11.47% relative abundance in Puy-de-Dôme samples, but less than 0.01% in Station-Nord and Storm-Peak samples. Average fungal Chao1 richness estimations were different between the sites ($P = 9.5 \times 10^{-14}$) with the highest value from Grenoble (742 +/- 106) and lowest values from Cape-Point (440 +/-152), Amsterdam-Island (225 +/- 45) and Station-Nord (104 +/- 51). The bacterial and fungal Chao1 richness estimations correlated with the bacterial and fungal concentrations, respectively ($N = 81$, $R = 0.61$, $P = 1.0 \times 10^{-9}$ and $N = 79$, $R = 0.58$, $P = 2.0 \times 10^{-8}$, respectively).

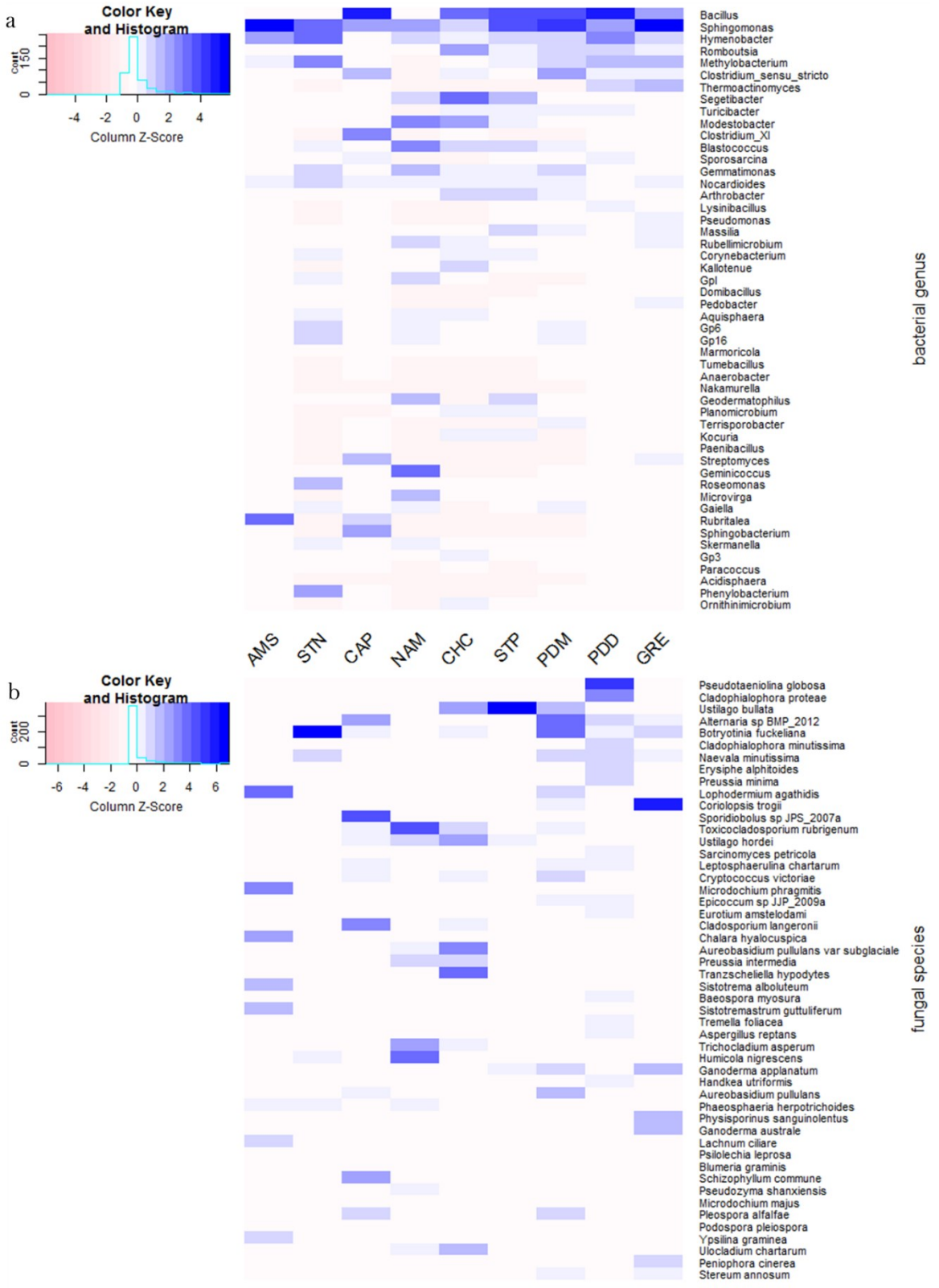


Figure 2: Heatmaps of the relative abundances (the relative abundances are centered and scaled) of the fifty most abundant bacterial genera (a) and fungal species (b) in the dataset. The fifty bacterial genera and fungal species are in order of decreasing relative abundance from top to bottom.

The different temporal samples grouped mainly by their site of origin using the hierarchical cluster analyses based on both their bacterial and fungal community structure (**Fig. 3;** for an expanded view see supplementary **Fig. S1**) and using the bacterial and fungal community profiles averaged over the multiple samples per site (**Fig. 4** and supplementary **Fig. S2**). The sites grouped into three distinct clusters on both the bacterial and fungal-based trees: one cluster included the marine site Amsterdam-Island, the second the polar site Station-Nord and the third one included all the terrestrial and non-polar sites including the coastal site Cape-Point (**Fig. 4** and supplementary **Fig. S2**). The different sites or groups of sites were characterized by bacterial genera and fungal species known to be associated to the environment type of the site in most cases (supplementary **Table S6**). The temporal variability of the composition of the bacterial and fungal communities at each site was different between the sites ($P < 2x10^{-16}$ for both bacterial and fungal communities; **Table 2**). Temporal variations in the bacterial community structure were not correlated to the variations in the fungal community structure ($R = 0.13$, $P = 0.74$). The highest temporal variability of the composition of the bacterial communities was observed in Cape-Point (similarity index of 3.9), Chacaltaya (5.4) and Station-Nord (5.5), while the lowest temporal variability was observed in Amsterdam-Island (17.5), Storm-Peak (17.4) and Namco (20.8). The highest temporal variability of the composition of the fungal communities was observed in Station-Nord (similarity index of 1.6), Puy-de-Dôme (2.95) and Storm-Peak (4.7), while the lowest temporal variability was observed in Amsterdam-Island (9.3), Cape-Point (9.3) and Pic-du-Midi (9.6; **Table 2**). We observed eight "outlier" samples (*i.e*. samples which did not group with the other samples of their respective site) out of eighty-two samples using the hierarchical cluster analysis based on the bacterial community structure of the individual samples (**Fig. 3**). Only one outlier was observed using the hierarchical cluster analysis based on the fungal community structure (**Fig. 3**).

Table 2: Temporal variability of the microbial community structure and meteorological conditions at each site. The similarity index was calculated as following: (1 – Bray-Curtis dissimilarity value averaged per site) / standard deviation. The maximum wind speed (m/s) per site, the variability of the wind direction (degree), relative humidity (%) and temperature (°C) within a week and between the weeks, the number of different landscapes within a perimeter of 50km and the landscape evenness are shown.

Site	Bacterial community structure similarity between weeks (similarity index)	Fungal community structure similarity between weeks (similarity index)	Number of different landscapes within a 50km perimeter	Landscape evenness (Pielou's evenness)	Maximum wind speed (m/s)	Wind direction variability within weeks (degree)	Wind direction variability between weeks (degree)	Relative humidity variability within weeks (%)	Relative humidity variability between weeks (%)	Temperature variability within weeks (°C)	Temperature variability between weeks (°C)
AMS	17.5	9.3	1	1	22	62.3	29.8	10.7	4	1.2	1
CAP	3.9	9.3	4	0.47	26.7	79.5	33.4	11.8	5.6	2.2	1
GRE	12.5	7.6	5	0.65	9.6	94.5	15.3	17.8	5.8	4.9	2.9
STN	5.5	1.6	2	0.80	12.1	81.9	28.6	7.7	5.3	2.7	19.4
PDD	6.81	2.95	4	0,71	31.1	91	34.6	17.6	11.1	3.92	23
PDM	9.6	9.6	4	0.76	32.5	81.8	32.4	20.3	13.4	2.9	3.3
CHC	5.4	7.2	6	0.59	20.1	85.3	47.6	21.6	22.4	1.2	1.2
NAM	20.8	5.3	2	0.66	6.1	29.9	23.2	8.2	6.9	1.2	2.3
STP	17.4	4.7	3	0.84	13	111.1	37	19.5	10.1	3	2.1

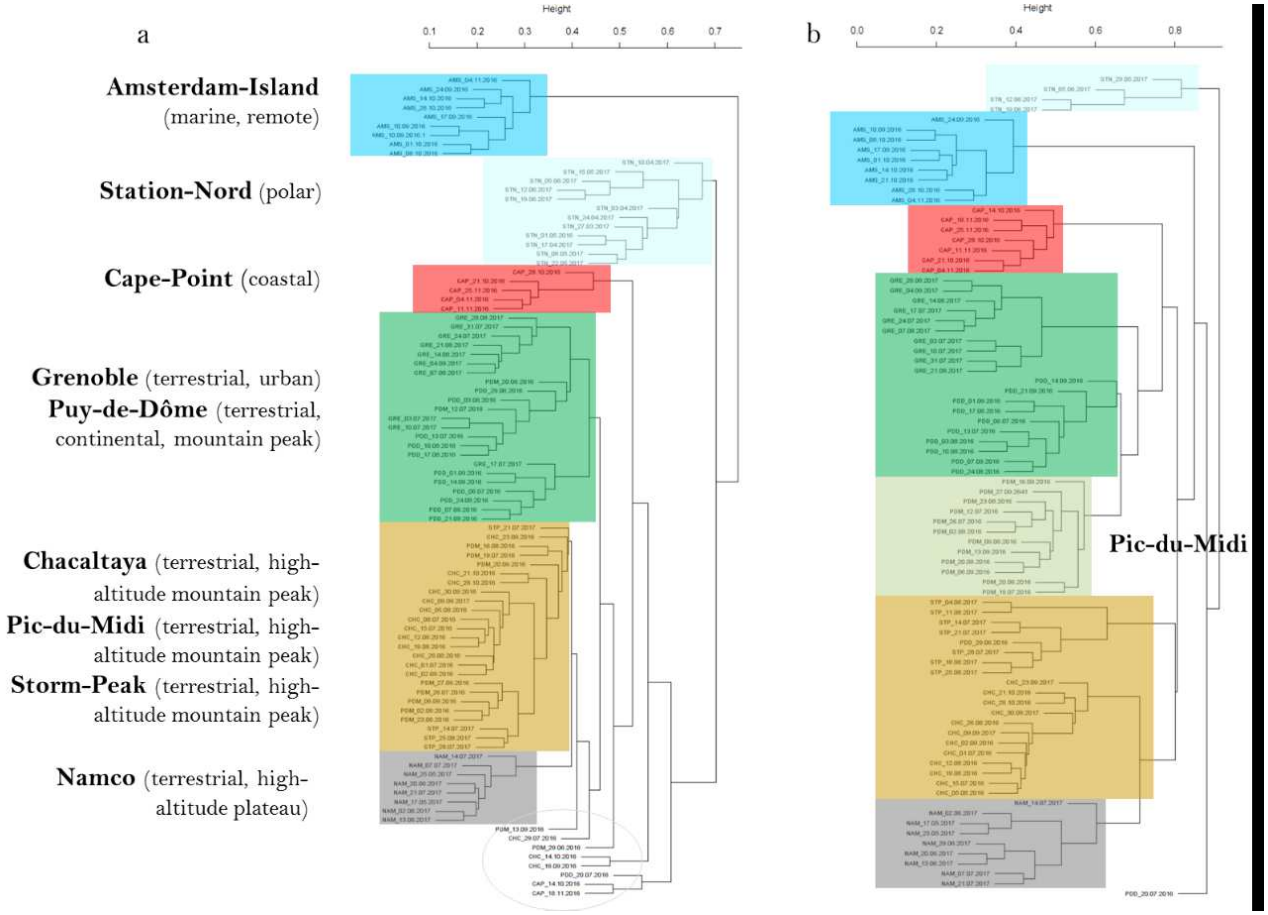


Figure 3: Hierarchical cluster analysis (average method) of the Bray-Curtis dissimilarity matrices based on the V3-V4 region of (a) the 16S rRNA gene (genus level) and (b) ITS region (species level). Colored rectangles correspond to samples of the same site or group of sites stated beside the rectangles. Outlier samples (samples which were outside their expected group) are circled in grey. Samples are named as follows: site_date.of.sampling.

Potential factors driving chemistry and microbiology

Samples from the same site tended to group together based on the PM10 chemistry using hierarchical cluster analysis (supplementary **Fig. S3**). Therefore, we calculated the average chemical profile for each site and redid the cluster analysis (**Fig. 4**). The sites were separated into three main clusters using a hierarchical cluster analysis based on the average chemical profiles (**Fig. 4**): the first cluster (Amsterdam-Island, Cape-Point) was characterized by high relative concentrations of sea salts (Cl, Na), the second (Puy-de-Dôme, Grenoble, Pic-du-Midi, Chacaltaya, Namco, Storm-Peak) by higher concentrations of organic carbon, polyols and sugars, and the third (Station-Nord) by very low relative concentrations of polyols and sugars and lower relative concentrations of sea salts compared to the first cluster (supplementary **Table S7**). The PM10 chemistry was correlated to the bacterial and fungal community

structure averaged over the multiple samples per site (Mantel test $R = 0.79$ $P = 0.007$ and $R = 0.68$ and $P = 0.02$, respectively) (see RDA figures in supplementary **Fig. S4**). The differences between the marine site (Amsterdam-Island), coastal site (Cape-Point), polar site (Station-Nord) and terrestrial sites in terms of PM10 chemistry and microbial communities might create this correlation. Thus, we looked at the correlation using only the terrestrial sites and the correlation between the PM10 chemistry and the bacterial and fungal community structure was weaker ($R = 0.16$, $P = 0.33$ and $R = 0.51$, $P = 0.03$ for bacterial and fungal community structure, respectively).

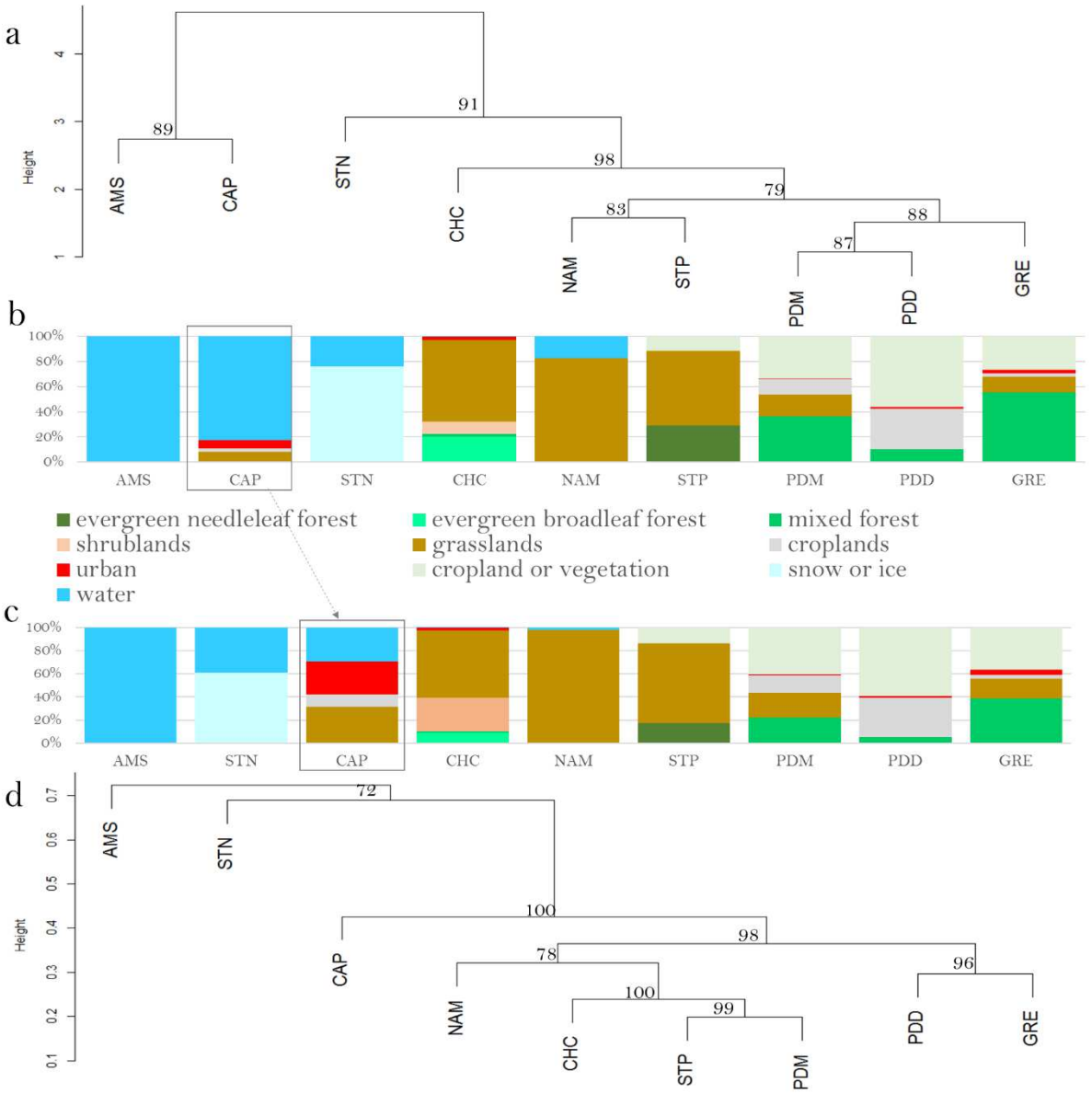


Figure 4: Distribution of the sites based on the different data sets (bacterial community structure, PM10 chemistry, landscapes, relative contributions of the different landscapes in the aerial emission of bacterial cells).
(a) Hierarchical cluster analysis (average method) on the Euclidean distance matrix based on the PM10 chemistry.
(b) Relative surfaces of the different landscapes surrounding the sites (perimeter of 50 km) based on the MODIS land cover approach.
(c) Relative contributions of the different landscapes in the aerial emission of bacterial cells (based on the study of Burrows *et al.*, 2009).
(d) Hierarchical cluster analysis (average method) on the Bray-Curtis dissimilarity matrix based on the bacterial community structure (genus level).
Bootstrap values in percentage are indicated over each node on both cluster analyses.

The different landscapes (e.g., forest, cropland, ocean, etc.) surrounding the sites were identified using the MODIS land cover approach (**Fig. 4**). We weighted the relative surface of the different landscapes by their associated bacterial cell concentration reported by Burrows *et al.* (2009)[40,41] to predict the relative contribution of each landscape to the aerial emission of bacterial cells at each site (**Fig. 4**). Terrestrial sites grouped together based on the high contribution of their terrestrial landscape to the aerial emission of bacterial cells. The average PM10 chemical profile for the different sites was only correlated to the relative surfaces of the different landscapes when considering all the sites (R = 0.77, P = 0.001), but not when considering only the terrestrial sites (R = 0.32, P = 0.1). Both the bacterial and fungal communities were correlated to the landscapes over all the sites (R = 0.80, P = 0.001; R = 0.64, P = 0.003, respectively). However, unlike the chemistry, the bacterial community structure averaged over the multiple samples was correlated to the relative surfaces of the different landscapes for the terrestrial sites (R = 0.65, P = 0.004). The same was observed for the fungal community structure averaged over the multiple temporal samples which was also correlated to the landscapes of the terrestrial sites (R = 0.47, P = 0.05). Although the bacterial community distribution was correlated to the fungal community distribution (R = 0.82, P = 0.006), some differences were observed. The mountain sites (Namco, Chacaltaya, Storm-Peak and Pic-du-Midi) grouped together in the cluster analysis based on their bacterial community structure, while the sites characterized by croplands as one of their surrounding landscapes (Grenoble, Pic-du-Midi, Puy-de-Dôme and Cape-Point) grouped together in the cluster analysis based on the fungal community diversity (**Fig. 4** and supplementary **Fig. S2** for the cluster analysis based on the bacterial and fungal diversity, respectively).

Meteorological characteristics (wind speed and hourly variations in wind direction, relative humidity, and temperature over one week and between weeks) were different between the sites (**Table 2**). Puy-de-Dôme, Pic-du-Midi and Cape-Point showed episodes of high wind speeds (between 26 and 32 m/s), while the wind speed did not reach 10 m/s in Grenoble and Namco. The wind direction varied within the weeks of sampling and also between the multiple weekly samples per site, and these variations were not necessarily correlated. For example, the wind direction in Grenoble varied continually within the weeks (mean variability within weeks: 94.5°), and this variability did not change much between the weeks (mean variability between weeks: 34.6°). The wind direction in Chacaltaya changed less within the weeks (35.3°) than between the weeks (47.6°). The different temporal variability of both the wind speed and

direction led to different wind roses and backward air mass trajectories with different directions and lengths at each site (supplementary **Fig. S5** and **S6**). The relative humidity and temperature also showed a different temporal (within and between weeks) variability depending on the site (**Table 2**). Multiple linear regressions between the level of temporal variability of both the bacterial and fungal community structures at each site and the meteorological and surrounding landscape characteristics (number of different landscapes and landscape evenness) were calculated. For the variation of bacterial composition over time, the regression included the wind direction variability within and between weeks, the temperature variability between weeks and the landscape evenness (*R* squared = 0.93, adjusted *R* squared = 0.82, *P* = 0.06). For the variation of fungal composition over time, the wind speed and temperature variability between weeks were the best parameters (*R* squared = 0.87, adjusted *R* squared = 0.83, *P* = 0.002) (supplementary **Table S8** and **Fig. S7**).

Discussion

Airborne particulate matter (PM) is thought to undergo both short and long-range transport (*i.e*. transport over hundreds of kilometers) depending mainly on particle size[42–44]. Airborne microbial cells might exist both as free cells and in association to PM in the atmosphere[45–47]. This means that the sources, transport duration and deposition processes could be different between PM and microbial cells, especially if their size distribution is different. Furthermore, the potential transformation processes that they might undergo during atmospheric transport are likely to be different. While photochemistry and gas condensation will change the chemical composition and size distribution of PM[48], selective processes associated mainly to UV radiation, desiccation and cold temperatures could lead to the death, sporulation and genetic mutations of microbial cells, although the occurrence of these processes in the atmosphere have been rarely explored[17,49]. Thus, a correlation between airborne microbial community structure and the chemical composition of PM is unlikely due to their expected differential fate during transportation in the atmosphere. So while PM10 chemistry correlated to both the bacterial (*R* = 0.79, *P* = 0.007) and fungal (*R* = 0.68, *P* = 0.02) community structures over all the sites, the correlation was considerably less for the terrestrial sites for bacteria (*R* = 0.16, *P* = 0.33) and fungi (*R* = 0.51, *P* = 0.03). The strong correlation observed when considering all the sites might be more indicative of strong differences between the different ecosystems (marine, polar, coastal and terrestrial) in terms of chemistry and microbial

communities than a similar behavior and/or driving processes of the PM10 and microbial cell distribution.

This weak correlation between PM10 chemistry and airborne bacterial communities over the terrestrial sites has been observed previously[1,18], but certain chemicals might affect specific microbial populations without influencing the overall correlation. For example, methylsulfonic acid (MSA) was correlated to the genus *Acholeplasma* (R = 0.88, P = 1.7x10^{-3}). Certain polyol concentrations were correlated to the relative abundance of specific fungi (*e.g.*, mannitol-arabitol and the species *Zalerion arboricola* R = 0.85, P = 3.9x10^{-3}). Polyols can be major components of fungal biomass and have been proposed as a suitable marker for fungal spores[50]. We currently do not know whether these chemistry-microbial genera/species correlations are due to similar sources for both PM10 and microorganisms or that the chemistry is selecting for these microorganisms during atmospheric transportation.

The similarity of the sites based on the PM10 chemistry globally matched the similarity of the sites based on the characteristics of the landscapes surrounding the sites (**Fig. 4**). However, since the correlation between the chemistry and the characteristics of the surrounding area was not very high when considering only the terrestrial sites (R = 0.32, P = 0.1), distant sources and/or photochemical transformations (formation of secondary aerosols) likely also contribute to the PM10 chemistry distribution. A strong correlation was observed between both airborne bacterial (R = 0.80, P = 0.001 with all the sites; R = 0.65, P = 0.004 with only the terrestrial sites) and fungal community structure (R = 0.64, P = 0.003 with all the sites; R = 0.47, P = 0.05 with only the terrestrial sites) and the landscapes (**Fig. 4** and supplementary **Fig. S2**). This correlation illustrates the important contribution of the local landscapes to the geographical distribution of airborne microorganisms. The landscapes with the most impact on the distribution of atmospheric microorganisms between the sites were ice, water, forest, grassland and cropland. These landscapes might emit distinct communities as compared to the others such as urban landscapes which did not seem to significantly contribute to the observed geographical microbial distribution of the sites. Regional environmental factors, and specifically, a combination of climate and soil characteristics (but no effect from urbanization) have been reported to correlate to the geographical distribution of dust-associated microbial communities[10]. Reports of other smaller-scale studies[2,13,16,18] suggested a strong influence of the local sources in the spatial distribution of airborne microbial communities. Airborne microbial community structure at the coastal site Cape-Point was more similar to that for

terrestrial sites than those from marine or polar sites. This might reflect that terrestrial landscapes contribute more biomass to airborne microbial communities than oceanic landscapes, as was proposed by Burrows *et al.* (2009)[40,41].

Our airborne bacterial concentrations were mainly correlated to the surrounding landscape. We observed the highest atmospheric bacterial concentrations at the grassland sites (Namco, Storm-Peak) and urban and cropland sites (Grenoble, Puy-de-Dôme), followed by the coastal site (Cape-Point), marine site (Amsterdam-Island) and polar site (Station-Nord). We did not observe the highest average concentration of bacteria at the most urban site (Grenoble) as expected based on the study of Burrows *et al.* (2009)[40,41], although Grenoble might not be as industrial as others reported in the literature[1,11]. Instead, the highest average concentration of bacteria was observed at Namco, which is a remote high-altitude and semi-arid grassland site (grassland covers > 80% of the surrounding landscape over 50 km of diameter). Pic-du-Midi (cropland/vegetation) and Chacaltaya (grassland), which are two high-altitude mountain peaks, had relatively low average bacterial concentrations that were comparable to the average bacterial concentration of the coastal site Cape-Point. The elevation and steep slopes of Pic-du-Midi and Chacaltaya could explain the relatively low airborne bacterial concentrations, since they might limit upward migration of aerosolized bacteria from land surfaces to peaks. Another explanation could be that Chacaltaya and Pic-du-Midi aerosols were sampled mainly from the free-troposphere that has fewer microorganisms than the planetary boundary layer. At the Namco site, the calm meteorological conditions (low wind speed, low relative humidity and a low precipitation rate), in association with the high dust content of the surrounding landscape could explain its high bacterial concentrations. The effect of meteorological conditions on airborne microbial concentrations have been investigated previously[51–53], but the wind speed, for example, could lead to either an increase or a decrease in the bacterial concentration depending on its direction, speed and site characteristics.

We hypothesized that meteorological conditions could have an influence on the temporal variation of airborne microbial community structure by affecting the relative inputs of different microbial populations from different surrounding landscapes. Our data showed that the temporal variability of microbial community structure was significantly different between the sites and could be correlated ($R > 0.80$) with wind condition variability, temperature variability and/or landscape characteristics (number of different landscapes and landscape

evenness). Wind conditions and temperature are both known to affect the aerosolization process[54,55]. The wind speed was an important factor explaining fungal variability as different wind speeds might lift up different fungal spore sizes and weights[56,57]. The characteristics of the surrounding landscape (number of different landscapes and landscape evenness) were also important factors determining the temporal variability of airborne microbial communities. Namco and Amsterdam-Island had relatively low temporal variability of airborne bacterial community structure in concordance with their relatively monospecific landscapes, grassland and oceanic, respectively. Although Amsterdam-Island site is characterized by high wind speed, the presence of an oceanic surface over a large perimeter around the site imposed a predominance of homogeneous atmospheric microbial communities. Consequently, wind speed and direction could be of importance when the surrounding landscape is diverse. Although Grenoble has five different surrounding landscapes, the low wind speed (< 9.6 m/s) in Grenoble would lead to a relatively low temporal variability of the airborne bacterial community structure. Conversely, the association of different surrounding landscapes and a medium (between 12 m/s and 13 m/s) or high wind speed (> 22 m/s), as found at the other sites would increase the variability of the composition of bacterial communities over time as a function of the wind speed, wind direction variability and the number of different landscapes. We think that the changes in the composition of airborne bacterial communities might be due to the combined effect of changes in landscapes and local meteorological conditions. Changes in seasons will likely affect landscapes and meteorological conditions and, therefore, influence microbial community composition[13–16,18,58].

Conclusion

This is the first worldwide-scale study investigating airborne microbial communities at diverse sites in terms of latitudinal position, type of ecosystem, surrounding landscapes and local meteorological conditions. We also investigated the atmospheric particulate matter, surrounding landscapes via the MODIS land cover approach, and local meteorology to assess their role in defining the atmospheric microbial communities. We observed that airborne microbial communities were correlated to the surrounding landscapes, although some (minor fraction) of the microbial cells might travel over long distances. While two sites sharing similar surrounding landscapes will likely get a similar airborne microbial profile, different local

meteorological conditions will control the stability of this microbial profile through time. In the context of global warming and land use changes atmospheric microbial (including viruses) communities should be continually monitored around our planet.

Acknowledgements

This program was funded by ANR-15-CE01-0002–INHALE, French Polar Institute IPEV (program 1028 and 399), Région Auvergne-Rhône Alpes and CAMPUS France. The chemical analyses were performed at the IGE AirOSol platform. This work was hosted by the following stations: Chacaltaya, Namco, Puy-de-Dôme, Cape-Point, Pic-du-Midi, Amsterdam-Island, Storm-Peak, Villum RS and we thank I.Jouvie, G.Hallar, I.McCubbin, Benny and Jesper, B.Jensen, A.Nicosia, M.Ribeiro, L.Besaury, L.Bouvier, M.Joly, I.Moreno, M.Rocca, F.Velarde for sampling and station management. We thank our project partners: K.Sellegri, P.Amato, M.Andrade, Q.Zhang, C.Labuschagne and L.Martin, J. Sonke. We thank R.Edwards, J. Schauer and C.Worley for lending their HV sampler. We thank L.Pouilloux for computing assistance and maintenance of the Newton supercalculator.

Data availability

Sequences reported in this paper have been deposited in ftp://ftp-adn.ec-lyon.fr/aerobiology_amplicon_INHALE/. A file has been attached explaining the correspondence between file names and samples.

References

1. Zhen, Q. *et al.* Meteorological factors had more impact on airborne bacterial communities than air pollutants. *Sci. Total Environ.* **601–602**, 703–712 (2017).
2. Šantl-Temkiv, T., Gosewinkel, U., Starnawski, P., Lever, M. & Finster, K. Aeolian dispersal of bacteria in southwest Greenland: their sources, abundance, diversity and physiological states. *FEMS Microbiol. Ecol.* **94**, (2018).
3. Klein, A. M., Bohannan, B. J. M., Jaffe, D. A., Levin, D. A. & Green, J. L. Molecular Evidence for Metabolically Active Bacteria in the Atmosphere. *Front. Microbiol.* **7**, (2016).
4. Michaud, J. M. *et al.* Taxon-specific aerosolization of bacteria and viruses in an experimental ocean-atmosphere mesocosm. *Nat Commun* **9**, 2017 (2018).

5. Griffin, D. W., Gonzalez-Martin, C., Hoose, C. & Smith, D. J. Global-Scale Atmospheric Dispersion of Microorganisms. in *Microbiology of Aerosols* 155–194 (John Wiley & Sons, Ltd, 2017). doi:10.1002/9781119132318.ch2c
6. Smith, D. J., Griffin, D. W. & Jaffe, D. A. The high life: Transport of microbes in the atmosphere. *Eos Trans. AGU* **92**, 249–250 (2011).
7. Smith, D. J. *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* **9**, 1752 (2018).
8. Maki, T. *et al.* Assessment of composition and origin of airborne bacteria in the free troposphere over Japan. *Atmospheric Environment* **74**, 73–82 (2013).
9. Els, N., Baumann-Stanzer, K., Larose, C., Vogel, T. M. & Sattler, B. Beyond the planetary boundary layer: Bacterial and fungal vertical biogeography at Mount Sonnblick, Austria. *Geo: Geography and Environment* **6**, e00069 (2019).
10. Barberán, A. *et al.* Continental-scale distributions of dust-associated bacteria and fungi. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5756–5761 (2015).
11. Innocente, E. *et al.* Influence of seasonality, air mass origin and particulate matter chemical composition on airborne bacterial community structure in the Po Valley, Italy. *Sci. Total Environ.* **593–594**, 677–687 (2017).
12. Uetake, J. *et al.* Seasonal changes of airborne bacterial communities over Tokyo and influence of local meteorology. *bioRxiv* 542001 (2019). doi:10.1101/542001
13. Bowers, R. M., McLetchie, S., Knight, R. & Fierer, N. 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 (2011).
14. Bowers, R. M., McCubbin, I. B., Hallar, A. G. & Fierer, N. Seasonal variability in airborne bacterial communities at a high-elevation site. *Atmospheric Environment* **50**, 41–49 (2012).
15. Bowers, R. M. *et al.* Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. *Environ. Sci. Technol.* **47**, 12097–12106 (2013).
16. Mhuireach, G. Á., Betancourt-Román, C. M., Green, J. L. & Johnson, B. R. Spatiotemporal Controls on the Urban Aerobiome. *Front. Ecol. Evol.* **7**, (2019).
17. Cáliz, J., Triadó-Margarit, X., Camarero, L. & Casamayor, E. O. A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12229–12234 (2018).
18. Gandolfi, I. *et al.* 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 (2015).
19. Fröhlich-Nowoisky, J. *et al.* Biogeography in the air: fungal diversity over land and oceans. *Biogeosciences* **9**, 1125–1136 (2012).
20. Mayol, E. *et al.* Long-range transport of airborne microbes over the global tropical and subtropical ocean. *Nature Communications* **8**, 201 (2017).
21. Fisher, M. C. *et al.* Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**, 186–194 (2012).
22. Brown, J. K. M. & Hovmøller, M. S. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* **297**, 537–541 (2002).
23. Żukiewicz-Sobczak, W. A. The role of fungi in allergic diseases. *Postepy Dermatol Alergol* **30**, 42–45 (2013).

24. Vaïtilingom, M. *et al.* Long-term features of cloud microbiology at the puy de Dôme (France). *Atmospheric Environment* **56**, 88–100 (2012).
25. Yamamoto, N. *et al.* Particle-size distributions and seasonal diversity of allergenic and pathogenic fungi in outdoor air. *ISME J* **6**, 1801–1811 (2012).
26. Fierer, N. *et al.* Short-Term Temporal Variability in Airborne Bacterial and Fungal Populations. *Appl Environ Microbiol* **74**, 200–207 (2008).
27. Dommergue, A. *et al.* Methods to investigate the global atmospheric microbiome. *Front. Microbiol.* **10**, (2019).
28. Carslaw, D. Tools for the Analysis of Air Pollution Data. (2019).
29. Draxler, R. R. & Hess, G. D. An Overview of the HYSPLIT_4 Modelling System for Trajectories, Dispersion, and Deposition. 25
30. Fierer, N., Jackson, J. A., Vilgalys, R. & Jackson, R. B. Assessment of Soil Microbial Community Structure by Use of Taxon-Specific Quantitative PCR Assays. *Appl. Environ. Microbiol.* **71**, 4117–4120 (2005).
31. Chemidlin Prévost-Bouré, N. *et al.* Validation and application of a PCR primer set to quantify fungal communities in the soil environment by real-time quantitative PCR. *PLoS ONE* **6**, e24166 (2011).
32. Louca, S., Doebeli, M. & Parfrey, L. W. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome* **6**, (2018).
33. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology* **73**, 5261–5267 (2007).
34. Oksanen, J. *et al.* Community Ecology Package. (2019).
35. Suzuki, R. & Shimodaira, H. Hierarchical Clustering with P-Values via Multiscale BootstrapResamplin. (2015).
36. Dray, S., Dufour, A.-B. & Thioulouse, J. Analysis of Ecological Data: Exploratory and Euclidean Methodsin Environmental Science. (2018).
37. Harrell, F. E. & Dupont, C. Harrell Miscellaneous - Package 'Hmisc'. (2019).
38. Shannan, S., Collins, K. & Emanuel, W. R. Global mosaics of the standard MODIS land cover type data. (2014).
39. Friedl, M. A. *et al.* Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment* **83**, 287–302 (2002).
40. Burrows, S. M., Elbert, W., Lawrence, M. G. & Pöschl, U. Bacteria in the global atmosphere – Part 1: Review and synthesis of literature data for different ecosystems. *Atmos. Chem. Phys.* **9**, 9263–9280 (2009).
41. Burrows, S. M. *et al.* Bacteria in the global atmosphere – Part 2: Modeling of emissions and transport between different ecosystems. *Atmospheric Chemistry and Physics* **9**, 9281–9297 (2009).
42. Alebic-Juretic, A. & Mifka, B. Secondary Sulfur and Nitrogen Species in PM10 from the Rijeka Bay Area (Croatia). *Bull Environ Contam Toxicol* **98**, 133–140 (2017).
43. Pawar, H. *et al.* Quantifying the contribution of long-range transport to particulate matter (PM) mass loadings at a suburban site in the north-western Indo-Gangetic Plain (NW-IGP). *Atmospheric Chemistry and Physics* **15**, 9501–9520 (2015).
44. Kaneyasu, N. *et al.* Impact of long-range transport of aerosols on the PM2.5 composition at a major metropolitan area in the northern Kyushu area of Japan. *Atmospheric Environment* **97**, 416–425 (2014).

45. Fröhlich-Nowoisky, J. *et al.* Bioaerosols in the Earth system: Climate, health, and ecosystem interactions. *Atmospheric Research* **182**, 346–376 (2016).
46. Després, V. *et al.* Primary biological aerosol particles in the atmosphere: a review. *Tellus B: Chemical and Physical Meteorology* **64**, 15598 (2012).
47. Yamaguchi, N., Ichijo, T., Sakotani, A., Baba, T. & Nasu, M. Global dispersion of bacterial cells on Asian dust. *Scientific Reports* **2**, 525 (2012).
48. Pöschl, U. Atmospheric aerosols: composition, transformation, climate and health effects. *Angew. Chem. Int. Ed. Engl.* **44**, 7520–7540 (2005).
49. DeLeon-Rodriguez, N. Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. (2013). Available at: http://www.pnas.org/content/110/7/2575.full. (Accessed: 25th July 2017)
50. Bauer, H. *et al.* Arabitol and mannitol as tracers for the quantification of airborne fungal spores. *Atmospheric Environment* **42**, 588–593 (2008).
51. Crandall, S. G. & Gilbert, G. S. Meteorological factors associated with abundance of airborne fungal spores over natural vegetation. *Atmospheric Environment* **162**, 87–99 (2017).
52. Dong, L. *et al.* Concentration and size distribution of total airborne microbes in hazy and foggy weather. *Sci. Total Environ.* **541**, 1011–1018 (2016).
53. Jones, A. M. & Harrison, R. M. The effects of meteorological factors on atmospheric bioaerosol concentrations--a review. *Sci. Total Environ.* **326**, 151–180 (2004).
54. Joung, Y. S., Ge, Z. & Buie, C. R. Bioaerosol generation by raindrops on soil. *Nat Commun* **8**, 14668 (2017).
55. Pietsch, R. B., David, R. F., Marr, L. C., Vinatzer, B. & III, D. G. S. Aerosolization of Two Strains (Ice+ and Ice–) of Pseudomonas syringae in a Collison Nebulizer at Different Temperatures. *Aerosol Science and Technology* **49**, 159–166 (2015).
56. Kanaani, H., Hargreaves, M., Ristovski, Z. & Morawska, L. Fungal spore fragmentation as a function of airflow rates and fungal generation methods. *Atmospheric Environment* **43**, 3725–3735 (2009).
57. Kildesø, J. *et al.* Determination of fungal spore release from wet building materials. *Indoor Air* **13**, 148–155 (2003).
58. Gandolfi, I., Bertolini, V., Ambrosini, R., Bestetti, G. & Franzetti, A. Unravelling the bacterial diversity in the atmosphere. *Appl. Microbiol. Biotechnol.* **97**, 4727–4736 (2013).

Author contributions

AD, CL and TMV designed the experiment. RTP, AD, AT and OM conducted the sampling field campaign. RTP did the molecular biology, bioinformatic and statistical analyses. RTP, AD, CL and TMV analyzed the results. CK contributed to the amplicon sequencing. RTP, TM, AD and CL wrote the manuscript. All authors reviewed the manuscript.

Chapter 3: Global airborne microbial communities controlled by surrounding landscapes and wind conditions

Romie Tignat-Perrier[1,2*], Aurélien Dommergue[1], Alban Thollot[1], Christoph Keuschnig[2], Olivier Magand[1], Timothy M. Vogel[2], Catherine Larose[2]

[1]Institut des Géosciences de l'Environnement, Université Grenoble Alpes, CNRS, IRD, Grenoble INP, Grenoble, France

[2]Environmental Microbial Genomics, Laboratoire Ampère, École Centrale de Lyon, Université de Lyon, Écully, France

Supplementary Information

Supplementary Information Text

Material and Methods

Real-Time qPCR analyses

16S rRNA gene qPCR. The bacterial cell concentration was approximated by the number of 16S rRNA gene copies per cubic meter of air. The V3 region of the 16S rRNA gene was amplified using the SensiFast SYBR No-Rox kit (Bioline) and the following primers sequences: Eub 338f 5'-ACTCCTACGGGAGGCAGCAG-3' as the forward primer and Eub 518r 5'-ATTACCGCGGCTGCTGG-3' as the reverse primer[1] on a Rotorgene 3000 machine (Qiagen). The reaction mixture of 20µL contained 10µL of SYBR master mix, 2µL of DNA and RNAse-free water to complete the final 20µL volume. The qPCR 2-steps program consisted of an initial step at 95°C for 2min for enzyme activation, then 35 cycles of 5sec at 95°C and 20sec at 60°C hybridization and elongation. A final step was added to obtain a denaturation from 55°C to 95°C with increments of 1°C.s^{-1}. The amplicon length was around 200 bp. PCR products obtained from DNA from a pure culture of *Escherichia coli* were cloned in a plasmid (pCR™2.1-TOPO® vector, Invitrogen) and used as standard after quantification with the Broad-Range Qubit Fluorometric Quantification (Thermo Fisher Scientific).

18S rRNA gene qPCR. The fungal cell concentration was approximated by the number of *18S rRNA* gene copies per cubic meter of air. The region located at the end of the SSU *18S rRNA* gene, near the ITS1 region, was quantified using the SensiFast SYBR No-Rox kit (Bioline) and the following

primers sequences: FR1 5'-AICCATTCAATCGGTAIT-3' as the forward primer and FF390 5'-CGATAACGAACGAGACCT-3' as the reverse primer [2] on a Rotorgene 3000 machine (Qiagen). The reaction mixture of 20µL contained 10µL of SYBR master mix, 2µL of DNA and RNAse-free water to complete the final 20µL volume. The qPCR 2-steps program consisted of an initial step at 95°C for 5min for enzyme activation, then 35 cycles of 15sec at 95°C and 30sec at 60°C hybridization and elongation. A final step was added to obtain a denaturation from 55°C to 95°C with increments of $1°C.s^{-1}$. The amplicon length was around 390 bp. PCR products obtained from DNA from a soil sample were cloned in a plasmid (pCR™2.1-TOPO® vector, Invitrogen) and used as standard after quantification with the Broad-Range Qubit Fluorometric Quantification (Thermo Fisher Scientific).

MiSeq Illumina amplicon sequencing

Library preparation. The V3-V4 region of the 16S rRNA gene was amplified using the Platinum Taq Polymerase (ThermoFisher Scientific) using the following primer sequences: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3' as the forward primer sequence, and 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3' as the reverse primer sequence. The PCR program used was: 95°C for 3 minutes, 35 cycles of 95°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds, then a final step of 72°C for 5 minutes. The average amplicon size was 550 bp.

To test the reproducibility of the molecular biology analyses, we divided by two (technical duplicates) two samples (AMS_10.09.2016 and PDD_21.12.2016) and did separately the different analyses from the DNA extraction to the 16s rRNA gene sequencing.

The ITS2 region was amplified using the Platinum Taq Polymerase (ThermoFisher Scientific) using the following primer sequences: ILL_5.8S_Fun 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAACTTTYRRCAAYGGATCWCT 3' as the forward primer sequence, and ILL_ITS4_Fun 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGCCTCCGCTTATTGATATGCTTAART 3' as the reverse primer sequence[3]. The PCR program used was: 95°C for 3 minutes, 35 cycles of 95°C for 30 seconds, 56°C for 30 seconds and 72°C for 60 seconds, then a final step of 72°C for 10 minutes. The average amplicon size was 510 bp.

The other steps of the library preparation (amplicon PCR clean-up, index PCR, index PCR clean-up, normalization and pooling) were performed following the Illumina library preparation protocol

("16S Metagenomic Sequencing Library Preparation"). The amplicons were sequenced by a paired-end MiSeq sequencing using the technology V3 (16S rRNA gene) and V2 (ITS) of Illumina with 2 x 250 cycles. The adapter sequences were removed by internal Illumina software at the end of the sequencing. Samples under 6000 raw reads were removed from the dataset with the exception of the arctic Station-Nord samples.

Reads quality filtering and taxonomic annotation. The base quality of the reads 1 and reads 2 was controlled (quality filtering using Q20) using tools of the FASTX-Toolkit software (http://hannonlab.cshl.edu/fastx_toolkit/). PANDAseq[4] was used to assemble the read 1 and the read 2 using the RDP algorithm, a minimum and maximum length of the resulting sequence of 410 bp and 500 bp for 16S rRNA gene sequencing, and 390 bp and 500 bp for ITS region sequencing, a minimum and maximum overlap length of 20 bp and 100 bp. The resulting sequences were stripped out from the primers and annotated at the genus or species level by RDP Classifier[5] using the RDP 16srrna and fungallsu databases for 16S rRNA gene and ITS sequencing, respectively, and an assignment confidence cutoff of 0.6. Sequence analyses were done on the Newton supercalculator of the Ecole Centrale de Lyon. The number of sequences per sample and the percentage of sequences annotated at the genus (bacteria) and species (fungi) level were evaluated using a home-made R script (Supplementary **Table S1**). The sequences annotated as chloroplasts were removed.

Supplementary Information Figures

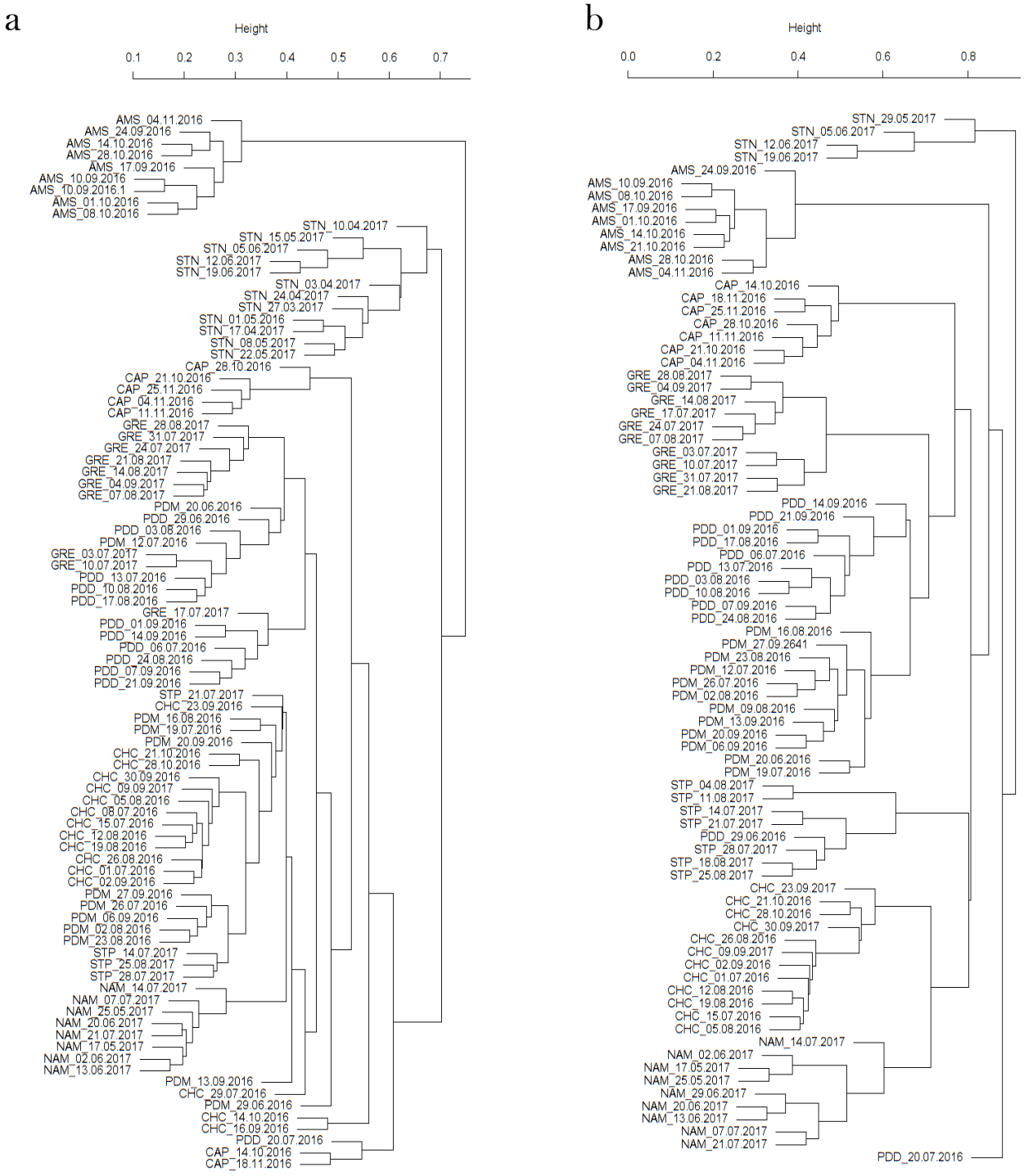


Figure S1: Hierarchical cluster analysis (average method) of the Bray-Curtis dissimilarity matrices based on the V3-V4 region of the 16S rRNA gene (a) and ITS region (b). Samples are named as follows: site_date.of.sampling.

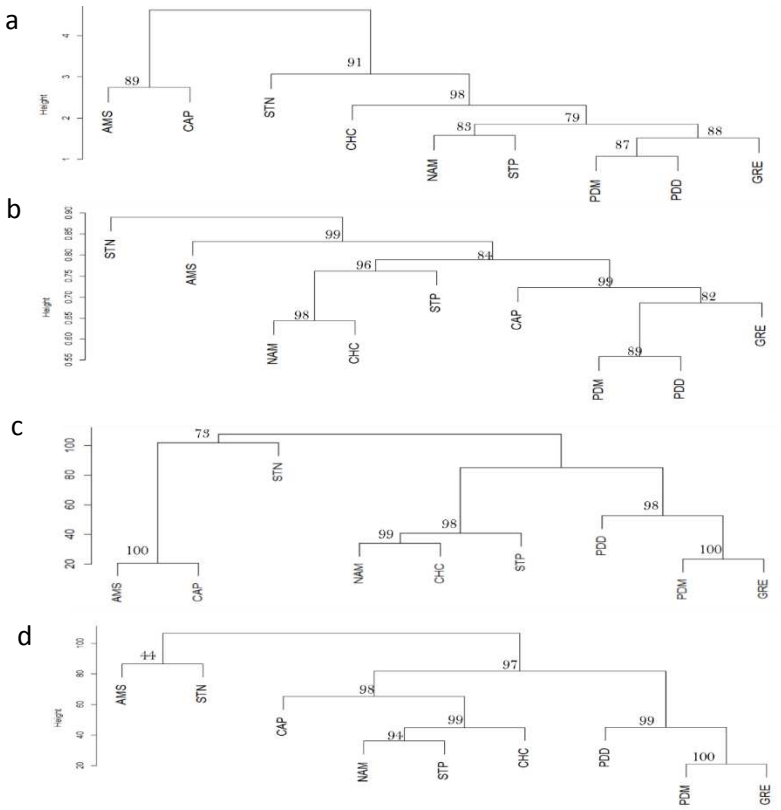


Figure S2: Distribution of the sites based on their (a) average chemical profile, (b) average fungal community profile, (c) relative surfaces of the different landscapes surrounding the sites, (d) relative contributions of the landscapes in the emission of bacterial cells based on the concentration estimates per landscape reported in Burrows *et al.* (2009)[6].

(a) Hierarchical cluster analysis (average method) on the Euclidean dissimilarity matrix calculated on the average composition in chemical species at each site.

(b) Hierarchical cluster analysis (average method) on the Bray-Curtis dissimilarity matrix calculated on the average composition in fungal species at each site.

(c) Hierarchical cluster analysis (average method) on the Euclidean distance matrix calculated on the relative surfaces of the different landscapes

(d)Hierarchical cluster analysis (average method) on the Euclidean distance matrix calculated on the relative contributions of the different landscapes in the aerial emission of bacterial cells. The expected average bacterial concentrations found above each landscape are based on the study of Burrows *et al.* (2009)[6].

Bootstrap values in percentage are indicated over each node.

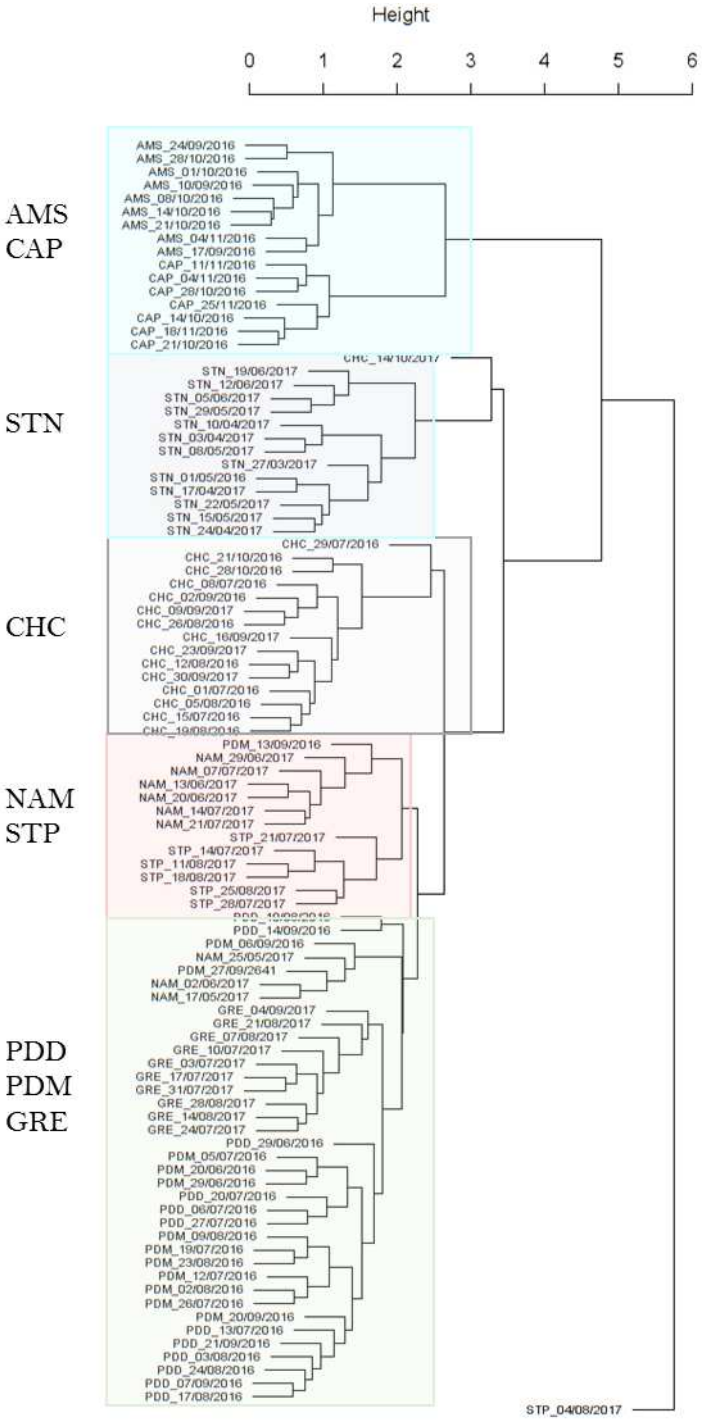


Figure S3: Hierarchical cluster analysis (average method) on the Euclidean distance matrix calculated on the PM10 chemistry data. Color rectangles correspond to the samples of the same site or group of sites stated beside the rectangles.

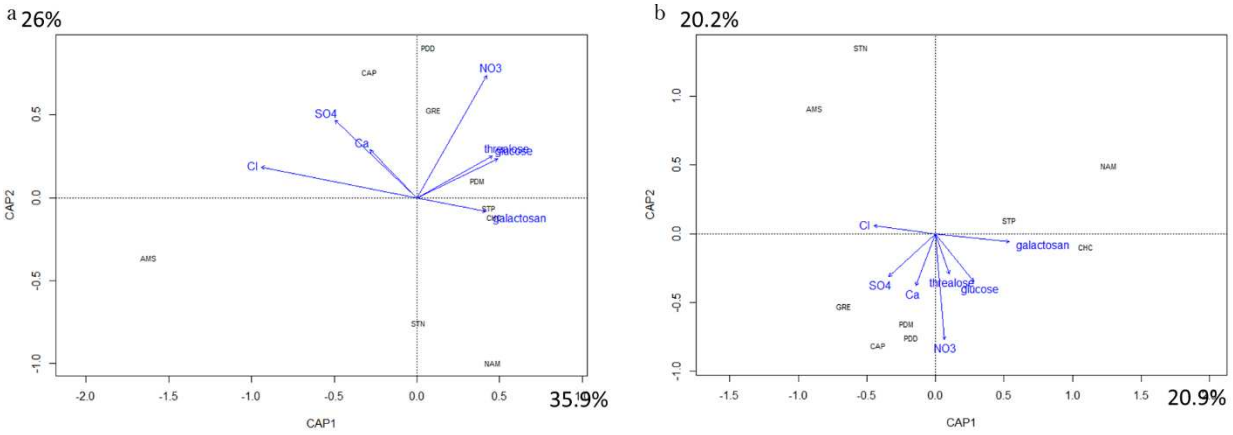


Figure S4: Distance-based RDA analyses. Part of variation in the distribution of both the bacterial (a) and fungal (b) community profile averaged by site explained by several chemical variables (non collinear variables having a VIF > 10) on the first two axes.

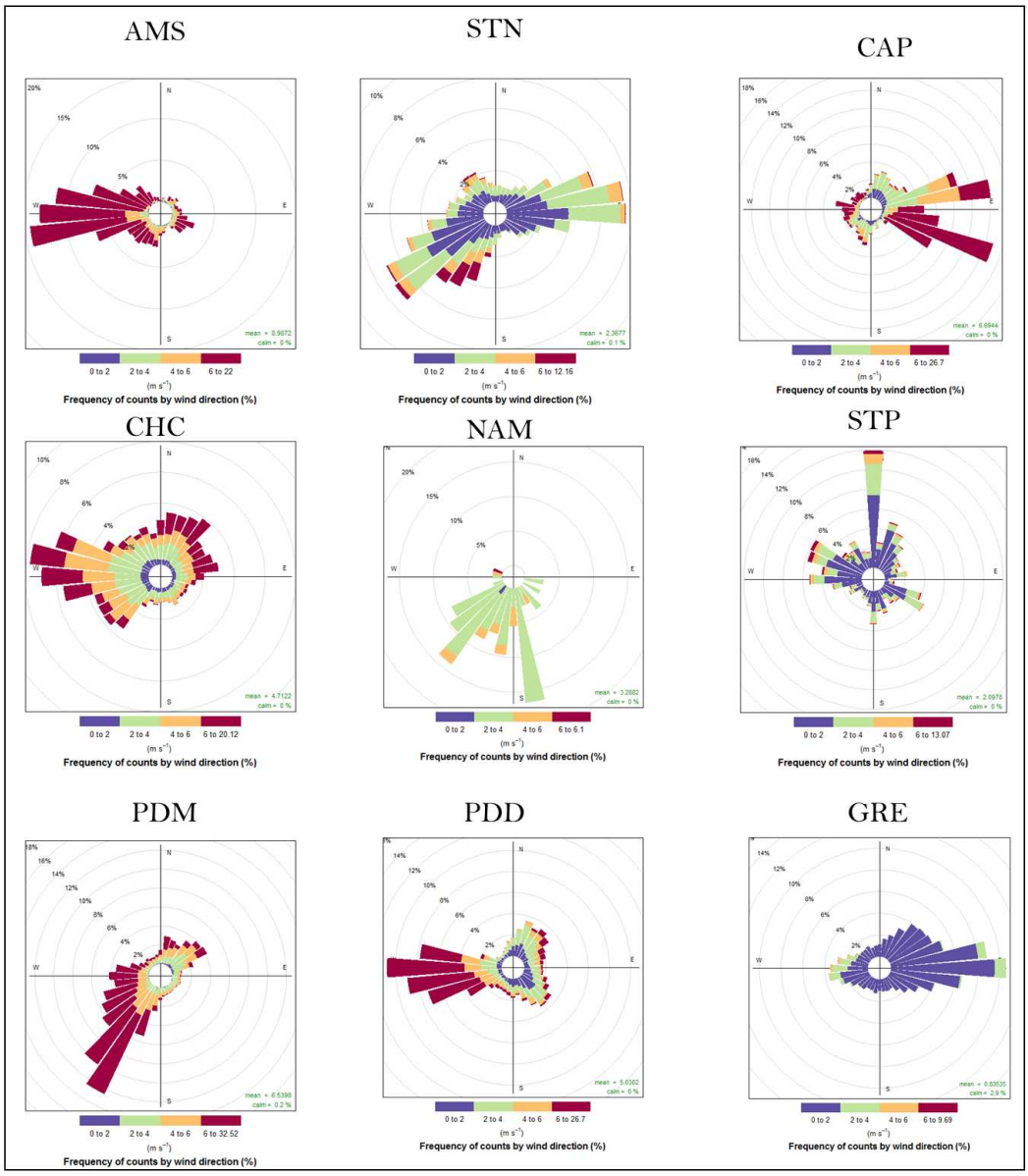


Figure S5: Wind roses covering the sampling time at each site.

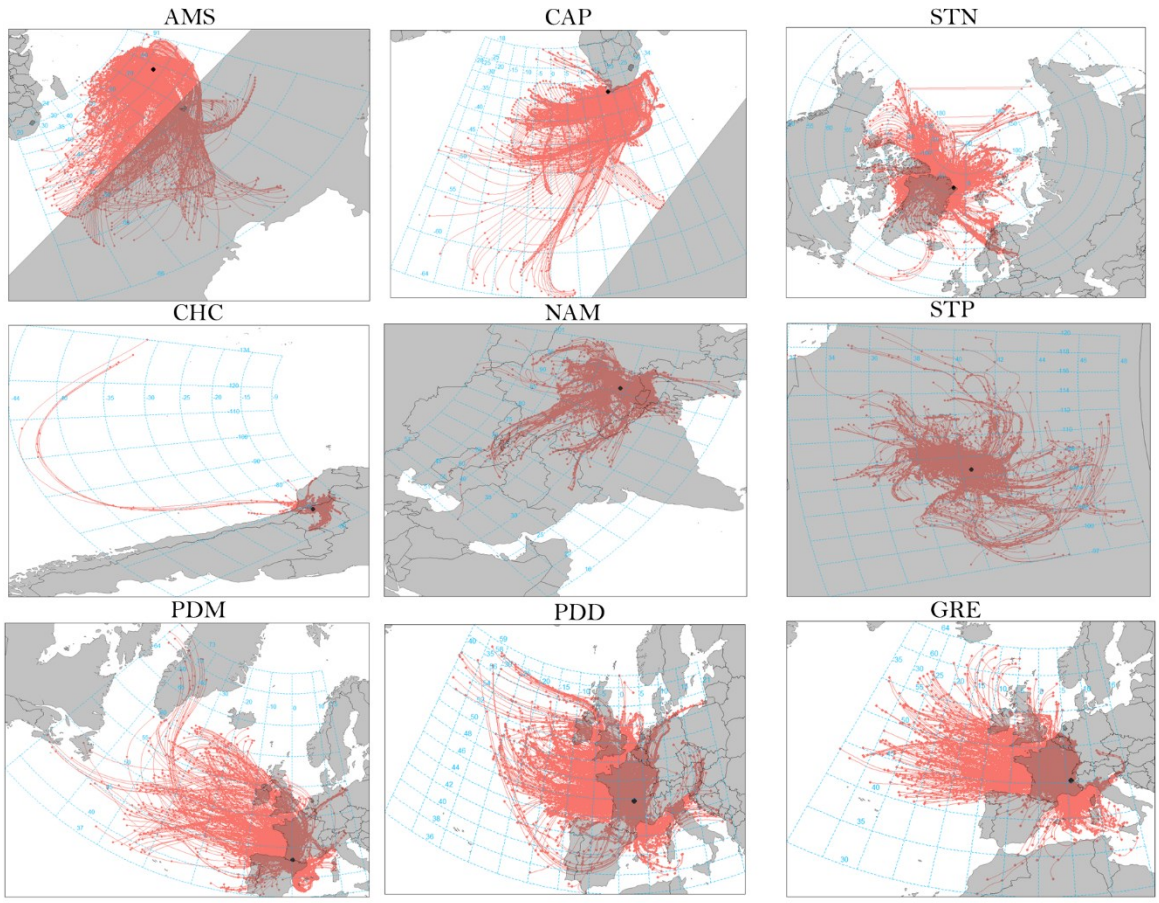


Figure S6: Backward trajectories calculated over 3 days (maximum height from sea level: 1 km) at each site using HYSPLIT.

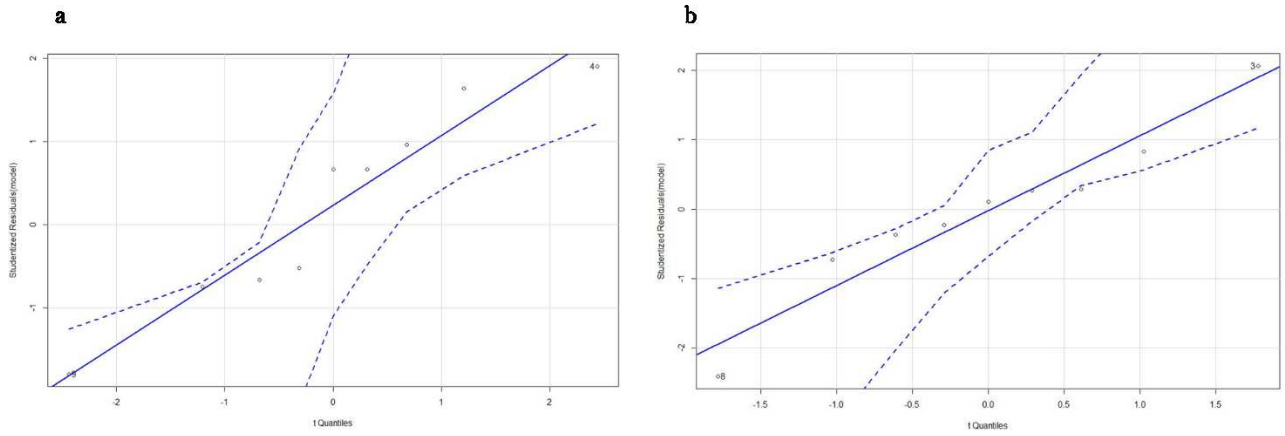


Figure S7: Q-Q plots of the multiple linear regressions. (a) Comparison between the observed temporal variability values based on bacterial communities and the theoretical model (explaining variables: wind direction variability between and within weeks, temperature variability between weeks and landscape evenness); (b) Comparison between the observed temporal variability values based on fungal communities and the theoretical model (explaining variables: maximum wind speed and temperature variability between weeks).

Supplementary Information Tables

Table S1: Sampling starting date, collected volume, raw read number and percentage of annotated sequences using RDP Classifier for both 16s rRNA gene and ITS sequencings for each sample. Blank cells indicate that the sample was removed from the dataset because of low read number (<6000 reads, except arctic Station-Nord samples which might have <6000 reads).

				16s rRNA gene sequencing		ITS sequencing	
Sample name	**Site number**	**Standardized collected air volume (m^3)**	**Sampling starting date (ending date 7 days after, same hour) (month/day/year)**	**Raw read number**	**Percentage of filtered sequences annotated to the genus level using RDP Classifier (%)**	**Raw read number**	**Percentage of filtered sequences annotated to the species level using RDP Classifier (%)**
AMS_10/09/2016	AMS S1	5232	09/10/16	62953	61.5	36823	41.8
AMS_17/09/2016	AMS S2	5266	09/17/16	12741	58.9	37244	42.7
AMS_24/09/2016	AMS S3	4985	09/24/16	78891	50.7	6534	40.8
AMS_01/10/2016	AMS S4	5182	10/01/16	45843	46.9	21370	40.7
AMS_08/10/2016	AMS S5	4449	10/08/16	85337	51.5	31399	33.1
AMS_14/10/2016	AMS S6	5282	10/14/16	58765	49.2	29509	43.6
AMS_21/10/2016	AMS S7	5059	10/21/16			31671	41.5
AMS_28/10/2016	AMS S8	5153	10/28/16	83908	43.4	26055	42.1
AMS_04/11/2016	AMS S9	5035	11/04/16	26816	54.7	20906	42.8
CAP_14/10/2016	CAP S1	4630	10/14/16	29049	64.6	14092	55.3
CAP_21/10/2016	CAP S2	4679	10/21/16	126203	54.9	36185	49.7
CAP_28/10/2016	CAP S3	545	10/28/16	109434	57.8	32413	50.5
CAP_04/11/2016	CAP S4	4695	11/04/16	150604	55.7	35135	58.2
CAP_11/11/2016	CAP S5	4667	11/11/16	108178	63.0	34074	47.4
CAP_18/11/2016	CAP S6	4673	11/18/16	108286	61.3	20212	47.8
CAP_25/11/2016	CAP S7	4668	11/25/16	129262	58.1	22404	41.1
CHC_01/07/2016	CHC S1	1284	07/01/16	90867	57.5	22514	32.4
CHC_09/09/2017	CHC S10	1157	09/09/17	192437	62.4	24981	31.8
CHC_16/09/2016	CHC S11	824	09/16/16	8310	64.8		
CHC_23/09/2016	CHC S12	1148	09/23/16	22575	63.3	12635	24.0
CHC_30/09/2016	CHC S13	1163	09/30/16	189657	58.4	59249	27.2
CHC_14/10/2016	CHC S14	1151	10/14/16	38877	64.1		
CHC_21/10/2016	CHC S15	1154	10/21/16	90828		35825	23.8
CHC_28/10/2016	CHC S16	1220	10/28/16	14370		38071	24.5
CHC_08/07/2016	CHC S2	1284	07/08/16	180240	60.2		
CHC_15/07/2016	CHC S3	1289	07/15/16	89332	61.2	27833	30.5
CHC_29/07/2016	CHC S4	1155	07/29/16	25325	58.6		
CHC_05/08/2016	CHC S5	1157	08/05/16	90048	61.0	24885	32.3
CHC_12/08/2016	CHC S6	1160	08/12/16	163307	60.8	26773	28.7
CHC_19/08/2016	CHC S7	1158	08/19/16	94984	60.4	32254	29.2
CHC_26/08/2016	CHC S8	1158	08/26/16	112875	59.1	40067	32.5
CHC_02/09/2016	CHC S9	1156	09/02/16	48009	56.9	32250	30.0
GRE_03/07/2017	GRE S16	4688	07/03/17	42508	62.5	37691	38.6
GRE_10/07/2017	GRE S17	4717	07/10/17	50677	63.7	40657	24.8
GRE_17/07/2017	GRE S18	4677	07/17/17	64385	61.7	35547	34.4
GRE_24/07/2017	GRE S19	4718	07/24/17	34806	60.9	20997	52.5
GRE_31/07/2017	GRE S20	4665	07/31/17	53090	68.2	27444	49.3
GRE_07/08/2017	GRE S21	4762	08/07/17	37646	60.4	27978	48.2
GRE_14/08/2017	GRE S22	4729	08/14/17	57103	64.5	30992	42.8
GRE_21/08/2017	GRE S23	4707	08/21/17	35248	65.3	19205	54.0
GRE_28/08/2017	GRE S24	4744	08/28/17	49012	62.4	37441	51.4
GRE_04/09/2017	GRE S25	4742	09/04/17	31545	61.8	34026	52.4
NAM_17/05/2017	NAM S1	5511	05/17/17	44324	50.3	16320	34.4
NAM_25/05/2017	NAM S2	5503	05/25/17	41233	54.9	19277	32.7
NAM_02/06/2017	NAM S3	5513	06/02/17	50773	50.2	21024	29.1
NAM_13/06/2017	NAM S4	4218	06/13/17	47395	51.1	23504	31.8
NAM_20/06/2017	NAM S5	5418	06/20/17	19614	57.4	19270	34.0
NAM_29/06/2017	NAM S6	5415	06/29/17			20615	23.8
NAM_07/07/2017	NAM S7	5483	07/07/17	42211	57.2	21025	19.7
NAM_14/07/2017	NAM S8	5413	07/14/17	52028	58.5	12525	23.5
NAM_21/07/2017	NAM S9	5465	07/21/17	48018	55.1	22636	33.6
PDD_29/06/2016	PDD S1	8610	06/29/16	63115	67.9	15813	48.5

PDD_01/09/2016	PDD S10	8610	09/01/16	42068	60.8	25941	17.8
PDD_07/09/2016	PDD S11	8578	09/07/16	45669	53.3	39370	15.7
PDD_14/09/2016	PDD S12	8675	09/14/16	59341	56.7	26614	46.3
PDD_21/09/2016	PDD S13	8463	09/21/16	58601	60.8	42164	12.3
PDD_06/07/2016	PDD S2	8230	07/06/16	51661	42.2	17016	44.1
PDD_13/07/2016	PDD S3	8591	07/13/16	51967	61.9	27368	38.8
PDD_20/07/2016	PDD S4	8769	07/20/16	82471	65.7	31202	54.7
PDD_03/08/2016	PDD S6	8447	08/03/16	41648	67.1	21891	24.3
PDD_10/08/2016	PDD S7	8476	08/10/16	50687	65.3	25038	24.4
PDD_17/08/2016	PDD S8	8552	08/17/16	39687	64.3	24683	20.0
PDD_24/08/2016	PDD S9	8817	08/24/16	57870	66.3	20751	18.5
PDM_20/06/2016	PDM S1	9664	06/20/16	51437	82.0	16290	30.3
PDM_23/08/2016	PDM S10	7956	08/23/16	54036	61.2	27868	18.1
PDM_13/09/2016	PDM S11	7931	09/13/16	85531	65.1	26033	17.0
PDM_20/09/2016	PDM S12	7853	09/20/16	28412	62.5	34125	11.7
PDM_06/09/2016	PDM S13	7867	09/06/16	66574	61.6	20450	15.0
PDM_27/09/2641	PDM S14	7985	09/27/41	29506	60.7	21092	13.9
PDM_29/06/2016	PDM S2	6803	06/29/16	22167	70.0		
PDM_12/07/2016	PDM S4	7550	07/12/16	171749	72.0	27356	18.5
PDM_19/07/2016	PDM S5	8040	07/19/16	38153	64.2	13360	32.2
PDM_26/07/2016	PDM S6	7794	07/26/16	41907	63.9	29215	18.4
PDM_02/08/2016	PDM S7	8103	08/02/16	61278	61.4	9943	19.2
PDM_09/08/2016	PDM S8	7747	09/08/16			8295	25.0
PDM_16/08/2016	PDM S9	8100	08/16/16	54431	67.1	4386	25.9
STN_27/03/2017	STN S1	5153	03/27/17	8937	70.7		
STN_29/05/2017	STN S10	5273	05/29/17	155	20.6	2638	43.2
STN_05/06/2017	STN S11	5333	06/05/17	10926	67.9	3166	25.5
STN_12/06/2017	STN S12	5319	06/12/17	15440	63.7	25940	29.1
STN_19/06/2017	STN S13	5315	06/19/17	28385	61.6	17579	30.8
STN_03/04/2017	STN S2	5130	04/03/17	970	24.5		
STN_10/04/2017	STN S3	5085	04/10/17	8738	66.7		
STN_17/04/2017	STN S4	5128	04/17/17	1315	58.6		
STN_24/04/2017	STN S5	5171	04/24/17	5856	64.4		
STN_01/05/2016	STN S6	5154	05/01/16	2343	64.7		
STN_08/05/2017	STN S7	5186	05/08/17	1744	74.1		
STN_15/05/2017	STN S8	5246	05/15/17	47026	43.1		
STN_22/05/2017	STN S9	5257	05/22/17	1951	64.5		
STP_14/07/2017	STP S1	11213	07/14/17	38363	64.0	17061	17.6
STP_21/07/2017	STP S2	9333	07/21/17	36037	67.0	23616	13.5
STP_28/07/2017	STP S3	5702	07/28/17	43072	63.3	30811	18.4
STP_04/08/2017	STP_S4	5702	04/08/17			13396	30.9
STP_11/08/2017	STP_S5	5702	11/08/17			15977	24.8
STP_18/08/2017	STP_S6	5702	08/18/17			26082	29.5
STP_25/08/2017	STP S7	5702	08/25/17	27320	62.2	33239	34.0

Table S2: Presentation of the different MODIS land covers from Friedl *et al.* (2002)[7].

MODIS land cover	Definition
Evergreen needleleaf forest	Lands dominated by needleleaf woody vegetation with a percent cover > 60% and height exceeding 2 m. Almost all trees remain green all year. Canopy is never without green foliage.
Evergreen broadleaf forest	Lands dominated by broadleaf woody vegetation with a percent cover > 60% and height exceeding 2 m. Almost all trees and shrubs remain green year round. Canopy is never without green foliage.
Mixed forest	Lands dominated by trees with a percent cover > 60% and height exceeding 2 m. Consists of tree communities with interspersed mixtures or mosaics of the other four forest types.
Closed shrublands	Lands with woody vegetation less than 2 m tall and with shrub canopy cover > 60%. The shrub foliage can be either evergreen or deciduous.
Open shrublands	Lands with woody vegetation less than 2 m tall and with shrub canopy cover between 10% and 60%. The shrub foliage can be either evergreen or deciduous.
Grasslands	Lands with herbaceous types of cover. Tree and shrub cover is less than 10%.
Croplands	Lands covered with temporary crops followed by harvest and a bare soil period (e.g., single and multiple cropping systems)
Urban	Land covered by buildings and other man-made structures.
Cropland or vegetation	Lands with a mosaic of croplands, forests, shrubland and grasslands in which no one component comprises more than 60% of the landscape.
Snow or ice	Lands under snow/ice cover throughout the year.
Water	Oceans, seas, lakes, reservoirs, and rivers. Can be either fresh or salt water bodies.

Table S3: Estimation of mean bacterial cell concentration per cubic meter of air in near-surface air above the different landscapes reported in Burrows *et al.* (2009)[6,8]. The « best estimates » have been used.

Earth surface	Bacterial concentration estimation in cells/m^3 of air above the Earth surface, by Burrows *et al.* (2009)
Land ice	$5.0x10^3$
Seas	$1.0x10^4$
Coastal	$7.6x10^4$
Forests	$5.6x10^4$
Crops	$1.1x10^5$
Grasslands	$1.1x10^5$
Urban	$1.2x10^5$
Shrubs	$3.5x10^5$

Table S4: Total abundance (number of annotated sequences), contribution in each site and relative abundance per site of the first fifty most abundant bacterial genera

				Genus contribution in each site (%)									Relative abundance of each genus per site (%)								
Bacterial genus	Bacterial family	Total number of each genus	Genus contribution (%)	AMS_%	CAP_%	CHC_%	GRE_%	NAM_%	PDD_%	PDM_%	STN_%	STP_%	AMS	CAP	CHC	GRE	NAM	PDD	PDM	STN	STP
Total number of sequences per site													220769	419353	810507	232126	172895	1107323	240389	72244	52500
Bacillus	Bacillaceae	273852	8.23	0.03	11.67	13.63	4.07	0.71	63.21	5.44	0.15	1.08	0.03	7.62	4.61	4.8	1.12	15.63	6.2	0.57	5.63
Sphingomonas	Sphingomonadaceae	187036	5.62	6.83	7.86	9.35	15.13	3.75	43.7	9.69	2.12	1.57	5.78	3.5	2.16	12.19	4.06	7.38	7.54	5.48	5.58
Hymenobacter	Cytophagaceae	143844	4.32	3.94	2.29	8.78	4.65	2.83	70.7	3.73	2.42	0.66	2.56	0.79	1.56	2.88	2.35	9.18	2.23	4.82	1.8
Romboutsia	Peptostreptococcaceae	92185	2.77	0.03	3.37	30.62	3.24	1.53	53.89	6.3	0.25	0.77	0.01	0.74	3.48	1.29	0.81	4.49	2.41	0.32	1.36
Methylobacterium	Methylobacteriaceae	87488	2.63	2.1	2.99	9.81	10.16	0.48	64.25	5.57	3.91	0.73	0.83	0.62	1.06	3.83	0.24	5.08	2.03	4.74	1.22
Clostridium_sensu_stricto	Clostridiaceae	72660	2.18	0.07	17.05	17.46	3.78	0.32	46.81	13.08	0.65	0.8	0.02	2.95	1.56	1.18	0.13	3.07	3.95	0.65	1.1
Thermoactinomyces	Thermoactinomycetaceae	62956	1.89	0.1	0.9	1.93	14.45	0.06	80.94	1.33	0.07	0.21	0.03	0.13	0.15	3.92	0.02	4.6	0.35	0.06	0.25
Segetibacter	Chitinophagaceae	50313	1.51	0.03	0.85	80.4	0.33	8.41	3.27	3.45	0.56	2.7	0.01	0.1	4.99	0.07	2.45	0.15	0.72	0.39	2.58
Turicibacter	Erysipelotrichaceae	43272	1.3	0.02	5.08	27.17	3.12	0.67	50.52	11	0.26	2.16	0	0.52	1.45	0.58	0.17	1.97	1.98	0.15	1.78
Modestobacter	Geodermatophilaceae	40533	1.22	0.01	1.1	71.28	1.42	19.63	1.69	2.45	0.41	2	0	0.11	3.56	0.25	4.6	0.06	0.41	0.23	1.55
Clostridium_XI	Peptostreptococcaceae	39826	1.2	0.32	45.31	12.33	0.93	0.41	37.96	2.15	0.01	0.56	0.06	4.3	0.61	0.16	0.1	1.37	0.36	0.01	0.43
Blastococcus	Geodermatophilaceae	37783	1.14	0.17	5.16	45.98	5.17	21.08	5.08	12.27	2.25	2.83	0.03	0.46	2.14	0.84	4.61	0.17	1.93	1.18	2.04
Sporosarcina	Planococcaceae	36476	1.1	0.18	18.53	12.13	2.06	1.44	58.62	5.46	0.61	0.96	0.03	1.61	0.55	0.32	0.3	1.93	0.83	0.31	0.67
Gemmatimonas	Gemmatimonadaceae	33981	1.02	0.05	10.09	41.84	2.81	15.8	7.9	15.01	3.86	2.66	0.01	0.82	1.75	0.41	3.1	0.24	2.12	1.81	1.72
Nocardioides	Nocardioidaceae	32978	0.99	3.13	12.97	35.96	8.92	6.33	15.03	10.29	4.87	2.49	0.47	1.02	1.46	1.27	1.21	0.45	1.41	2.22	1.57
Arthrobacter	Micrococcaceae	31139	0.94	0.2	6.29	55.43	3.16	6.56	11.33	12	1.09	3.94	0.03	0.47	2.13	0.42	1.18	0.32	1.55	0.47	2.34
Lysinibacillus	Planococcaceae	26959	0.81	0	7.12	10.75	7.22	0.76	65.45	7.16	0.02	1.51	0	0.46	0.36	0.84	0.12	1.59	0.8	0.01	0.78
Pseudomonas	Pseudomonadaceae	26068	0.78	2.06	6.5	12.08	10.96	0.85	61.14	4.93	0.33	1.17	0.24	0.4	0.39	1.23	0.13	1.44	0.53	0.12	0.58
Massilia	Oxalobacteraceae	25505	0.77	0.33	3.68	22.55	17.93	2.47	34.12	13	1.76	4.14	0.04	0.22	0.71	1.97	0.36	0.79	1.38	0.62	2.01
Rubellimicrobium	Rhodobacteraceae	25361	0.76	0.02	9	40.2	13.53	14.77	8.98	9.21	2.1	2.19	0	0.54	1.26	1.48	2.17	0.21	0.97	0.74	1.06
Corynebacterium	Corynebacteriaceae	25258	0.76	0.26	8.07	57.17	7.38	6.24	9.16	5.85	3.42	2.46	0.03	0.49	1.78	0.8	0.91	0.21	0.61	1.19	1.18
Kallotenue	Kallotenuaceae	24017	0.72	0.01	1.53	82.41	1.1	6.87	1.74	4.8	0.1	1.44	0	0.09	2.44	0.11	0.95	0.04	0.48	0.03	0.66
GpI	Family I	22539	0.68	0.15	6.15	34.33	6.57	21.62	24.5	1.4	4.32	0.97	0.01	0.33	0.95	0.64	2.82	0.5	0.13	1.35	0.42
Domibacillus	Bacillaceae	21850	0.66	0.01	5.63	6.76	1.37	0.34	72.77	11.64	0.65	0.82	0	0.29	0.18	0.13	0.04	1.44	1.06	0.2	0.34
Pedobacter	Sphingobacteriaceae	21674	0.65	2.69	2.54	13.87	12.28	1.73	56.59	7.01	1.29	1.99	0.26	0.13	0.37	1.15	0.22	1.11	0.63	0.39	0.82
Aquisphaera	Planctomycetaceae	21573	0.65	0.68	4.58	61.16	2.48	11.38	5.99	5.81	5.22	2.7	0.07	0.24	1.63	0.23	1.42	0.12	0.52	1.56	1.11
Gp6	-	20902	0.63	0.12	13	27.06	6.07	14.4	13.34	17.18	6.06	2.77	0.01	0.65	0.7	0.55	1.74	0.25	1.49	1.75	1.1
Gp16	-	20886	0.63	0.08	6.45	36.46	4.12	15.01	12.97	16.48	6.64	1.79	0.01	0.32	0.94	0.37	1.81	0.24	1.43	1.92	0.71
Marmoricola	Nocardioidaceae	20014	0.6	2.21	11.98	44.44	6.86	8.87	12.45	8.27	2.58	2.33	0.2	0.57	1.1	0.59	1.03	0.22	0.69	0.72	0.89
Tumebacillus	Alicyclobacillaceae	19727	0.59	0.01	6.34	5.01	3.53	0.2	78.26	5.51	0.03	1.12	0	0.3	0.12	0.3	0.02	1.39	0.45	0.01	0.42
Anaerobacter	Ruminococcaceae	19298	0.58	0.12	15.49	16.58	3.39	0.4	48.8	14.44	0.09	0.69	0.01	0.71	0.39	0.28	0.05	0.85	1.16	0.02	0.25
Nakamurella	Nakamurellaceae	18866	0.57	0.61	1.41	16.44	2.33	0.82	74.03	3.15	0.22	0.99	0.05	0.06	0.38	0.19	0.09	1.26	0.25	0.06	0.36
Geodermatophilus	Geodermatophilaceae	18516	0.56	0.75	5.28	21.72	5.47	27.64	14.23	15.23	2.93	6.77	0.06	0.23	0.5	0.44	2.96	0.24	1.17	0.75	2.39
Planomicrobium	Planococcaceae	18505	0.56	0	1.1	72.73	1.6	7.71	3.27	9.53	0.02	4.03	0	0.05	1.66	0.13	0.83	0.05	0.73	0.01	1.42

Terrisporobacter	Peptostreptococcaceae	18212	0.55	0.02	8.15	23.01	2.6	0.43	48.91	16.62	0.01	0.24	0	0.35	0.52	0.2	0.05	0.8	1.26	0	0.08
Kocuria	Micrococcaceae	17994	0.54	0.04	6.2	77.68	3.01	1.91	2.36	4.83	0.31	3.66	0	0.27	1.72	0.23	0.2	0.04	0.36	0.08	1.25
Paenibacillus	Paenibacillaceae	17951	0.54	0.03	9.43	13.98	4.95	0.53	65.33	4.92	0.03	0.8	0	0.4	0.31	0.38	0.06	1.06	0.37	0.01	0.27
Streptomyces	Streptomycetaceae	17828	0.54	0.83	61.18	10.07	16.1	2.05	4.52	4.15	0.09	1.01	0.07	2.6	0.22	1.24	0.21	0.07	0.31	0.02	0.34
Geminicoccus	-	17678	0.53	0.53	5.71	23.01	2.73	56.46	2.33	6.68	1.38	1.18	0.04	0.24	0.5	0.21	5.77	0.04	0.49	0.34	0.4
Roseomonas	Acetobacteraceae	17545	0.53	0.56	7.89	38.38	7.31	9.69	14.48	7.19	11.38	3.12	0.04	0.33	0.83	0.55	0.98	0.23	0.52	2.76	1.04
Microvirga	Methylobacteriaceae	17530	0.53	0	6.09	30.29	3.57	35.22	7.42	14.5	0.22	2.69	0	0.25	0.66	0.27	3.57	0.12	1.06	0.05	0.9
Gaiella	Rubrobacteraceae	17391	0.52	0.1	5.52	21.48	6.21	17	17.91	24.18	6.12	1.49	0.01	0.23	0.46	0.47	1.71	0.28	1.75	1.47	0.49
Rubritalea	Rubritaleaceae	16949	0.51	49.96	49.85	0.01	0.03	0	0.11	0	0	0.05	3.84	2.01	0	0	0	0	0	0	0.02
Sphingobacterium	Sphingobacteriaceae	16777	0.5	0.01	91.94	1.17	2.65	0.46	2.66	0.9	0	0.21	0	3.68	0.02	0.19	0.05	0.04	0.06	0	0.07
Skermanella	Rhodospirillaceae	16319	0.49	0	2.87	45.74	3.59	21.48	5.48	14.07	4.81	1.95	0	0.11	0.92	0.25	2.03	0.08	0.96	1.09	0.61
Gp3		16292	0.49	0.06	6.3	70.8	1.52	5.71	5.68	6.21	2.12	1.61	0	0.24	1.42	0.11	0.54	0.08	0.42	0.48	0.5
Paracoccus	Rhodobacteraceae	16202	0.49	2.17	22.55	46.54	9.96	2.04	7.65	6.79	1.08	1.22	0.16	0.87	0.93	0.7	0.19	0.11	0.46	0.24	0.38
Acidisphaera	Acetobacteraceae	15887	0.48	0.22	0.36	13.57	0.57	0.14	84.38	0.6	0.09	0.06	0.02	0.01	0.27	0.04	0.01	1.21	0.04	0.02	0.02
Phenylobacterium	Caulobacteraceae	15797	0.47	0.81	11.92	22.74	3.02	1.27	35.85	6.9	16.43	1.06	0.06	0.45	0.44	0.21	0.12	0.51	0.45	3.59	0.32
Ornithinimicrobium	Intrasporangiaceae	15755	0.47	0.06	12.19	66.07	2.93	3.2	3.81	7.88	0.3	3.56	0	0.46	1.28	0.2	0.29	0.05	0.52	0.07	1.07

Table S5: Total abundance (number of annotated sequences), contribution in each site and relative abundance per site of the first fifty most abundant fungal species

				Genus contribution in each site (%)									Relative abundance of each genus per site (%)								
Fungal species	Fungal family	Total number of each genus	Genus contribution (%)	AMS_%	CAP_%	CHC_%	GRE_%	NAM_%	PDD_%	PDM_%	STN_%	STP_%	AMS	CAP	CHC	GRE	NAM	PDD	PDM	STN	STP
Total number of sequences per site													98965	97585	108560	135543	51672	507133	39325	14920	39006
Pseudotaeniolina globosa	Capnodiales_Incertae sedis	59485	5.44	1.25	0.7	0.22	0.02	0.01	97.77	0.04	0	0	0.75	0.42	0.12	0.01	0.01	11.47	0.07	0	0
Cladophialophora proteae	Herpotrichiellaceae	40075	3.67	0	0	0	0	0	99.99	0	0	0	0	0	0	0	0	7.9	0	0	0
Ustilago bullata	Ustilaginaceae	35197	3.22	0	0.19	16.23	0.03	0.05	14.11	3.48	0	65.9	0	0.07	5.26	0.01	0.04	0.98	3.12	0	59.47
Alternaria sp BMP_2012	Pleosporaceae	28415	2.6	0	16.66	0.01	4.58	0	71.68	7	0.05	0	0	4.85	0	0.96	0	4.02	5.06	0.1	0
Botryotinia fuckeliana	Sclerotiniaceae	26986	2.47	3.6	2.87	5.94	10.96	1.16	48.21	8.56	17.27	1.42	0.98	0.79	1.48	2.18	0.6	2.57	5.88	31.23	0.98
Cladophialophora minutissima	Herpotrichiellaceae	20181	1.85	0	0	0.11	0	0	99.86	0.02	0	0	0	0	0.02	0	0	3.97	0.01	0	0
Naevala minutissima	Dermateaceae	18791	1.72	0.01	0	0.01	6.39	0.45	83.2	4.08	5.1	0.76	0	0	0	0.89	0.16	3.08	1.95	6.43	0.36
Erysiphe alphitoides	Erysiphaceae	15581	1.43	0.01	0.01	0.03	0.06	0.03	99.22	0.63	0	0.03	0	0	0	0.01	0.01	3.05	0.25	0	0.01
Preussia minima	Sporormiaceae	15218	1.39	0	0.09	1.89	1.11	0.04	96.15	0.63	0	0.09	0	0.01	0.27	0.12	0.01	2.89	0.24	0	0.03
Lophodermium agathidis	Rhytismataceae	15133	1.38	92.59	0.27	0.11	0.53	0.5	1.08	4.51	0.01	0.39	14.16	0.04	0.02	0.06	0.15	0.03	1.73	0.01	0.15
Coriolopsis trogii	Polyporaceae	14697	1.35	0	1.07	0.33	94.69	0.22	1.51	2.17	0	0.01	0	0.16	0.04	10.27	0.06	0.04	0.81	0	0.01
Sporidiobolus sp JPS_2007a	Sporidiobolaceae	12897	1.18	7.09	75.21	0.04	3.25	0.01	12.88	1.25	0.01	0.26	0.92	9.94	0	0.31	0	0.33	0.41	0.01	0.09
Toxicocladosporium rubrigenum	Capnodiales_Incertae sedis	12331	1.13	1.1	13.09	29.25	1.18	40.35	10.99	2.72	0.43	0.89	0.14	1.65	3.32	0.11	9.63	0.27	0.85	0.36	0.28
Ustilago hordei	Ustilaginaceae	11219	1.03	0	9.61	59.06	0.01	10.46	11.28	1.05	0	8.53	0	1.1	6.1	0	2.27	0.25	0.3	0	2.45
Sarcinomyces petricola	Pezizomycotina_Incertae sedis	10886	1	0.02	0.12	1.4	0.06	0.45	97.61	0.31	0	0.03	0	0.01	0.14	0.01	0.09	2.1	0.09	0	0.01
Leptosphaerulina chartarum	Didymellaceae	10866	0.99	0.02	16.54	3.43	5.37	0	69.83	4.77	0	0.05	0	1.84	0.34	0.43	0	1.5	1.32	0	0.01
Cryptococcus victoriae	Tremellaceae	10825	0.99	7.64	12.38	9.53	2.84	1.26	57.93	7.5	0	0.92	0.84	1.37	0.95	0.23	0.26	1.24	2.06	0	0.26
Microdochium phragmitis	Hyponectriaceae	10701	0.98	95.65	0.19	0.61	0	0.02	2.51	0.98	0.04	0	10.34	0.02	0.06	0	0	0.05	0.27	0.03	0
Epicoccum sp JJP_2009a	Pleosporales_Incertae sedis	10308	0.94	0.56	0.28	2.58	2.49	0.62	88.32	4.74	0.06	0.34	0.06	0.03	0.25	0.19	0.12	1.8	1.24	0.04	0.09
Eurotium amstelodami	Trichocomaceae	9965	0.91	0	0.09	2.17	2.09	0.17	94.17	1.19	0	0.12	0	0.01	0.2	0.15	0.03	1.85	0.3	0	0.03
Cladosporium langeronii	Mycosphaerellaceae	9854	0.9	0.58	70.13	11.8	9.38	0.73	6.32	0.9	0	0.15	0.06	7.08	1.07	0.68	0.14	0.12	0.23	0	0.04
Chalara hyalocuspica	Pezizomycotina_Incertae sedis	9675	0.89	91.49	0.23	0.06	0.6	0.85	5.75	0.81	0.18	0.04	8.94	0.02	0.01	0.04	0.16	0.11	0.2	0.11	0.01
Aureobasidium pullulans var subglaciale	Dothioraceae	9601	0.88	0.01	0.08	86.67	0.26	5.17	4.67	1.9	0	1.25	0	0.01	7.66	0.02	0.96	0.09	0.46	0	0.31
Preussia intermedia	Sporormiaceae	9413	0.86	0	0	33.72	0	15.97	48.56	0.72	0.01	1.02	0	0	2.92	0	2.91	0.9	0.17	0.01	0.25
Tranzscheliella hypodytes	Ustilaginaceae	9405	0.86	0	0	96.12	0.02	0	1.45	0.52	0	1.89	0	0	8.33	0	0	0.03	0.12	0	0.46
Sistotrema alboluteum	Sistotremataceae	9255	0.85	82.83	0.43	0.02	2.09	0.31	13.31	0.81	0.01	0.18	7.75	0.04	0	0.14	0.06	0.24	0.19	0.01	0.04

Baeospora myosura	Marasmiaceae	8790	0.8	0	0	0.14	0.22	0	99.28	0.36	0	0	0	0	0.01	0.01	0	1.72	0.08	0	0
Sistotremastrum guttuliferum	Hydnodontaceae	8695	0.8	86.6	2.4	0	3.75	0.06	6.36	0.54	0.01	0.28	7.61	0.21	0	0.24	0.01	0.11	0.12	0.01	0.06
Tremella foliacea	Tremellales_Incertae sedis	8602	0.79	0.02	0	0.17	0.07	0	99.7	0.03	0	0	0	0	0.01	0	0	1.69	0.01	0	0
Aspergillus reptans	Trichocomaceae	8571	0.78	0.01	0.04	9.61	1.58	0.92	87.06	0.64	0	0.14	0	0	0.76	0.1	0.15	1.47	0.14	0	0.03
Trichocladium asperum	Chaetomiaceae	8265	0.76	0	0.19	15.33	0.22	38.73	44.49	0.92	0.01	0.11	0	0.02	1.17	0.01	6.19	0.73	0.19	0.01	0.02
Humicola nigrescens	Chaetomiaceae	7754	0.71	0.03	0.08	1.99	0.05	59.8	35.79	0.43	1.65	0.19	0	0.01	0.14	0	8.97	0.55	0.08	0.86	0.04
Ganoderma applanatum	Ganodermataceae	7677	0.7	0.01	0.01	0.01	60.79	0.39	10.52	11.84	0	16.41	0	0	0	3.44	0.06	0.16	2.31	0	3.23
Handkea utriformis	Lycoperdaceae	7057	0.65	0	0	0	0.3	0.68	95.32	3.26	0	0.44	0	0	0	0.02	0.09	1.33	0.58	0	0.08
Aureobasidium pullulans	Dothioraceae	7022	0.64	0.68	18.58	0.93	9.87	0	55.72	14	0	0.21	0.05	1.34	0.06	0.51	0	0.77	2.5	0	0.04
Phaeosphaeria herpotrichoides	Phaeosphaeriaceae	6916	0.63	43.36	0.17	6.84	4.79	7.89	23.15	3.72	2.56	7.52	3.03	0.01	0.44	0.24	1.06	0.32	0.65	1.19	1.33
Physisporinus sanguinolentus	Meripilaceae	6358	0.58	3.38	0.36	0.66	88.42	0.38	5.65	1.15	0	0	0.22	0.02	0.04	4.15	0.05	0.07	0.19	0	0
Ganoderma australe	Ganodermataceae	6288	0.58	0.02	0	0.06	91.78	0	4.04	4.1	0	0	0	0	0	4.26	0	0.05	0.66	0	0
Lachnum ciliare	Hyaloscyphaceae	6247	0.57	82.52	0.05	0	2.75	0	14.5	0.16	0.02	0	5.21	0	0	0.13	0	0.18	0.03	0.01	0
Psilolechia leprosa	Micareaceae	5876	0.54	0	0	0.02	0	0	99.98	0	0	0	0	0	0	0	0	1.16	0	0	0
Blumeria graminis	Erysiphaceae	5794	0.53	0	0.14	0.03	0.05	0	99.69	0.07	0	0.02	0	0.01	0	0	0	1.14	0.01	0	0
Schizophyllum commune	Schizophyllaceae	5635	0.52	0.02	89.55	0.05	7.93	0	1.54	0.91	0	0	0	5.17	0	0.33	0	0.02	0.13	0	0
Pseudozyma shanxiensis	Ustilaginaceae	5235	0.48	4.01	0.08	6.63	0.23	17.94	69.21	1.6	0	0.31	0.21	0	0.32	0.01	1.82	0.71	0.21	0	0.04
Microdochium majus	Hyponectriaceae	5201	0.48	0	0.02	2.75	1.23	0.12	94.81	1.02	0	0.06	0	0	0.13	0.05	0.01	0.97	0.13	0	0.01
Pleospora alfalfae	Pleosporaceae	5177	0.47	1.51	47.94	1.95	0.41	0	32.82	14.93	0	0.44	0.08	2.54	0.09	0.02	0	0.34	1.97	0	0.06
Podospora pleiospora	Lasiosphaeriaceae	4938	0.45	0	0	4.35	0	0.02	95.38	0.22	0	0.02	0	0	0.2	0	0	0.93	0.03	0	0
Ypsilina graminea	Helotiaceae	4785	0.44	95.01	0.02	0.04	0	0.02	4.79	0.1	0.02	0	4.59	0	0	0	0	0.05	0.01	0.01	0
Ulocladium chartarum	Pleosporaceae	4750	0.43	0	0.08	84.91	0.06	8.08	2.27	3.28	0	1.31	0	0	3.71	0	0.74	0.02	0.4	0	0.16
Peniophora cinerea	Peniophoraceae	4564	0.42	0.2	5	0.31	85.78	0	7.32	1.29	0	0.11	0.01	0.23	0.01	2.89	0	0.07	0.15	0	0.01
Stereum annosum	Stereaceae	4360	0.4	0	2.41	0	38.26	0	52	7.04	0	0.3	0	0.11	0	1.23	0	0.45	0.78	0	0.03

Table S6: Bacterial genera and fungal species characterizing the different sites or groups of sites identified using hierarchical cluster analyses based on both bacterial and fungal community structures. We indicated potential associated environmental sources based on articles showing the presence (and mainly isolation) of the bacterial genus/fungal species in these potential sources.

Site or group of sites	Bacterial genera	Potential source	References	Site or group of sites	Fungal species	Potential source	References
AMS	*Aquimarina*	Sea water	[9,10]	AMS	*Lophodermium agathidis*	Plants	[11]
	Ktedonobacter	Soil	[12]		*Microdochium phragmitis*	Aquatic, plants	[13]
	Cocleimonas	Sea water, sand snail	[14,15]		*Chalara hyalocuspica*	Plants	[16]
	Rubritalea	Sea water, marine chordates	[17–19]		*Sistotremastrum guttuliferum*	Oceanic islands	[20]
STN	*Methylobacterium*	Soil, plants	[21,22]	STN	*Botryotinia fuckeliana (Botrytis cinerea)*	Plants	[23]
	Sediminibacterium	Different environments (soil, sediment, reservoir)	[24,25]		*Bullera variabilis*	Plants	[26]
	Phenylobacterium	Soil, rhizosphere, sludge	[27,28]		*Cryptococcus sp AL_V*		
	Bradyrhizobium	Plants	[29]		*Penicillium corylophilum*	Different environments (buildings)	[30]
CAP	*Sphingobacterium*	Soil, compost	[31,32]	CAP	*Cladosporium langeronii*	Human skin, insects	[33]
	Coraliomargarita	Sea water	[17]		*Sporidiobolus sp JPS_2007a*		
	Clostridium XI	Feces-associated	[34]		*Entyloma dahlia*	Plants	
	Streptomyces	Soil	[35,36]		*Cladosporium salinae*	Saline environments	[37]
GRE/PDD	*Sphingomonas*	Different environments	[38,39]	GRE/PDD/PDM	*Alternaria sp BMV_2012*	Plants	[40]
					Coriolopsis troggi (trametes trogii)	Soil	
CHC/PDM/STP	*Segetibacter*	Soil	[41]	CHC/STP/NAM	*Ustilago bullata*	Plants	[42]

NAM	*Clostridium sensu-stricto*	Different environments		*Ustilago hordei*	Plants	[43]
	Modestobacter	Surfaces, extreme conditions	[44–46]	*Toxicocladosporium rubrigenum*	Plants	[47]
	Blastococcus	Different environments (sandstone, sea, soil pland, snow)	[48]			
	Geminicoccus	Biofilter of a marine aquaculture system	[49]			

Table S7: Average concentration (and standard deviation) of the chemicals per site in ng/m^3 of air. Part of these data was published in Dommergue *et al.* (2019)[50] (Chapter 2).

Site	Organic carbon	Elemental carbon	MSA	Cl	NO3	SO4	Oxalate	Na	NH4	K	Mg	Ca	Inositol	Glycerol	Erythriol	Xylitol	Mannitol + arabitol	Sorbitol	Threalose	Levoglucosan + mannosan	Galactosan	Rhamnose	Glucose
AMS	1.96 ± 0.09	0	1.2 ± 0.19	3.99 ± 0.16	1.66 ± 0.16	3.13 ± 0.16	1.05 ± 0.11	3.73 ± 0.15	0.14 ± 0.27	2.27 ± 0.17	2.66 ± 0.16	2.35 ± 0.2	< 0.007	0.08 ± 0.17	0.02 ± 0.02	0 ± 0.01	0.21 ± 0.16	< 0.007	0.1 ± 0.08	< 0.01	< 0.007	< 0.02	0.34 ± 0.22
CAP	2.4 ± 0.14	0.01 ± 0.01	1.69 ± 0.11	3.71 ± 0.17	2.8 ± 0.2	3.2 ± 0.11	1.72 ± 0.2	3.53 ± 0.11	2.05 ± 0.23	2.11 ± 0.08	2.46 ± 0.11	2.26 ± 0.24	< 0.008	0.56 ± 0.16	0.16 ± 0.11	0.11 ± 0.09	0.22 ± 0.16	0.28 ± 0.12	< 0.02	0.24 ± 0.23	0.01 ± 0.01	< 0.02	0.21 ± 0.15
CHC	3.04 ± 0.14	0.04 ± 0.01	0.86 ± 0.09	1.06 ± 0.39	2.1 ± 0.6	2.99 ± 0.31	1.85 ± 0.16	1.47 ± 0.45	2.52 ± 0.24	1.77 ± 0.16	0.85 ± 0.25	1.79 ± 0.21	0.14 ± 0.1	0.57 ± 0.36	0.33 ± 0.11	< 0.07	0.55 ± 0.09	< 0.06	< 0.035	1.63 ± 0.3	0.62 ± 0.32	0.01 ± 0.02	0.6 ± 0.14
GRE	3.55 ± 0.09	0.22 ± 0.05	1.54 ± 0.16	1.03 ± 0.36	2.51 ± 0.22	3.15 ± 0.14	2.29 ± 0.19	2.16 ± 0.27	2.16 ± 0.29	2.02 ± 0.13	1.42 ± 0.19	2.45 ± 0.22	0.24 ± 0.07	< 0.30	0.51 ± 0.12	0.13 ± 0.08	1.58 ± 0.22	0.57 ± 0.14	1.07 ± 0.22	0.97 ± 0.24	0.02 ± 0.03	< 0.02	1.3 ± 0.23
NAM	2.9 ± 0.15	0.04 ± 0.01	0.55 ± 0.11	0.87 ± 0.39	2.33 ± 0.45	2.81 ± 0.29	2.03 ± 0.13	1.81 ± 0.34	1.89 ± 0.3	1.28 ± 0.23	1.19 ± 0.26	2.31 ± 0.41	0.09 ± 0.06	< 0.96	0.13 ± 0.05	< 0.05	0.51 ± 0.23	0.04 ± 0.11	0.61 ± 0.15	0.69 ± 0.09	0.06 ± 0.02	< 0.05	1.03 ± 0.3
PDD	3.12 ± 0.11	0.05 ± 0.01	1.4 ± 0.31	1.26 ± 0.63	2.61 ± 0.34	2.98 ± 0.11	2.16 ± 0.12	2.23 ± 0.21	2.42 ± 0.25	1.62 ± 0.13	1.42 ± 0.18	1.97 ± 0.25	0.2 ± 0.16	< 0.16	0.39 ± 0.13	0.39 ± 0.16	1.22 ± 0.21	0.88 ± 0.19	0.78 ± 0.28	0.62 ± 0.21	0.05 ± 0.07	< 0.01	0.98 ± 0.2
PDM	3.06 ± 0.21	0.03 ± 0.02	0.93 ± 0.34	1.05 ± 0.32	2.74 ± 0.3	2.92 ± 0.23	2.15 ± 0.22	1.97 ± 0.2	2.3 ± 0.34	1.48 ± 0.27	1.37 ± 0.23	2.38 ± 0.3	0.19 ± 0.11	0.11 ± 0.28	0.36 ± 0.18	0.18 ± 0.14	0.72 ± 0.24	0.93 ± 0.25	0.72 ± 0.25	0.4 ± 0.24	0.03 ± 0.05	< 0.04	0.64 ± 0.25
STN	2.15 ± 0.23	0.01 ± 0	0.94 ± 0.23	1.28 ± 0.74	1.51 ± 0.39	2.67 ± 0.35	0.92 ± 0.42	1.72 ± 0.5	2.01 ± 0.24	0.79 ± 0.32	0.94 ± 0.42	1.38 ± 0.38	0.02 ± 0.01	0.1 ± 0.14	0 ± 0.01	0.01 ± 0.02	0.02 ± 0.04	< 0.006	< 0.008	0.02 ± 0.05	< 0.006	< 0.02	0.04 ± 0.06
STP	3.08 ± 0.5	0.02 ± 0.01	0.85 ± 0.33	0.58 ± 0.34	2.12 ± 0.58	2.3 ± 1.03	2.02 ± 0.48	1.22 ± 0.53	1.98 ± 0.58	1.36 ± 0.57	1.06 ± 0.45	1.83 ± 0.59	0.16 ± 0.1	0.33 ± 0.27	0.59 ± 0.3	0.12 ± 0.1	1.3 ± 0.6	< 0.05	1.31 ± 0.6	0.74 ± 0.58	0.16 ± 0.18	< 0.05	0.97 ± 0.44

Table S8: Multiple linear regression results.

Response variable	Explanatory variables	*R* squared	Adjusted *R* squared	Pvalue (*P*)
Airborne bacterial community temporal variability (similarity index)	-Wind direction variability within weeks -Wind direction variability between weeks -Landscape evenness (Pielou's evenness) -Temperature variability between weeks	0.93	0.82	0.06
Airborne fungal community temporal variability (similarity index)	-Maximum wind speed -Temperature variability between weeks	0.87	0.83	0.002

References

1. Fierer, N., Jackson, J. A., Vilgalys, R. & Jackson, R. B. Assessment of Soil Microbial Community Structure by Use of Taxon-Specific Quantitative PCR Assays. *Appl. Environ. Microbiol.* **71**, 4117–4120 (2005).
2. Chemidlin Prévost-Bouré, N. *et al.* Validation and application of a PCR primer set to quantify fungal communities in the soil environment by real-time quantitative PCR. *PLoS ONE* **6**, e24166 (2011).
3. Taylor, D. L. *et al.* Accurate Estimation of Fungal Diversity and Abundance through Improved Lineage-Specific Primers Optimized for Illumina Amplicon Sequencing. *Appl. Environ. Microbiol.* **82**, 7217–7226 (2016).
4. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**, 31 (2012).
5. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology* **73**, 5261–5267 (2007).
6. Burrows, S. M. *et al.* Bacteria in the global atmosphere – Part 2: Modeling of emissions and transport between different ecosystems. *Atmospheric Chemistry and Physics* **9**, 9281–9297 (2009).
7. Friedl, M. A. *et al.* Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment* **83**, 287–302 (2002).
8. Burrows, S. M., Elbert, W., Lawrence, M. G. & Pöschl, U. Bacteria in the global atmosphere – Part 1: Review and synthesis of literature data for different ecosystems. *Atmos. Chem. Phys.* **9**, 9263–9280 (2009).
9. Nedashkovskaya, O. I. *et al.* Description of Aquimarina muelleri gen. nov., sp. nov., and proposal of the reclassification of [Cytophaga] latercula Lewin 1969 as Stanierella latercula gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* **55**, 225–229 (2005).
10. Yi, H. & Chun, J. Aquimarina addita sp. nov., isolated from seawater. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* **61**, 2445–2449 (2011).
11. Ortiz-García, S. *et al.* Phylogenetics of Lophodermium from Pine. *Mycologia* **95**, 846–59 (2003).

12. Wang, H. *et al.* Distribution and diversity of bacterial communities and sulphate-reducing bacteria in a paddy soil irrigated with acid mine drainage. *J. Appl. Microbiol.* **121**, 196–206 (2016).
13. Liu, Y., Zachow, C., Raaijmakers, J. M. & De Bruijn, I. Elucidating the Diversity of Aquatic Microdochium and Trichoderma Species and Their Activity against the Fish Pathogen Saprolegnia diclina. *International Journal of Molecular Sciences* **17**, 140 (2016).
14. Tanaka, N., Romanenko, L. A., Iino, T., Frolova, G. M. & Mikhailov, V. V. Cocleimonas flava gen. nov., sp. nov., a gammaproteobacterium isolated from sand snail (Umbonium costatum). *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* **61**, 412–416 (2011).
15. Zeng, Y., Yu, Y., Qiao, Z.-Y., Jin, H.-Y. & Liu, Q. Diversity of bacterioplankton in coastal seawaters of Fildes Peninsula, King George Island, Antarctica. *Archives of microbiology* **196**, (2014).
16. Koukol, O. New species of Chalara occupying coniferous needles. *Fungal Diversity* **49**, 75 (2011).
17. Yoon, J. *et al.* Cerasicoccus arenae gen. nov., sp. nov., a carotenoid-producing marine representative of the family Puniceicoccaceae within the phylum 'Verrucomicrobia', isolated from marine sand. *Int. J. Syst. Evol. Microbiol.* **57**, 2067–2072 (2007).
18. Kasai, H. *et al.* Rubritalea squalenifaciens sp. nov., a squalene-producing marine bacterium belonging to subdivision 1 of the phylum 'Verrucomicrobia'. *Int. J. Syst. Evol. Microbiol.* **57**, 1630–1634 (2007).
19. Scheuermayer, M. Rubritalea marina gen. nov., sp. nov., a marine representative of the phylum 'Verrucomicrobia', isolated from a sponge (Porifera). *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* **56**, 2119–2124 (2006).
20. Telleria, M. *et al.* Sistotremastrum guttuliferum: A new species from the Macaronesian islands. *Mycological Progress* (2013). doi:10.1007/s11557-012-0876-0
21. Knief, C., Frances, L. & Vorholt, J. A. Competitiveness of diverse Methylobacterium strains in the phyllosphere of Arabidopsis thaliana and identification of representative models, including M. extorquens PA1. *Microb. Ecol.* **60**, 440–452 (2010).
22. Renier, A. *et al.* Nodulation of Crotalaria podocarpa DC. by Methylobacterium nodulans displays very unusual features. *J Exp Bot* **62**, 3693–3697 (2011).
23. Debieu, D., Bach, J., Hugon, M., Malosse, C. & Leroux, P. The hydroxyanilide fenhexamid, a new sterol biosynthesis inhibitor fungicide efficient against the plant pathogenic fungus Botryotinia fuckeliana (Botrytis cinerea). *Pest Management Science* **57**, 1060–1067 (2001).
24. Kang, H., Kim, H., Lee, B.-I., Joung, Y. & Joh, K. Sediminibacterium goheungense sp. nov., isolated from a freshwater reservoir. *Int. J. Syst. Evol. Microbiol.* **64**, 1328–1333 (2014).
25. Qu, J.-H. & Yuan, H.-L. Sediminibacterium salmoneum gen. nov., sp. nov., a member of the phylum Bacteroidetes isolated from sediment of a eutrophic reservoir. *Int. J. Syst. Evol. Microbiol.* **58**, 2191–2194 (2008).
26. Nakase, T. *et al.* Bullera begoniae sp. nov. and Bullera setariae sp. nov., two new species of ballistoconidium-forming yeasts in the Bullera variabilis (Bulleribasidium) cluster isolated from plants in Taiwan. *Mycoscience* **45**, 287–294 (2004).
27. Farh, M. E.-A., Kim, Y.-J., Singh, P., Hoang, V.-A. & Yang, D.-C. Phenylobacterium panacis sp. nov., isolated from the rhizosphere of rusty mountain ginseng. *Int. J. Syst. Evol. Microbiol.* **66**, 2691–2696 (2016).
28. Kanso, S. & Patel, B. K. C. Phenylobacterium lituiforme sp. nov., a moderately thermophilic bacterium from a subsurface aquifer, and emended description of the genus Phenylobacterium. *Int. J. Syst. Evol. Microbiol.* **54**, 2141–2146 (2004).
29. Nguyen, H. D. T., Cloutier, S. & Bromfield, E. S. P. Complete Genome Sequence of Bradyrhizobium ottawaense OO99T, an Efficient Nitrogen-Fixing Symbiont of Soybean. *Microbiol Resour Announc* **7**, e01477-18 (2018).
30. McMullin, D. R., Nsiama, T. K. & Miller, J. D. Secondary metabolites from Penicillium corylophilum isolated from damp buildings. *Mycologia* **106**, 621–628 (2014).

31. Yoo, S.-H. *et al.* Sphingobacterium composti sp. nov., isolated from cotton-waste composts. *Int. J. Syst. Evol. Microbiol.* **57**, 1590–1593 (2007).
32. Wei, W., Zhou, Y., Wang, X., Huang, X. & Lai, R. Sphingobacterium anhuiense sp. nov., isolated from forest soil. *Int. J. Syst. Evol. Microbiol.* **58**, 2098–2101 (2008).
33. Sun, T. *et al.* A Lethal Fungus Infects the Chinese White Wax Scale Insect and Causes Dramatic Changes in the Host Microbiota. *Scientific Reports* **8**, (2018).
34. Kubasova, T. *et al.* Effects of host genetics and environmental conditions on fecal microbiota composition of pigs. *PLOS ONE* **13**, e0201901 (2018).
35. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **417**, 141–147 (2002).
36. Higginbotham, S. J. & Murphy, C. D. Identification and characterisation of a Streptomyces sp. isolate exhibiting activity against methicillin-resistant Staphylococcus aureus. *Microbiological Research* **165**, 82–86 (2010).
37. Zalar, P. *et al.* Phylogeny and ecology of the ubiquitous saprobe Cladosporium sphaerospermum, with descriptions of seven new species from hypersaline environments. *Stud Mycol* **58**, 157–183 (2007).
38. Kaur, J., Kaur, J., Niharika, N. & Lal, R. Sphingomonas laterariae sp. nov., isolated from a hexachlorocyclohexane-contaminated dump site. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* **62**, 2891–2896 (2012).
39. Koskinen, R. *et al.* Characterization of Sphingomonas isolates from Finnish and Swedish drinking water distribution systems. *Journal of Applied Microbiology* **89**, 687–696 (2000).
40. Eram, D., Arthikala, M.-K., Melappa, G. & Santoyo, G. Alternaria species: endophytic fungi as alternative sources of bioactive compounds. *Italian Journal of Mycology* **47**, 40–54 (2018).
41. An, D.-S., Lee, H.-G., Im, W.-T., Liu, Q.-M. & Lee, S.-T. Segetibacter koreensis gen. nov., sp. nov., a novel member of the phylum Bacteroidetes, isolated from the soil of a ginseng field in South Korea. *Int. J. Syst. Evol. Microbiol.* **57**, 1828–1833 (2007).
42. Meyer, S. E., Nelson, D. L., Clement, S. & Ramakrishnan, A. Ecological genetics of the Bromus tectorum (Poaceae)-Ustilago bullata (Ustilaginaceae) pathosystem: A role for frequency-dependent selection? *Am. J. Bot.* **97**, 1304–1312 (2010).
43. Oksanen, J. *et al.* Community Ecology Package. (2019).
44. Sghaier, H. *et al.* Stone-dwelling actinobacteria *Blastococcus saxobsidens*, *Modestobacter marinus* and *Geodermatophilus obscurus* proteogenomes. *The ISME Journal* **10**, 21–29 (2016).
45. Normand, P. *et al.* Genome sequence of radiation-resistant Modestobacter marinus strain BC501, a representative actinobacterium that thrives on calcareous stone surfaces. *J. Bacteriol.* **194**, 4773–4774 (2012).
46. Busarakam, K. *et al.* Modestobacter caceresii sp. nov., novel actinobacteria with an insight into their adaptive mechanisms for survival in extreme hyper-arid Atacama Desert soils. *Syst. Appl. Microbiol.* **39**, 243–251 (2016).
47. Bezerra, J. D. P. *et al.* New endophytic Toxicocladosporium species from cacti in Brazil, and description of Neocladosporium gen. nov. *IMA Fungus* **8**, 77–97 (2017).
48. Castro, J. F. *et al.* Blastococcus atacamensis sp. nov., a novel strain adapted to life in the Yungay core region of the Atacama Desert. *Int. J. Syst. Evol. Microbiol.* **68**, 2712–2721 (2018).
49. Foesel, B. U., Gössner, A. S., Drake, H. L. & Schramm, A. Geminicoccus roseus gen. nov., sp. nov., an aerobic phototrophic Alphaproteobacterium isolated from a marine aquaculture biofilter. *Syst. Appl. Microbiol.* **30**, 581–586 (2007).
50. Dommergue, A. *et al.* Methods to investigate the global atmospheric microbiome. *Front. Microbiol.* **10**, (2019).

Chapter 4: Seasonal changes in local landscapes drive airborne microbial community variation

In preparation for submission in FEMS

Romie Tignat-Perrier[1,2], Aurélien Dommergue[1], Alban Thollot[1], Olivier Magand[1], Pierre Amato[3], Timothy M. Vogel[2], Catherine Larose[2]

[1]Institut des Géosciences de l'Environnement, Université Grenoble Alpes, CNRS, IRD, Grenoble INP, Grenoble, France

[2]Environmental Microbial Genomics, CNRS UMR 5005 Laboratoire Ampère, École Centrale de Lyon, Université de Lyon, Écully, France

[3]Institut de Chimie de Clermont-Ferrand, UMR6096 CNRS Université Clermont Auvergne-Sigma, Clermont-Ferrand, France

Abstract

Microorganisms are ubiquitous in the atmosphere. Thousands to millions of diverse microbial cells are transported into and through the air. Global investigations on the geographical and temporal distribution of airborne microbial communities are critical for determining the sources and factors controlling the geographical and temporal composition of airborne microbial communities. At mid-latitude sites, a seasonal shift in both the concentration and diversity of airborne microbial communities in the planetary boundary layer has been systematically observed. While the factors suspected of affecting this seasonal change were hypothesized (*e.g.,* changes in the surface conditions, meteorological parameters and global air circulation), our understanding on how these factors affect the temporal variation of airborne microbial communities, especially at the microbial taxon level, remains limited. Here, we investigated the distribution of both airborne bacterial and fungal communities over one year. The changes in microbial community structure were correlated to the surrounding landscapes, local meteorology and particulate matter chemistry at the mid-latitude and continental site of Puy-de-Dôme (France; 1465 m altitude from sea level). The changes in the landscapes (cropland and natural vegetation) were correlated to the seasonal shifts in

microbial community composition and structure. The microbial taxa that were most affected during seasonal changes in airborne microbial communities trended differently trends throughout the seasons depending on their trophic mode. In addition, the strong and variable local meteorological conditions found in Puy-de-Dôme were likely responsible for the within season variability observed in the composition of airborne microbial communities.

Introduction

Depending on geography and time of year, thousands to millions of a diverse range of microbial cells are transported in the air. The global distribution of airborne microbial communities is still misunderstood although its understanding is critical to identify how airborne microorganisms might impact human and crop health as well as meteorological processes.

Recent geographical and spatial investigations showed that the local landscapes and local environmental factors might be mainly responsible for the observed distribution of airborne microbial communities of the planetary boundary layer[1,2] (Chapter 3). The composition of airborne microbial communities might be closely related to the nature of the surrounding landscapes (ocean, agricultural soil, forest etc.). Besides, local meteorology might change the Earth's surface conditions (*i.e.* vegetation, snow cover etc.) as well as control microbial cell emission rates from the different surrounding landscapes, impacting the composition of the aerosolized microbial cells indirectly and directly, respectively.

A few experiments of planetary boundary layer microbial communities (both dry phase and liquid phase associated microorganisms) at mid-latitude sites over the course of one year reported a seasonal change in both the concentration and structure of airborne microbial communities[3–8]. They suggested differing explanations such as changes in surface conditions (*i.e.* vegetation, snow cover etc.)[5,6], changes in meteorological conditions[3,4] or changes in the global air circulation (*i.e.* air mass origins)[4,8]. Yet, our understanding on how these potential factors affect the changes in the airborne microbial communities, and more specifically the microbial taxa individually, remains limited.

Here, we investigated the distribution of airborne microbial communities and especially important microbial taxa in relation to the surrounding landscapes, local meteorology and particulate matter chemistry at the mid-latitude and continental site Puy-de-Dôme (France;

+1465 m altitude) over an entire year. We monitored the diversity and abundance of both bacterial and fungal communities over 2016 and 2017 using Illumina MiSeq amplicon sequencing and qPCR analyses.

Puy-de-Dôme is surrounded mainly by cropland and vegetation (*i.e.* > 80% of the surrounding landscapes in a perimeter of 50 km) whose surface conditions change drastically over the different seasons. While the seasonal shift in the composition of airborne microbial communities would be expected to be linked to variations in the relative abundance of microorganisms associated with specific crops and vegetation (*e.g.,* plant pathogens in summer, soil and dead matter associated microorganisms in winter etc.), the within season variability of airborne microbial communities is expected to be associated to local meteorology.

Material and Methods

Sites and Sampling

A size selective high volume air sampling installed at the Puy-de-Dôme (PDD) meteorological station terrace was used to collect particulate matter on quartz fiber filters every week from June 2016 to August 2017 (supplementary **Table S1**). The sampler was equipped with a PM10 size-selective inlet in order to collect particulate matter smaller than 10 µm (PM10) and sampling was done as presented in Dommergue *et al.* (2019)[9]. The different samples and collected volumes are presented in **Table S1** in supplementary data. Quartz fiber filters were heated to 500°C for 8 hours to remove traces of organic carbon including DNA. All the material including the filter holders, aluminium foils and plastic bags in which the filters were transported were sterilized using UV radiation as detailed in Dommergue *et al.* (2019)[9]. A series of field and transportation blank filters were done to monitor and check the quality of the sampling protocol as presented in Dommergue *et al.* (2019)[9]. PDD is a mid-altitude (+ 1465 m) site surrounding by croplands, an urban area (Clermont-Ferrand) and forests within a 50 km perimeter (**Fig. 1**). Monthly NASA satellite images of Puy-de-Dôme surrounding surfaces (https://wvs.earthdata.nasa.gov/) are shown in **Figure S1** in supplementary data. The Atlantic coast and Mediterranean Sea are at around 320 km and 240 km from PDD, respectively.

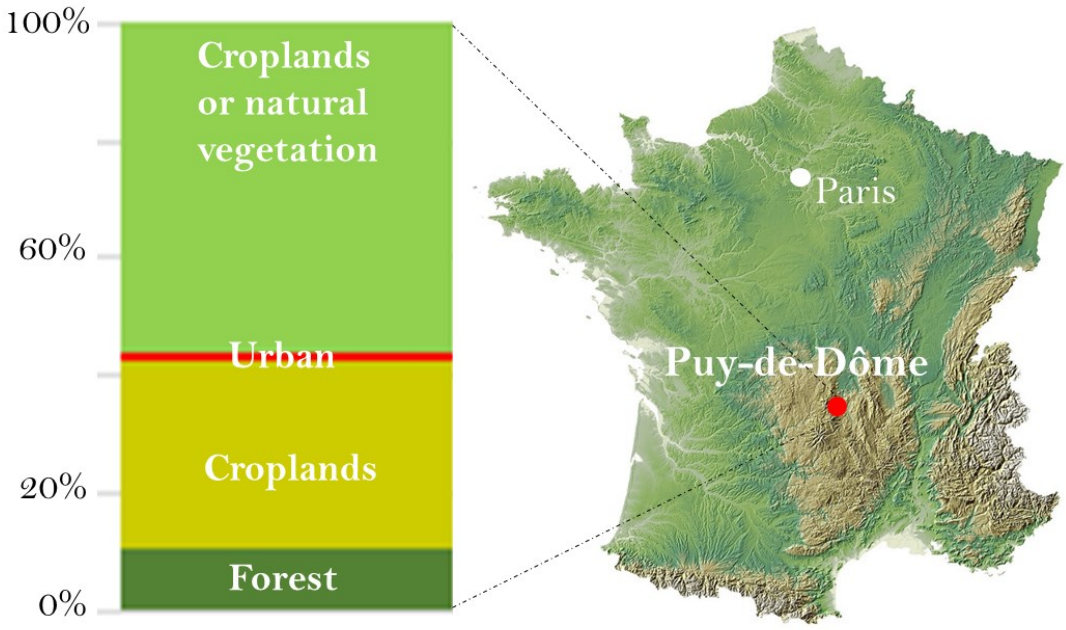


Figure 1: Central position of Puy-de-Dôme in France and relative surfaces of the landscapes surrounding the site in a perimeter of 50 km based on the MODIS land surfaces. Cropland and vegetation areas make > 80% of the surrounding landscapes, while forest and urban areas (mainly Clermont-Ferrand) make < 20% of the surrounding landscapes.

DNA extraction

We extracted DNA from 3 punches (diameter of one punch: 38 mm) from the quartz fiber filters using the DNeasy PowerWater kit as described in Chapter 2. DNA concentration was measured using the High Sensitive Qubit Fluorometric Quantification (Thermo Fisher Scientific) then DNA was stored at -20°C.

Real-Time qPCR analyses

The bacterial cell concentration was approximated by the number of 16S rRNA gene copies per cubic meter of air and the fungal cell concentration was approximated by the number of 18S rRNA gene copies per cubic meter of air. Primers and methodology are presented in Chapter 3.

MiSeq Illumina amplicon sequencing

16S rRNA gene sequencing: library preparation, reads quality filtering and taxonomic annotation. The V3-V4 region of the 16S rRNA gene was amplified and libraries were prepared

as presented in Chapter 3. The amplicons were sequenced by paired-end MiSeq sequencing using the V3 Illumina technology with 2 x 250 cycles. Reads were filtered based on quality using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), assembled using PANDAseq[10], and annotated using RDP Classifier[11] and the RDP 16srRNA database as detailed in Chapter 3. The number of sequences per sample and the percentage of sequences annotated at the genus level were evaluated using a home-made R script. The sequences annotated as chloroplasts were removed.

ITS sequencing: library preparation, reads quality filtering and taxonomic annotation. The ITS2 region of the ITS was amplified libraries were prepared as presented in the Chapter 3. The amplicons were sequenced by a paired-end MiSeq sequencing using the technology V2 of Illumina with 2 x 250 cycles. Reads were filtered based on quality using FASTX-Toolkit, assembled using PANDAseq, and annotated using RDP Classifier and the RDP fungallsu database as detailed in Chapter 3. The number of sequences per sample and the percentage of sequences annotated at the species level were evaluated using a home-made R script.

The number of reads per sample and per sequencing is presented in the **Table S1** in supplementary data. Samples with less than 6000 reads were removed. Rarefaction curves per season are presented in **Figure S2** in supplementary data.

Estimation of the trophic mode of the fungal species

We used the FUNGuild software[12] to assign the trophic mode of the fungal species (RDP classifier based annotation). Fungal species annotated to a trophic mode with the level confidence “possible” were grouped in the “not classified” fungi. Then, we calculated the percentage represented by each trophic mode per sample. Heatmaps were done using the R package gplots.

Chemical analyses

The elemental carbon (EC), organic carbon (OC), sugar anhydrides and alcohols (levoglucosan, mannosan, galactosan, inositol, glycerol, erythriol, xylitol, arabitol, sorbitol, mannitol, trehalose, rhamnose, glucose, fructose and sucrose), soluble anions (MSA, SO_4^{2-}, NO_3^-, Cl^-, Ox) and cations (Na^+, NH_4^+, K^+, Mg^{2+}, Ca^{2+}) concentrations were analyzed as presented in Dommergue *et al.* (2019)[9].

Meteorological data

Meteorological parameters such as wind speed and direction, temperature, relative humidity and UV were collected. For each sample, the backtrajectories of the air masses were calculated over 3 days (maximum height above sea level: 1 km) before the sampling using HYSPLIT[13].

Graphical and Statistical analyses

Environmental variables. Values in ng per cubic meter of air were used in the analyses. The chemical table was log10-transformed to approach a Gaussian distribution, and a hierarchical cluster analysis (average method) was done on the Euclidean distance matrix using the vegan and ade4 R packages. Meteorological data were used to do the wind roses using the openair R package[14]. Backtrajectories of the air masses over three days were plotted on maps using the openair R package, and the relative surfaces of the landscapes (MODIS land surfaces) air masses over flown were calculated.

Diversity statistics and Multivariate analyses. Before doing the multivariate analyses, the raw abundances of the taxa (bacterial genera and fungal species) were transformed into relative abundances to counter the heterogeneity in the number of sequences per sample. A hierarchical cluster analysis (average metric) on the Bray-Curtis dissimilarity matrix was done following Hellinger transformation of the data using the vegan and ade4 R packages[15]. To access the variability of microbial population structure within the seasons (winter: 22th of December to 19th of March; spring: 20th of March to 19th of June; summer: 20th of June to 22th of September; autumn: 23th of September to 21th of December), we averaged the degrees of dissimilarity obtained from the Bray-Curtis matrix for each pair of samples from the same season, subtracted these values to 1 to get similarity values and divided the similarity values by the standard deviation (as detailed in Chapter 3). Spearman correlations were calculated to test the correlation between microbial abundance and richness and quantitative environmental factors using the Hmisc R package[16]. ANOVA analyses were used to test the influence of qualitative factors such as season and year on both bacterial and fungal abundance and richness using the vegan R package, followed by TukeyHSD tests to identify which group revealed a significantly different mean. A distance-based redundancy analysis

(RDA – linear or non-linear correlation) was carried out to evaluate the part of the variance between the samples explained by the seasons, chemistry, meteorology and/or relative surfaces based on the backtrajectories, and an ANOVA was carried out to test each variable using the vegan and ade4 R packages. Venn diagrams using the R package VennDiagram were done to access the shared and unique bacterial genera and fungal species from each season after rarefaction on the raw abundances (rarefaction at the minimum number of reads). A Mantel test between the Bray-Curtis matrices based on the bacterial genera and fungal species was used to evaluate similarities in the distribution of the samples. A Mantel test was done between the Bray-Curtis matrix based on either the bacterial or the fungal diversity and the Euclidean distance matrix based on the chemical variables or meteorology or relative surfaces to evaluate the similarities in the distribution of the samples.

Results

Temporal distribution of airborne microbial communities

Airborne microbial abundance. Airborne bacterial and fungal concentration varied between $1.8x10^3$ and $2.1x10^7$ cells per cubic meter of air and 3 and $1.0x10^5$ cells per cubic meter of air, respectively. A significant correlation was observed between the bacterial and the fungal concentrations (r=0.77, pvalue=$7.9x10^{-11}$). The average bacterial and fungal concentrations did not differ significantly among the seasons (pvalues=0.20 and 0.38), although the highest values were found in spring and summer for both bacteria and fungi (**Fig. 2** and supplementary **Table S2**).

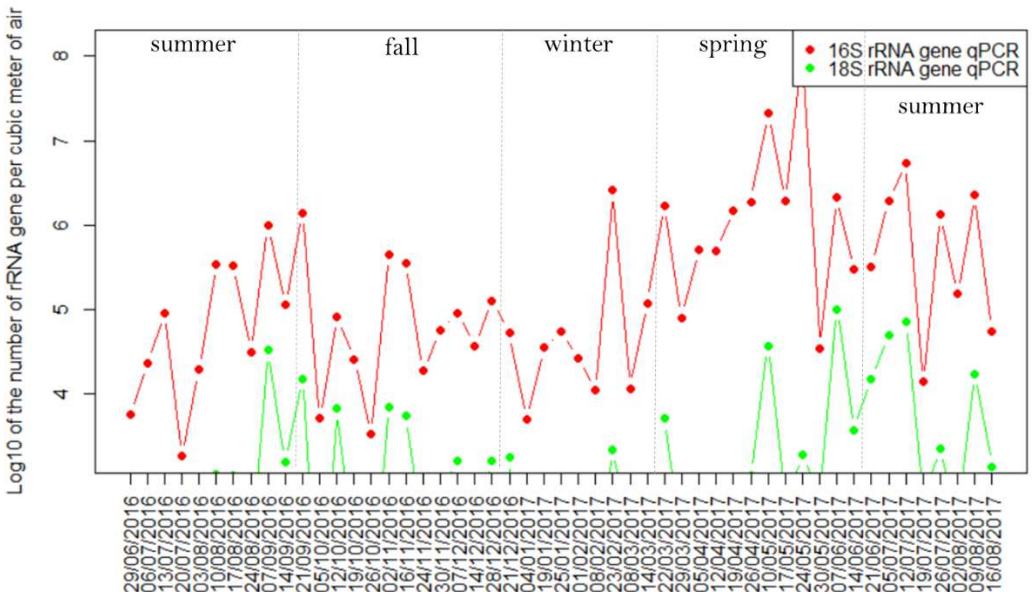


Figure 2: Log10 of the bacterial and fungal cell concentration estimated by the number of 16S and 18S rRNA gene copies per cubic meter of air and per sample.

Airborne microbial richness. The twenty-five most abundant bacterial genera and fungal species observed per season are found in **Table S3** of the supplementary data. The five most abundant bacterial genera were *Bacillus* (15.6%), *Hymenobacter* (9.2%), *Sphingomonas* (7.4%), *Methylobacterium* (5.1%) and *Thermoactinomyces* (4.6%). Their relative abundance varied between seasons; for some seasons they did not appear in the top five genera. For example, during the spring, after *Bacillus* (18.8%), the second and third most abundant genera were *Romboutsia* (8.2%) and *Clostridum_sensu_stricto* (7%). *Sphingomonas* was abundant in summer (12.5%), but much less abundant in the other seasons (autumn: 3.8%, spring: 3.6%, winter: 2.6%). The five most abundant fungal species were *Pseudotaeniolina globosa* (11.5%), *Cladophialophora proteae* (8.3%), *Alternaria sp BMP_2012* (4%), *Cladophialophora minutissima* (4%) and *Naevala minutissima* (3.1%). Their relative abundance varied seasonally. For example, the relative abundance of *Cladophialophora proteae* was 13.3% in winter and 0.3% in summer; the relative abundance of *Naevala minutissima* was 10.1% in spring and 0 in winter.

The number of observed bacterial genera and fungal species varied between 150 and 674 bacterial genera, and 84 and 649 fungal species. The Chao1 bacterial and fungal richness estimation varied between 234 and 897 bacterial genera, and 97 and 820 fungal species. The Chao1 bacterial richness estimation was significantly higher in summer compared to winter

and autumn (pvalue=4.5x10^{-4}; supplementary **Table S1**) and the within season variability was similar (pvalue>0.005). The Chao1 fungal richness estimation did not differ among the seasons (pvalue=0.18; supplementary **Table S1**) although higher in summer, and the within season variability was relatively lower in spring. A significant correlation was found between the Chao1 richness estimations and qPCR concentrations for both bacteria and fungi (r=0.41, pvalue=3.0x10^{-3} and r=0.50, pvalue=4.6x10^{-4}, respectively).

Summer had the highest percentage of unique bacterial genera (16%), followed by autumn (4.5%), spring (4.4%) and winter (4.3%) (supplementary **Fig. S3**). This was also observed for unique fungal species, with the highest percentage observed in summer (28.4%), followed by autumn (17.4%), spring (14.4%) and winter (8%) (supplementary **Fig. S3**). The majority of the bacterial genera and fungal species observed in each season was shared by all the seasons (supplementary **Fig. S3**).

Airborne microbial structure. The hierarchical cluster analyses of the samples based on both the bacterial and fungal community structures are shown in the **Figure 3**. The distribution of the samples based on the bacterial community structure revealed high similarities with the distribution of the samples based on the fungal community structure (r=0.50 and pvalue<0.001). The samples tended to group by season (anosim r=0.30 and r=0.49 with pvalues=0.001 for bacterial and fungal communities, respectively; **Fig. 3**), and the season explained 17% and 22% of the variability between the samples based on the bacterial and fungal community structure respectively over the 2 first components (pvalue=0.001). The distribution of the summer samples based on the fungal species were different between 2016 and 2017 (pvalue=0.001). Temporal variability in bacterial community structure was the highest in winter and the smallest in autumn, and the temporal variability in fungal community structure was the highest in autumn and the smallest in spring (supplementary **Table S4**).

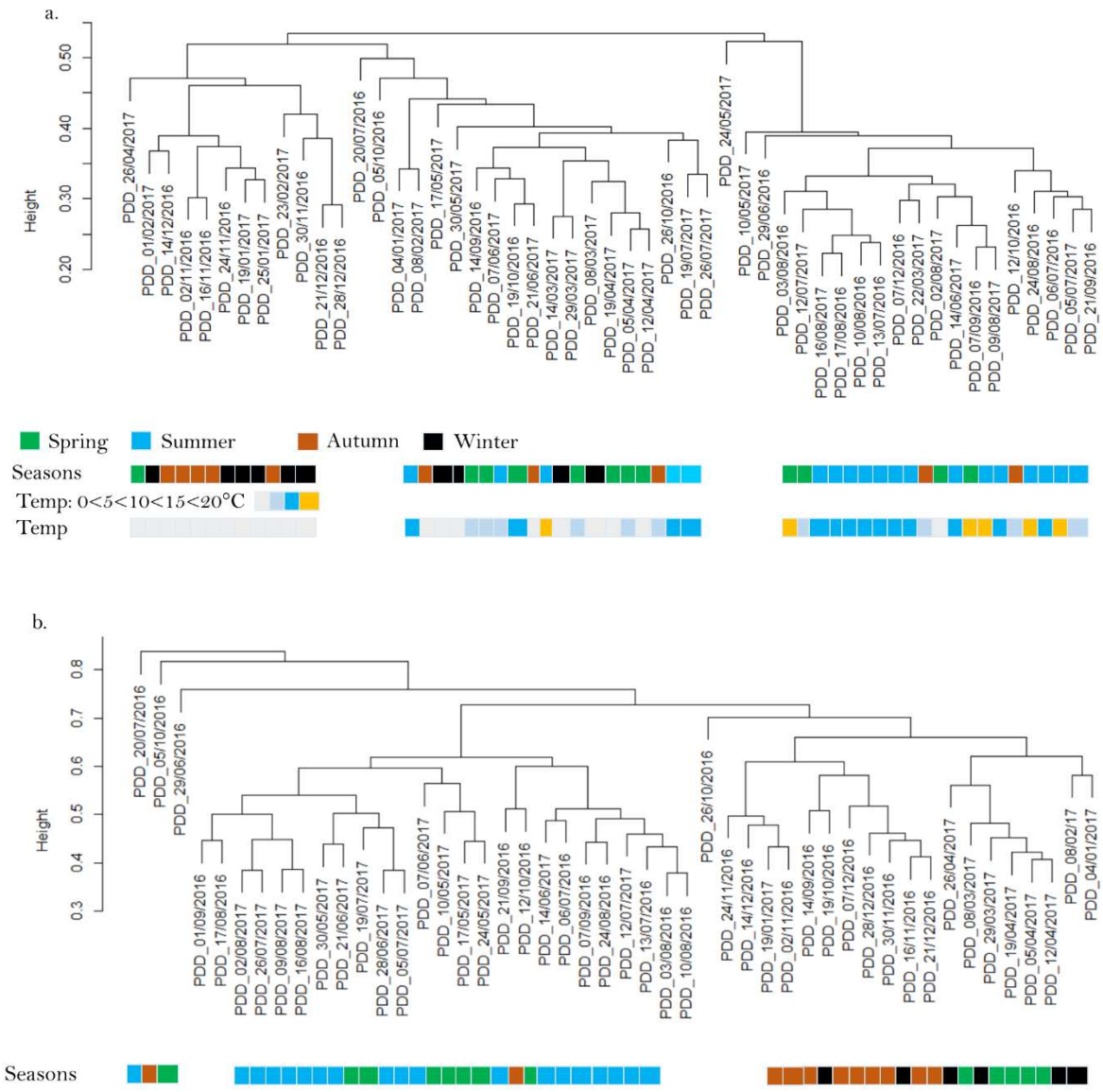


Figure 3: Hierarchical cluster analysis (average method) on the airborne bacterial community structure (16S rRNA gene – a.) and airborne fungal community structure (ITS region – b.) based on the Bray-Curtis dissimilarity matrices. The season to which each sample belongs (black=winter, blue=summer, green=spring and brown=autumn) (as well as the mean temperature for b.; grey between 0 and 5°C, light blue between 5 and 10°C, blue between 10 and 15°C and orange between 15 and 20°C) are indicated under each site in a frieze.

The relative abundance of fungal saprotrophs and symbiotrophs (like *Balospora myosura*, *Strobilurus albipilatus*, *Pseudotaeniolina globosa*) (**Fig. 4**) was higher in fall and winter, although not significative for the latter (pvalue=4.3×10^{-5} and 0.5 for saprotrophs and symbiotrophs, respectively). The relative abundance of fungal pathotrophs was higher in the

spring and summer periods although not significant (pvalue=0.51 - **Fig. 4**). Examples of phytopathogens increasing in relative abundance during the summer and/or spring are *Mycosphaerella graminicola* (wheat plant pathogen – supplementary **Fig**. **S4**), *Microdochium majus* (cereal pathogen – supplementary **Fig**. **S4**), *Blumeria graminis* (powdery mildew pathogen – supplementary **Fig**. **S4**), *Ustilago hordei* (maize pathogen – supplementary **Fig**. **S4**), *Botryotinia fuckeliana* (gray mold disease) and *Erysiphe alphitoides* (powdery mildew on oak trees). However, in terms of the number (richness) of fungal species associated to the trophic modes, spring and summer showed a higher richness in pathotrophs and saprotrophs; and summer and fall showed a higher richness in symbiotrophs (supplementary **Table S5**).

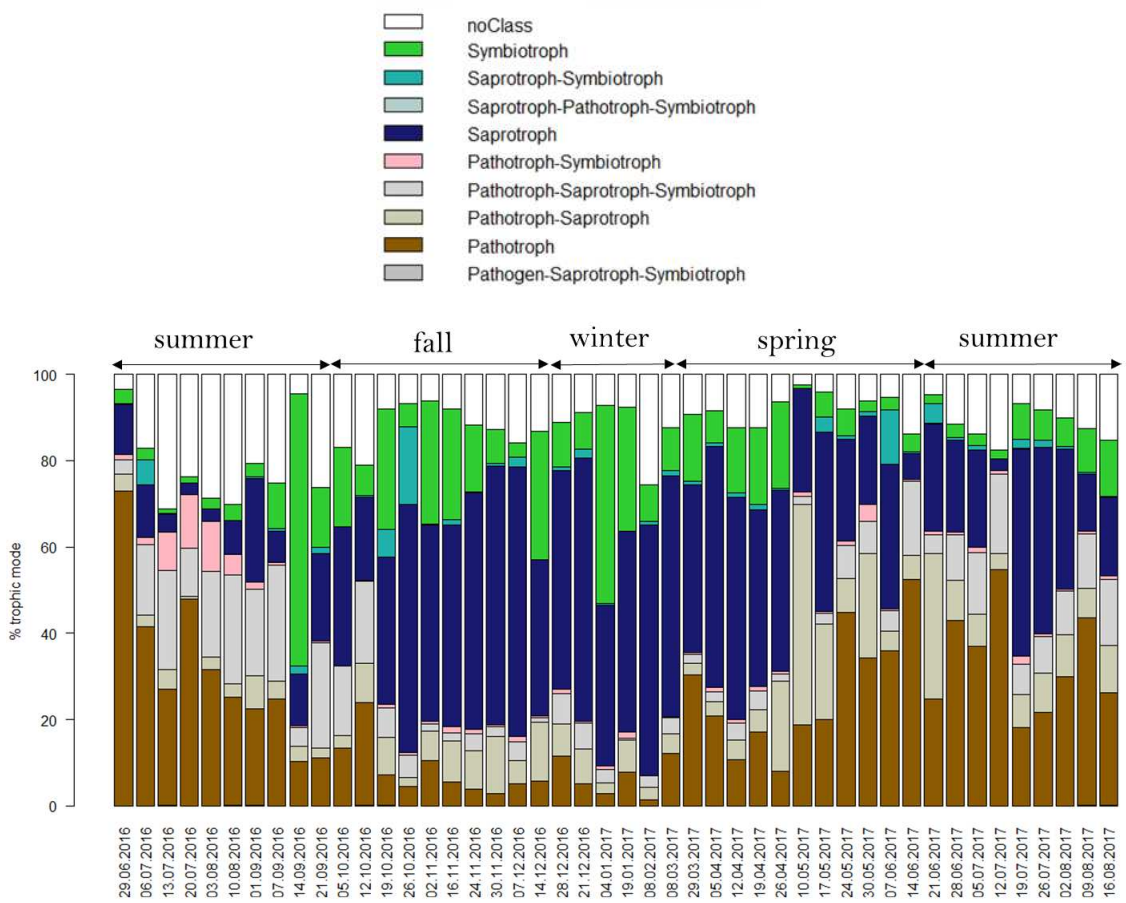


Figure 4: Trophic mode of the fungal species. Each fungal species detected per sample was associated to either a specific trophic mode, different trophic modes or no trophic mode (« noClass ») using the FUNGuild software, then the percentage represented by each trophic mode was calculated.

When looking at the fifty most abundant bacterial genera and fungal species of the dataset, which mainly control the distribution of the samples based on the microbial community structures, we observed three different patterns in the temporal variation of their relative abundance. 20% (10/50) of the bacterial genera and 34% (17/50) of the fungal species showed a higher relative abundance in winter and/or fall compared to the other seasons. This was the case with *Hymenobacter* and *Pseudotaeniolina globosa* for example, whose relative abundances increased during the period between 11/02/2016 and 02/23/2016, and between 10/19/2016 and 04/26/2016, respectively (**Fig**. **5**). The period during which the relative abundance increased varied in length and depended on the microorganism. Fungal taxa belonging to this group were almost exclusively saprotroph or pathotroph (supplementary **Table S6**). 42% (21/50) of the bacterial genera and 46% (23/50) of the fungal species had higher relative abundances in summer and/or spring and fall. For example, *Sphingomonas* and *Alternaria sp BMP_2002* relative abundances started to decrease in summer (08/17/2016) and fall (10/12/2016), respectively; and started to increase in the late spring (06/14/2017 and 06/10/2017, respectively) (**Fig**. **5**). Fungal taxa belonging to this group were pathotroph, symbiotroph or saprotroph (supplementary **Table S6**). Finally, 38% (19/50) of the bacterial genera and 20% (10/50) of the fungal species showed no clear pattern in their relative abundance (*i.e.* stable or highly variable over the year), see *Bacillus* and *Phaeosphaeria herpotrichoides* for example (**Fig**. **5**). Fungal taxa belonging to this group were saprotroph (supplementary **Table S6**). The temporal evolution in the relative abundances centered and scales of the fifty most abundant bacterial genera and fungal species is shown on the heatmaps of the **Figure S5** in supplementary data.

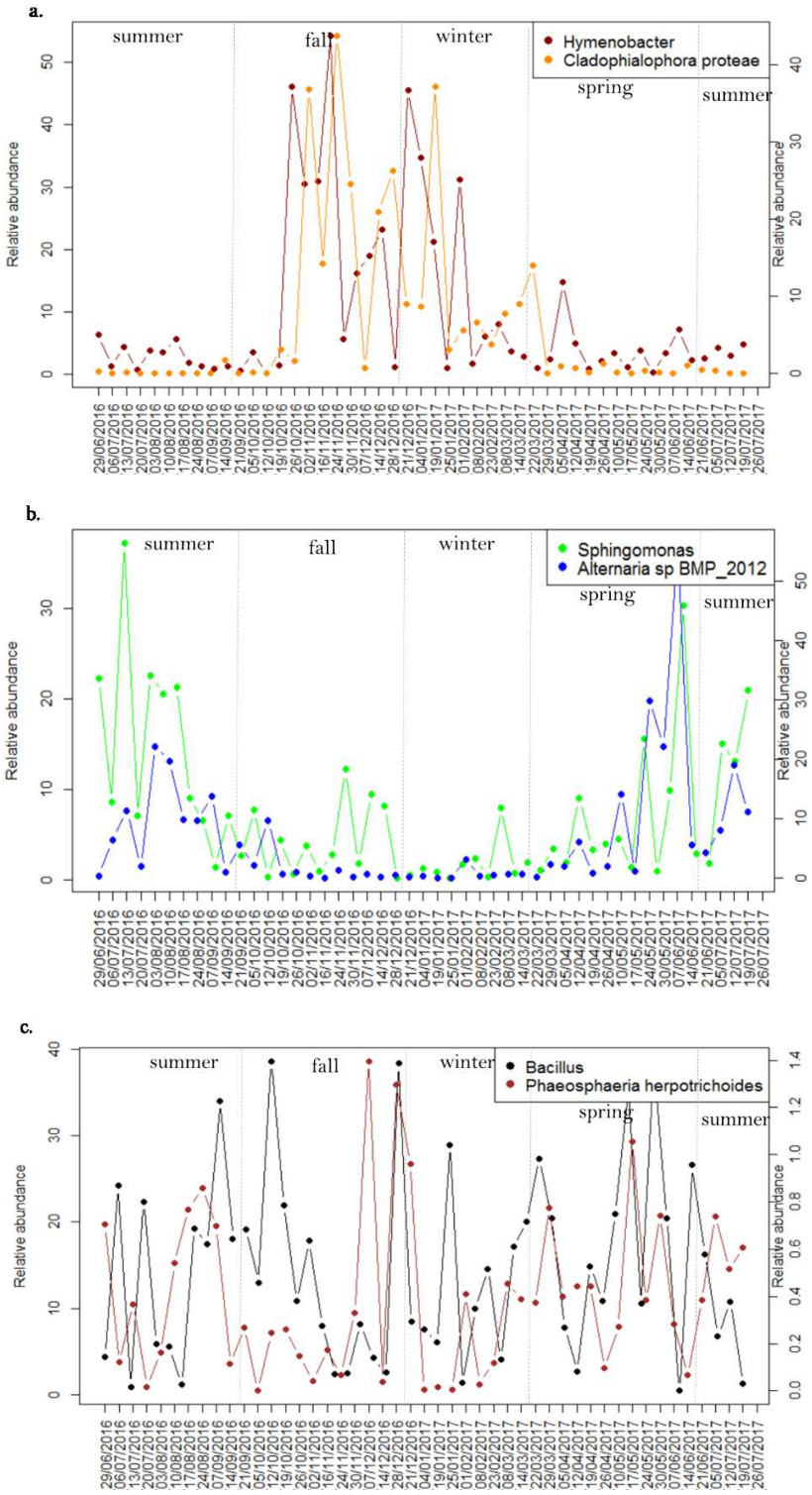


Figure 5: Temporal evolution of the relative abundance of several fungal species and bacterial genera over the year.

a) Higher relative abundance in fall and winter: examples of the fungal species *Cladophialophora proteae* and the bacterial genus *Hymenobacter*
b) Higher relative abundance in spring and summer: examples of the fungal species *Alternaria sp BMP 2012* and the bacterial genus *Sphingomonas*
c) No trend of the relative abundance over the year: examples of the fungal species *Phaeosphaeria herpotrichoides* and the bacterial genus *Bacillus*

Controlling factors

Atmospheric PM10 chemistry. The hierarchical cluster analysis of the samples based on the PM10 chemical profile is shown in the **Figure S6** in supplementary data. Chemical concentrations were significantly different depending on the seasons for all the substances except Cl^-, NO_3^-, NH_4^+ and rhamnose (supplementary **Table S7**). On average, summer showed higher concentrations of OC and Ca^{2+}, spring showed higher concentrations of NO_3^-, SO_4^{2-} and NH_4^+ and winter higher concentrations of Cl^- (supplementary **Table S7**). Seasons explained 38% of the variance in the distribution of the samples based on the chemistry over the first 2 components (pvalue=0.001). When looking at the weekly variation of the concentration of the different chemical substances, we observed two general trends with some chemicals whose concentrations increased (galactosan, levoglucosan and mannosan) or decreased (glucose, trehalose, sorbitol, xylitol, inositol, SO_4^{2-}, OC, MSA) during the winter (**Fig. 6**).

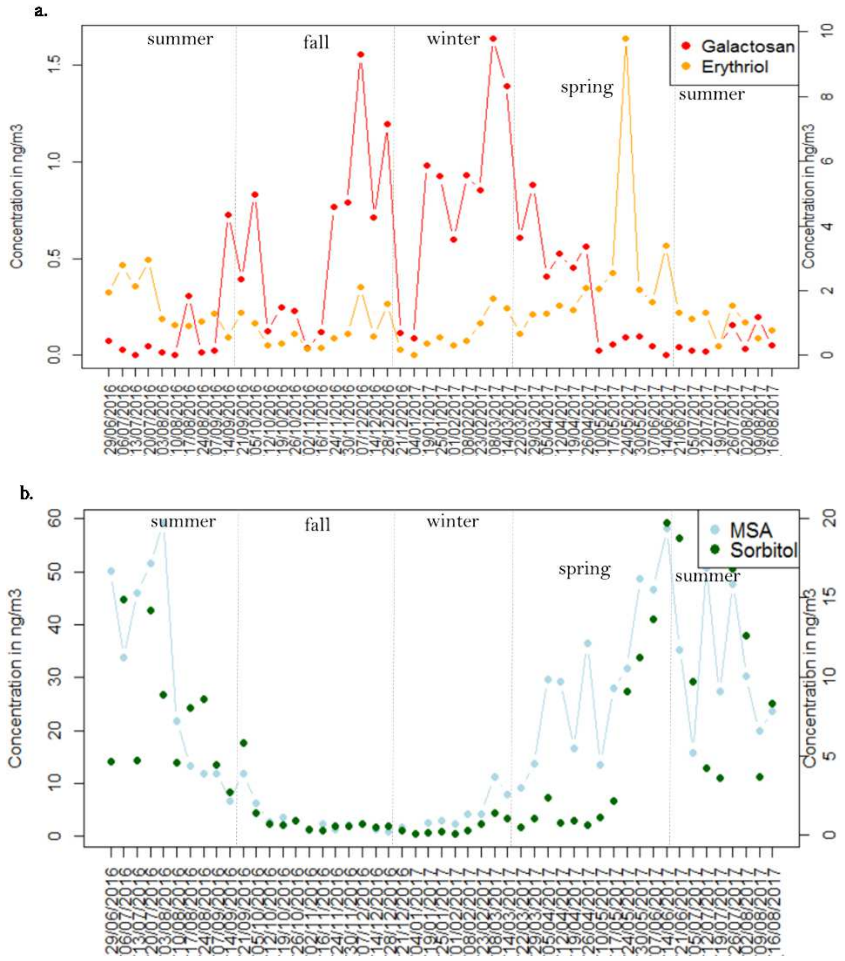


Figure 6: Temporal evolution of the concentration of chemical species over the year.

a) Galactosan and erythriol
b) MSA and sorbitol

Both distributions based on the bacterial and fungal communities were correlated to the distribution based on the overall PM10 chemistry (Mantel r=0.25 pvalue=0.001 and 0.41 pvalue=0.001, respectively) (see RDA analyses in supplementary **Fig. S7**). Based on the RDA, specific chemical compounds (*i.e.* the polyols, MSA and Na^+ for the fungal communities, and organic carbon, Na^+, Cl^- and erythriol for the bacterial communities) were the main contributors to these correlations.

We observed strong correlations between bacterial genera or fungal species, especially from the fifty most abundant bacterial genera and fungal species that showed seasonal patterns in their abundances, and chemical substances which showed also a general trend over the year. For examples, mannitol-arabitol and *Alternaria sp BMP_2002* r=0.67 and pvalue=$2x10^{-7}$ or mannitol-arabitol and *Pseudotaeniolina globose* r=-0.67 and pvalue=$3x10^{-7}$ (supplementary **Table S8**).

Weather characteristics: local meteorology and air mass backtrajectories (air mass origin).

Temperature, wind speed and UV radiation averaged per week varied over the year while wind direction and relative humidity did not. The mean temperature per week varied between -5.5 and 17.6°C and was significantly correlated to the seasons (pvalue=$1.4x10^{-11}$, supplementary **Fig. S8** and **S9**). The wind speed varied between 2.8 and 13.5 m/s on average over the week, and was significantly higher in winter as compared to summer and spring (pvalue=$2.1x10^{-3}$– supplementary **Fig**. **S9**), and significantly more variable in winter compared to the other seasons (pvalue=$5.4x10^{-4}$). The wind speed was positively correlated to the wind direction (r=0.44, pvalue=0.001): higher wind speed tended to come from the west of the sampling site. The mean direction of the wind varied between 115.4 and 248.8° over the week, and was not significantly different between seasons (pvalue=0.11 – supplementary **Fig**. **S9**). The wind roses and backtrajectories over the seasons are presented in supplementary **Figure S9** and **Figure S10**. The variability in wind conditions, as well in temperature and relative humidity for each season is shown in supplementary **Table S4**, in regards to other geographical sites from Chapter 2.

Bacterial concentration best correlated with UV radiation (r=0.38, pvalue=$5.4x10^{-3}$) and fungal concentration best correlated with temperature (r=0.38, pvalue=$5.6x10^{-3}$) (supplementary **Table S9**). The Chao1 bacterial richness estimation revealed a strong and significant

correlation with temperature (r=0.52, pvalue=$7.7x10^{-5}$), while the Chao1 fungal richness estimation showed a strong and significant correlation with the wind direction (r=0.50, pvalue=$4.6x10^{-4}$) (supplementary **Table S9**). A significant and strong correlation was found between the overall meteorology and both the bacterial and fungal community structures (r=0.57 pvalue=0.001 and r=0.48 pvalue=0.001). The main contributors were temperature and wind speed, followed by wind direction and relative humidity for bacterial communities; and the wind direction and speed, temperature and UV radiation for fungal communities.

The correlation between the overall PM10 chemistry and the relative surfaces corrected by the backtrajectories (both calculated over a 50km-perimeter and over 3 days) was weak (r=0.10, pvalue=0.001), as were the correlations between both the bacterial and fungal communities and the relative surfaces corrected by the backtrajectories (r=0.12 pvalue=0.001 and r=0.21 pvalue=0.001, respectively). These relative surfaces corrected by the backtrajectories were only partially correlated to the seasons (16% and 14% of the variance explained, pvalue=0.016 and 0.013 for 50 km backtrajectories and 3-days backtrajectories respectively).

Discussion

Previous investigations on the variation of airborne microbial communities over time have consistently shown a seasonal shift in terms of both concentration and structure[3–7]. This seasonal shift was explained mainly by changes in the landscape condition (*e.g.,* vegetation in summer, snow cover in winter *etc.*)[5,6], seasonal changes in local meteorology[3,4] and/or seasonal changes in the origin of the air masses (*i.e.* changes in the global air circulation)[4]. Our data showed that seasonality might be a significant factor influencing the temporal distribution of airborne microbial communities in Puy-de-Dôme in impacting microbial structure gradually over the year. The concentration of airborne bacterial and fungal cells showed a seasonal variation resulting in higher concentration in spring and lower concentrations in winter; and the bacterial and fungal richness varied throughout the year with a higher richness in summer, although no significant. Puy-de-Dôme is surrounded mainly by cropland and/or vegetation (> 80% in a 50 km perimeter; **Fig**. **1**) which undergo strong changes over the year in term of surface conditions (see Puy-de-Dôme monthly satellite images in supplementary **Fig. S1** for a glimpse). These changes are associated with crops and

vegetation (like trees) that cycle through their different phases on an annual basis. These changes are also strongly linked to the changes in local meteorology (*e.g.,* temperature and precipitation). This has consequently a critical influence on the terrestrial microbial communities which would change in parallel[17,18]. Our data suggest that surface conditions might have been critical for explaining changes in the airborne microbial community distribution over the year. Over the year, different terrestrial microorganisms of different trophic modes were likely aerosolized (**Fig. 4**). Fungal phytopathogens (like *Ustilago hordei*) and leaf-associated fungi (like *Naevala minutissima*) increased in relative abundance during the spring and summer periods when the crop plants grew and trees were green; while saprotrophs (like *Cladophialora protea*) increased in relative abundance during the fall and winter periods after crop harvesting and when dead and decomposing biological material (like leaves) covered terrestrial surfaces. Similarly to fungi, soil-associated bacterial taxa (like *Hymenobacter*) were observed in higher relative abundance in fall and winter, and bacterial phytopathogens and leaf-associated bacteria (like *Pseudomonas*) were observed in higher relative abundance in spring and summer. Changes in the microbial communities associated with specific crops and vegetation might have strongly influenced the composition of airborne microbial communities. The observation of three general trends (*i.e.* an increase of specific taxa in fall/winter; an increase of other taxa in summer/spring; stable or highly variable microbial taxa over the year – **Fig. 5**) might have driven the dichotomy (*i.e.* fall/winter samples and spring/summer samples – **Fig. 3**) observed in the distribution of the samples based on both the airborne bacterial and fungal community structures. In addition, the period during which we observed an increase in the relative abundance of a winter or summer-associated microbial taxon varied in length and might have been specific to the life strategy of the taxon (for example, a taxon would increase in relative abundance between June and August while another between July and August only). This might have partly driven the high variability observed within seasons. Microbial taxa whose relative abundance remained stable or varied highly over the year might partly belong to taxa associated to decomposing matter (the trophic mode of all the fungal taxa present in this group was saprotroph), which is necessarily present the whole year although much more abundant in winter. Airborne fungal community structure was more significantly correlated to the seasons than bacterial community structure, and a higher number of fungal taxa showed a seasonal pattern over the fifty most

abundant taxa as compared to bacterial taxa (40/50 and 31/50 for fungi and bacteria, respectively). The stronger seasonal shift for airborne fungi might be explained by a higher influence of vegetation on fungi as compared to bacteria[19].

Local meteorology might also have a role in the temporal distribution of airborne microbial communities, especially on their within season variability. In a previous study (Chapter 3), we highlighted the importance of local meteorology (especially wind speed, wind direction and temperature variability over time) in the variability of airborne microbial communities especially when the site was enriched in landscapes (*i.e.* high richness and low evenness). Puy-de-Dôme is characterized by strong wind speed and highly changing meteorological conditions partly explained by its central position and highest elevation in the Chaîne-des-Puys (+1465 m from sea level). The wind direction changed rapidly both within and between the weeks. The relative humidity and temperature also showed a high variability within and between the weeks within the seasons (supplementary **Table S4**). While the correlation between meteorological parameters and airborne microbial communities in Puy-de-Dôme could be explained by seasonality, these meteorological parameters might have also affected airborne microbial structure at a smaller scale than the seasons in impacting the aerosolisation process as discussed in Chapter 3. For example, specific temperatures might activate the sporulation of specific fungi[20]. The high correlation observed between temperature and airborne bacterial richness might be explained by the indirect effect of the temperature on the increase of the microbial diversity in the sources (*i.e.* surrounding landscapes)[21–24] and also by a more direct effect with the enhanced rising of warm air masses which might increase the aerosolisation rate of microorganisms. Wind parameters might also have played a role in the high variability in the composition of airborne microbial communities observed within seasons, with higher wind speeds potentially uplifting heavier cells or bringing cells from more distant sources. Wind parameters were the best meteorological parameters correlated to airborne fungal richness.

The wind direction and the backtrajectories did not drastically change throughout the seasons, and were highly variable over the year (supplementary **Fig. S9** and **S10**). We did not observe a correlation between airborne microbial structure and the origin of the air masses. In winter, during which free troposphere air masses have the larger influence[25–27] (and thus potentially long-range transport), the highest variability in the bacterial community structure was

observed, which was consistent with both potential larger inputs from distant sources and stronger meteorological conditions. Marine air masses have been shown to contribute the most (72%) during winter in Puy-de-Dôme[27]. Especially the sample which started in February the 2nd (2017) showed a marine signature. However, this sample was as concentrated in marine bacterial genera (*Corialomargarita*, *Rubritalea, Aquimarina*) as other samples which did not present a marine signature (March the 8th, March the 14th, May the 24th, July the 12th and August the 9th for *Corialomargarita*). Marine bacterial genera concentrations were very low whenever present (they were represented by 2 or 3 sequences over > 20000 sequences per sample). Uetake *et al.* (2019)[3] investigated airborne microbial communities in Tokyo (Japan) over a year, and have also highlighted the absence of oceanic related bacteria (*i.e.* SAR group, *Oceanospirillales*) which could have suggested a long-range transport of airborne microbial cells from the Pacific Ocean to Tokyo. We looked at the samples which presented highly dissimilar bacterial and fungal community structures compared to the other samples and tried to link their change in microbial community structure with a change in PM10 chemistry or a change in the backtrajectories, which might have suggested inputs from a distant source. We observed that they were either not correlated to a change in PM10 chemistry (for example the sample 05/24/2017 which showed a particular bacterial community structure) or correlated to slightly changes in PM10 chemistry (for examples the samples 07/20/2016 and 10/05/2016 which showed particular bacterial and fungal community structures showed also higher concentrations of SO_4^{2-}/NO_3^- and NO_3^- /levoglucosan-manosan compared to the samples from the same season, respectively). Similarly, a significant change of PM10 chemistry (for example the sample starting in August the 3rd 2017 which presented higher concentrations of MSA, NO_3^- and NH_4^+ compared to the other samples of the same season) was not necessarily correlated to a change in microbial community structure.

Thus, our data did not provide evidence that the long-range transport and the microbial inputs coming from distant sources had a significant impact on the temporal structuration of airborne microbial communities observed in Puy-de-Dôme. However, our sampling strategy (one-week sampling) might not have been adapted to identify microbial long-range transport events which, if they occurred, might have been greatly diluted in the local sources. Investigations on the temporal variability of airborne microbial communities which observed

a correlation between backtrajectories, the composition of airborne microbial communities and/or PM10 chemical composition suggesting a high contribution of distant sources in the composition of airborne microbial communities[4,8], had either a different sampling strategy then ours (*i.e.* shorter sampling duration such as 24h)[4] or did not sample microorganisms of the dry phase of the troposphere as we did[8]. Indeed, Cáliz et *al.* (2018)[8] investigated rain associated microbial communities whose controlling factors might be different than microbial communities of the dry phase of the lower part of the troposphere (for example rain associated microorganisms that originate from clouds might have travelled long distances in the higher part of the troposphere before falling as precipitation).

We observed a seasonal change in the particulate matter chemistry (PM10 chemistry). In Puy-de-Dôme, changes in PM10 chemistry through the seasons (for examples higher concentrations of SO_4^{2-} and NH_4^+ in spring and Na^+ Cl^- in autumn and winter) have been correlated to changes in air mass origins (local and distant) and changes in the vertical stratification of the atmospheric layers (*i.e.* increase of the height of the mixed layer in summer)[25,26]. We observed specific and strong correlations between some chemical species and microbial taxa (supplementary **Table S8**) which could explain the overall correlation observed between the overall PM10 chemistry and airborne microbial structure. In particular, we observed many correlations between specific microbial taxa and polyols. The concentration of atmospheric polyols might be due to the presence of airborne green plant debris as well as specific microbial taxa and especially fungal taxa which produce polyols[28–31].

Conclusion

We investigated changes in airborne microbial community concentration and structure at the elevated continental site Puy-de-Dôme, highest point of the Chaîne-des-Puys in the Massif Central of France (1465 m altitude above sea level) as a function of the surrounding landscapes, local meteorology and particulate matter chemistry. Puy-de-Dôme airborne microbial community structure shifted throughout the year in relation with the seasonal changes in surface conditions of the surrounding landscape of Puy-de-Dôme, characterized mainly by cropland and natural vegetation. The microbial taxa that drove the seasonal distribution of airborne microbial communities showed different trends throughout the seasons depending on their trophic mode. Crop-associated microorganisms, and especially

crop pathogens were in higher relative abundance in spring/summer while soil-associated microorganisms and dead material-associated microorganisms were found in higher relative abundance in fall/winter. Within the seasons, the temporal variability observed in the composition of airborne microbial communities was likely associated to the variable and strong meteorological conditions observed in Puy-de-Dôme.

References

1. Mhuireach, G. Á., Betancourt-Román, C. M., Green, J. L. & Johnson, B. R. Spatiotemporal Controls on the Urban Aerobiome. *Front. Ecol. Evol.* **7**, (2019).
2. Tignat-Perrier, R. *et al.* Global airborne microbial communities controlled by surrounding landscapes and wind conditions. *Scientific Reports* (in revision).
3. Uetake, J. *et al.* Seasonal changes of airborne bacterial communities over Tokyo and influence of local meteorology. *bioRxiv* 542001 (2019). doi:10.1101/542001
4. Innocente, E. *et al.* Influence of seasonality, air mass origin and particulate matter chemical composition on airborne bacterial community structure in the Po Valley, Italy. *Sci. Total Environ.* **593–594**, 677–687 (2017).
5. Bowers, R. M., McCubbin, I. B., Hallar, A. G. & Fierer, N. Seasonal variability in airborne bacterial communities at a high-elevation site. *Atmospheric Environment* **50**, 41–49 (2012).
6. Bowers, R. M. *et al.* Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. *Environ. Sci. Technol.* **47**, 12097–12106 (2013).
7. Franzetti, A., Gandolfi, I., Gaspari, E., Ambrosini, R. & Bestetti, G. Seasonal variability of bacteria in fine and coarse urban air particulate matter. *Appl. Microbiol. Biotechnol.* **90**, 745–753 (2011).
8. Cáliz, J., Triadó-Margarit, X., Camarero, L. & Casamayor, E. O. A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12229–12234 (2018).
9. Dommergue, A. *et al.* Methods to investigate the global atmospheric microbiome. *Front. Microbiol.* **10**, (2019).
10. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**, 31 (2012).
11. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology* **73**, 5261–5267 (2007).
12. Nguyen, N. H. *et al.* FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology* **20**, 241–248 (2016).
13. Draxler, R. R. & Hess, G. D. An Overview of the HYSPLIT_4 Modelling System for Trajectories, Dispersion, and Deposition. 25
14. Carslaw, D. Tools for the Analysis of Air Pollution Data. (2019).
15. Dray, S., Dufour, A.-B. & Thioulouse, J. Analysis of Ecological Data: Exploratory and Euclidean Methodsin Environmental Science. (2018).

16. Harrell, F. E. & Dupont, C. Harrell Miscellaneous - Package 'Hmisc'. (2019).
17. Constancias, F. *et al.* Mapping and determinism of soil microbial community distribution across an agricultural landscape. *Microbiologyopen* **4**, 505–517 (2015).
18. Zhang, Q. *et al.* Alterations in soil microbial community composition and biomass following agricultural land use change. *Scientific Reports* **6**, 36587 (2016).
19. Sun, S., Li, S., Avera, B. N., Strahm, B. D. & Badgley, B. D. Soil Bacterial and Fungal Communities Show Distinct Recovery Patterns during Forest Ecosystem Restoration. *Appl Environ Microbiol* **83**, (2017).
20. Pickersgill, D. A. *et al.* Lifestyle dependent occurrence of airborne fungi. *Biogeosciences Discussions* 1–20 (2017). doi:https://doi.org/10.5194/bg-2017-452
21. Dennis, P. G., Newsham, K. K., Rushton, S. P., O'Donnell, A. G. & Hopkins, D. W. Soil bacterial diversity is positively associated with air temperature in the maritime Antarctic. *Sci Rep* **9**, 1–11 (2019).
22. Zhou, J. *et al.* Temperature mediates continental-scale diversity of microbes in forest soils. *Nature Communications* **7**, 12083 (2016).
23. Nottingham, A. T. *et al.* Microbes follow Humboldt: temperature drives plant and soil microbial diversity patterns from the Amazon to the Andes. *Ecology* **99**, 2455–2466 (2018).
24. Tittensor, D. P. *et al.* Global patterns and predictors of marine biodiversity across taxa. *Nature* **466**, 1098–1101 (2010).
25. Venzac, H., Sellegri, K., Villani, P., Picard, D. & Laj, P. Seasonal variation of aerosol size distributions in the free troposphere and residual layer at the puy de Dôme station, France. *Atmospheric Chemistry and Physics* **9**, 1465–1478 (2009).
26. Bourcier, L., Sellegri, K., Chausse, P., M. Pichon, J. & Laj, P. Seasonal variation of water-soluble inorganic components in aerosol size-segregated at the puy de Dôme station (1,465 m a.s.l.), France. *Journal of Atmospheric Chemistry* **69**, (2012).
27. Freney, E. J. *et al.* Seasonal variations in aerosol particle composition at the puy-de-Dôme research station in France. *Atmospheric Chemistry and Physics* **11**, 13047–13059 (2011).
28. Medeiros, P. M., Conte, M. H., Weber, J. C. & Simoneit, B. R. T. Sugars as source indicators of biogenic organic carbon in aerosols collected above the Howland Experimental Forest, Maine. *Atmospheric Environment* **40**, 1694–1705 (2006).
29. Ruijter, G. J. G. *et al.* Mannitol Is Required for Stress Tolerance in Aspergillus niger Conidiospores. *Eukaryot Cell* **2**, 690–698 (2003).
30. Solomon, P. S., Waters, O. D. C. & Oliver, R. P. Decoding the mannitol enigma in filamentous fungi. *Trends Microbiol.* **15**, 257–262 (2007).
31. Lewis, D. H. & Smith, D. C. Sugar alcohols (polyols) in fungi and green plants. *New Phytologist* **66**, 143–184 (1967).

Chapter 4: Seasonal changes in local landscapes drive airborne microbial community variation

Romie Tignat-Perrier[1,2], Aurélien Dommergue[1], Alban Thollot[1], Olivier Magand[1], Pierre Amato[3], Timothy M. Vogel[2], Catherine Larose[2]

[1]Institut des Géosciences de l'Environnement, Université Grenoble Alpes, CNRS, IRD, Grenoble INP, Grenoble, France

[2]Environmental Microbial Genomics, CNRS UMR 5005 Laboratoire Ampère, École Centrale de Lyon, Université de Lyon, Écully, France

[3]Institut de Chimie de Clermont-Ferrand, UMR6096 CNRS Université Clermont Auvergne-Sigma, Clermont-Ferrand, France

Supplementary Information

Supplementary tables

Table S1: Sampling starting date, collected air volume and number of sequences obtained after the 16s rRNA gene and ITS sequencing and quality filtering. Blank cells represent samples which have been removed from the dataset because of low read number (< 6000 reads).

Sample number	Sample name	Sampling starting date (month/day/year). End date: 7 days later, same hour.	Collected standardized air volume (m^3)	Number of sequences for the 16s rRNA gene sequencing after quality filtering	Number of sequences for the ITS sequencing after quality filtering
PDD_S01	PDD_29/06/2016	06/29/16	8610	63115	15813
PDD_S02	PDD_06/07/2016	07/06/16	8230	51661	17016
PDD_S03	PDD_13/07/2016	07/13/16	8591	51967	27368
PDD_S04	PDD_20/07/2016	07/20/16	8769	82471	31202
PDD_S06	PDD_03/08/2016	08/03/16	8447	41648	21891
PDD_S07	PDD_10/08/2016	08/10/16	8476	50687	25038
PDD_S08	PDD_17/08/2016	08/17/16	8552	39687	24683
PDD_S09	PDD_24/08/2016	08/24/16	8817	57870	20751
PDD_S10	PDD_01/09/2016	09/01/16	estimated 8600	42068	25941
PDD_S11	PDD_07/09/2016	09/07/16	8578	45669	39370
PDD_S12	PDD_14/09/2016	09/14/16	8675	59341	26614
PDD_S13	PDD_21/09/2016	09/21/16	8463	58601	42164
PDD_S14	PDD_05/10/2016	10/05/16	8577	17726	81962
PDD_S15	PDD_12/10/2016	10/12/16	8627	59389	34328
PDD_S16	PDD_19/10/2016	10/19/16	8787	58277	26900
PDD_S17	PDD_26/10/2016	10/26/16	8826	63880	8240
PDD_S18	PDD_02/11/2016	11/02/16	14745	48262	35500
PDD_2-S01	PDD_16/11/2016	11/16/16	10346	51196	33643
PDD_2-S02	PDD_24/11/2016	11/24/16	7784	53821	13462
PDD_2-S03	PDD_30/11/2016	11/30/16	9077	66750	19034
PDD_2-S04	PDD_07/12/2016	12/07/16	8959	35240	11621
PDD_2-S05	PDD_14/12/2016	12/14/16	8755	88674	15678
PDD_2-S06	PDD_28/12/2016	12/28/16	9253	60388	8357
PDD_2-S07	PDD_21/12/2016	12/21/16	8985	55503	32651
PDD_2-S08	PDD_04/01/2017	01/04/17	18506	51159	37955
PDD_2-S09	PDD_19/01/2017	01/19/17	8571	30716	38356
PDD_2-S10	PDD_25/01/2017	01/25/17	8824	33035	
PDD_2-S11	PDD_01/02/2017	02/01/17	9080	78861	
PDD_2-S12	PDD_08/02/2017	02/08/17	7812	66259	60697
PDD_2-S13	PDD_23/02/2017	02/23/17	16094	52339	
PDD_2-S14	PDD_08/03/2017	03/08/17	7888	71333	34979
PDD_2-S15	PDD_14/03/2017	03/14/17	10041	65448	
PDD_2-S16	PDD_22/03/2017	03/22/17	8485	58594	

PDD_2-S17	PDD_29/03/2017	03/29/17	8967	49703	10955
PDD_2-S18	PDD_05/04/2017	04/05/17	9345	46706	26383
PDD_2-S19	PDD_12/04/2017	04/12/17	8244	33819	27646
PDD_2-S20	PDD_19/04/2017	04/19/17	6377	73626	17930
PDD_2-S21	PDD_26/04/2017	04/26/17	8841	41224	34559
PDD_2-S22	PDD_10/05/2017	05/10/17	9091	45674	24725
PDD_2-S23	PDD_17/05/2017	05/17/17	8652	67192	24698
PDD_2-S24	PDD_24/05/2017	05/24/17	7415	70375	22394
PDD_3-S01	PDD_30/05/2017	05/30/17	10232	58483	22260
PDD_3-S02	PDD_07/06/2017	06/07/17	8761	47217	20653
PDD_3-S03	PDD_14/06/2017	06/14/17	8360	50094	20060
PDD_3-S04	PDD_21/06/2017	06/21/17	8672	116853	18010
PDD_3-S05	PDD_28/06/2017	06/28/17	9012		29150
PDD_3-S06	PDD_05/07/2017	07/05/17	8034	37218	22442
PDD_3-S07	PDD_12/07/2017	07/12/17	8713	36945	24629
PDD_3-S08	PDD_19/07/2017	07/19/17	8620	50864	20170
PDD_3-S09	PDD_26/07/2017	07/26/17	8664	37942	25343
PDD_3-S10	PDD_02/08/2017	08/02/17	7399	35652	34048
PDD_3-S11	PDD_09/08/2017	08/09/17	9926	28142	26101
PDD_3-S12	PDD_16/08/2017	08/16/17	8657	37329	36039

Table S2: Bacterial and fungal concentration (number of 16S rRNA and 18S rRNA gene copies/m^3) and richness (number of bacterial genera and fungal species) averaged per season. Reference letters indicate the group membership based on Tukey's HSD post hoc tests.

Season	Average bacterial concentration (number of 16S rRNA gene copies/m^3 of air)	Average fungal concentration (number of 18S rRNA gene copies/m^3 of air)	Average Chao1 bacterial richness estimation (estimated number of bacterial genera)	Average Chao1 fungal richness estimation (estimated number of fungal species)
PDD summer	$8.1x10^5$ +- $1.3x10^6$	$1.2x10^4$ +- $2.0x10^4$	$6.2x10^2$ +- $1.4x10^{2\,a}$	$4.7x10^2$ +- $1.8x10^2$
PDD autumn	$1.1x10^5$ +- $1.6x10^5$	$2.2x10^3$ +- $2.9x10^3$	$4.0x10^2$ +- $1.0x10^{2\,b}$	$3.5x10^2$ +- $2.0x10^2$
PDD winter	$3.0x10^5$ +- $8.1x10^5$	$6.9x10^2$ +- $8.9x10^2$	$4.0x10^2$ +- $1.3x10^{2\,b}$	$2.8x10^2$ +- $1.4x10^2$
PDD spring	$2.9x10^6$ +- $6.1x10^6$	$1.4x10^4$ +- $3.1x10^4$	$4.9x10^2$ +- $1.5x10^{2\,ab}$	$4.6x10^2$ +- $1.0x10^2$

Table S3: Top 25 of the bacterial genera and fungal species observed in Puy-de-Dôme, and average percentages of these bacterial genera and fungal species per season.

Bacterial genus	Percentage of each genus (%)	Autumn (%)	Spring (%)	Summer (%)	Winter (%)	Fungal species	Percentage of each genus (%)	Autumn (%)	Spring (%)	Summer (%)	Winter (%)
Bacillus	15.6	13.6	18.8	17	13.9	*Pseudotaeniolina.globosa*	11.5	16.4	7.7	5.3	18.7
Hymenobacter	9.2	19.2	3.5	2.7	13.9	*Cladophialophora.proteae*	7.9	15.6	5.1	0.3	13.3
Sphingomonas	7.4	3.8	3.6	12.5	2.6	*Alternaria.sp.BMP_2012*	4	1	1.9	10.8	0.5
Methylobacterium	5.1	6.4	2.2	2.9	8.9	*Cladophialophora.minutissima*	4	2.6	7.4	0.3	6.2
Thermoactinomyces	4.6	7.1	4.6	1.2	8.5	*Naevala.minutissima*	3.1	0	10.1	1.8	0
Romboutsia	4.5	4.4	8.2	4.3	3.5	*Erysiphe.alphitoides*	3	0	4.9	6.1	0
Clostridium_sensu_stricto	3.1	2.7	7	2.1	3.8	*Preussia.minima*	2.9	2.6	2.4	2.1	4.7
Turicibacter	2	1.8	4	1.6	2	*Botryotinia.fuckeliana*	2.6	0.9	3.2	4.7	0.8
Sporosarcina	1.9	1.9	2.6	2.1	1.4	*Sarcinomyces.petricola*	2.1	4.1	1.6	0.1	3.2
Lysinibacillus	1.6	1.3	1.8	1.7	1.5	*Eurotium.amstelodami*	1.9	0.6	0.7	0.1	6.4
Pseudomonas	1.4	0.4	0.3	2.9	0.1	*Epicoccum.sp.JJP_2009a*	1.8	2.5	0.5	3.6	0.2
Domibacillus	1.4	0.8	0.4	1.2	2.9	*Baeospora.myosura*	1.7	6	0	0	1.7
Tumebacillus	1.4	0.7	1.1	1.9	1.2	*Tremella.foliacea*	1.7	3.3	0.5	1.1	2.3
Clostridium_XI	1.4	1.4	2.4	1.3	1.1	*Leptosphaerulina.chartarum*	1.5	1.6	0.2	3.4	0.3
Nakamurella	1.3	0.9	1.1	0.3	3.4	*Aspergillus.reptans*	1.5	0.4	2	0.2	3.6
Acidisphaera	1.2	0.9	1	0.1	3.7	*Handkea.utriformis*	1.3	0.1	0.6	1.1	3.5
Pedobacter	1.1	0.9	0.3	1.8	0.3	*Cryptococcus.victoriae*	1.2	0.3	0.5	3.4	0.2
Paenibacillus	1.1	0.9	1.3	1.1	1	*Psilolechia.leprosa*	1.2	1.2	1.2	1	1.4
Anaerobacter	0.9	0.8	2	0.6	1	*Blumeria.graminis*	1.1	0	3.9	0.1	0.6
Terrisporobacter	0.8	0.7	2	0.5	1.1	*Ustilago.bullata*	1	0	0.1	3.2	0
Massilia	0.8	0.5	0.3	1.3	0.3	*Microdochium.majus*	1	0	0.1	3.2	0
Rhodopila	0.7	1.6	0.3	0.1	1.2	*Podospora.pleiospora*	0.9	1.4	1.3	1	0
Neorhizobium	0.7	0.4	0.2	1.3	0	*Preussia.intermedia*	0.9	0.5	1.5	0.2	1.4
Acidicaldus	0.6	1.1	0.3	0.1	1.3	*Mycosphaerella.graminicola*	0.8	0	0	2.6	0
Rathayibacter	0.5	0.3	0.1	1	0.1	*Aureobasidium.pullulans*	0.8	0.2	0.7	1.6	0.3

Table S4: Weekly variability of the bacterial and fungal composition per season ((1 – Bray-Curtis dissimilarity) / standard deviation). The maximum wind speed (m/s) per season and the variability of the wind direction (degree), relative humidity (%) and temperature (°C) within a week and between the weeks in each season are also shown.

Site	Bacterial populations similarity between weeks	Fungal populations similarity between weeks	Maximum wind speed (m/s)	Wind direction variability within weeks (degree)	Wind direction variability between weeks (degree)	Relative humidity variability within weeks (%)	Relative humidity variability between weeks (%)	Temperature variability within weeks (°C)	Temperature variability between weeks (°C)
PDD summer	5.6	3.5	26.7	83.9	34.0	17.0	9.7	3.9	2.6
PDD autumn	6.2	2.7	26.3	80.8	42.2	20.4	23.2	3.1	3.0
PDD winter	5.3	4.1	31.1	86.9	38.2	21.3	18.3	3.3	3.1
PDD spring	5.6	4.4	25.0	94.2	27.6	19.0	9.0	3.9	5.5

Table S5: Richness in fungal species annotated as pathotroph, symbiotroph and saprotroph averaged per season and according to FUNGuild.

Season	Averaged number of pathotrophs	Averaged number of saprotrophs	Averaged number of symbiotrophs
Fall	39	141	26
Winter	25	100	10
Spring	84	165	21
Summer	92	145	23

Table S6: Grouping of the fifty most abundant bacterial genera and fungal species based of the trend of their relative abundance over the year. The first trend groups taxa whose relative abundances increase in fall and/or winter; the second trend groups taxa whose relative abundances increase in spring and/or summer; and the third trend groups taxa whose relative abundances remain stable or are highly variable over the year. For fungal species, the trophic mode of each taxon has been given based on the FUNGuild software.

Grouping of the fifty most abundant fungal species in three general trends

Trend 1	Trophic mode	Trend 2	Trophic mode	Trend 3	Trophic mode
Pseudotaeniolina globosa	symbiotroph	*Aternaria sp BMP_2012*	patotroph	*Torula caligans*	pathotroph-saprotroph
Cladophialophora proteae	saprotroph	*Botryotinia fickeliana*	patotroph	*Podospora tetraspora*	saprotroph
Cladophialophora minutissima	saprotroph	*Erysiphe alphitoides*	patotroph	*Schizopora radula*	saprotroph
Sarcinomyces petricola	pathotroph-saprotroph	*Naevala minutissima*	pathotroph-saprotroph	*Stereum annosum*	saprotroph
Baeospora myosura	saprotroph	*Epicoccum sp JJP_2009a*	pathotroph-saprotroph	*Phaeosphaeria herpotrichoides*	saprotroph
Tremella foliacea	pathotroph	*Leptosphaerulina chartarum*	symbiotroph	*Pyrenochaeta sp CF_2008*	saprotroph
Blumeria graminis	pathotroph	*Cryptococcus victoriae*	pathotroph	*Trichocladium asperum*	saprotroph
Aspergillus reptans	saprotroph	*Podospora pleiospora*	pathotroph-saprotroph-symbiotroph	*Preussia minima*	saprotroph
Eurotium amstelodami	saprotroph	*Aureobasidium pullulans*	pathotroph-saprotroph-symbiotroph	*Ustilago bullata*	saprotroph
Preussia intermedia	saprotroph	*Microdochium majus*	pathotroph-saprotroph	*Humicola nigrescens*	saprotroph
Psilolechia leprosa	symbiotroph	*Epicoccum nigrum*	pathotroph-saprotroph-symbiotroph		
Coprinopsis radiata	saprotroph	*Pseudozyma shanxiensis*	patotroph		
Blumeria graminis f sp hordei	pathotroph	*Handkea utriformis*	saprotroph		
Strobilurus albipilatus	saprotroph	*Pleospora alfalfae*	patotroph		
Phlebia radiata	saprotroph	*Dioszegia crocea*	pathotroph-saprotroph-symbiotroph		
Hyphodontia alutaria	saprotroph	*Mycosphaerella graminicola*	patotroph		
Sarcomyxa serotina	saprotroph	*Bovista plumbea*	saprotroph		
		Sporidiobolus sp JPS_2007a	pathotroph-saprotroph		
		Phoma paspali	pathotroph-saprotroph-symbiotroph		
		Ustilago hordei	patotroph		
		Alternaria brassicae	pathotroph-saprotroph-symbiotroph		
		Melampsora epitea	patotroph		
		Cryptococcus chernovii	pathotroph-saprotroph-symbiotroph		

Grouping of the fifty most abundant bacterial genera in three general trends

Trend 1	Trend 2	Trend 3
Beijerinckia	*Brevundimonas*	*Gaeilla*
Spirosoma	*Dyabobacter*	*Flavobacterium*
Laceyella	*Solibacillus*	*Arthrobacterium*
Acidicaldus	*Bradyrhizobium*	*Planococcus*
Rhodopila	*Terrisporobacter*	*Aureimonas*
Acidisphaera	*Gemmatimonas*	*Sediminibacter*
Nakamurella	*Clavibacter*	*Ammoniphilus*
Thermoactinomyces	*Devosia*	*Caryophanon*

Methylobacterium
Hymenobacter
Sodalis
Kineococcus
Rhdococcus
Rathayibacter
Neorhizobium
Massilia
Anaerobacter
Pedobacter
Clostridium XI
Turicibacter
Clostridium sensu stricto
Romboutsia
Sphingomonas
Staphylococcus
Phenylobacterium
Gpl
Nocardium
Domibacillus
Paenibacillus
Tumebacillus
Lysinibacillus
Pseudomonas
Sporosarcina
Bacillus

Table S7: Chemical concentrations averaged per season.

Chemical compounds (ng/m^3)	spring	summer	autumn	winter
OC	1222.4 +/- 699.89	1441.9 +/- 451.88	589.51 +/- 228.28	442.05 +/- 165.07
EC	0.12 +/- 0.05	0.13 +/- 0.04	0.06 +/- 0.02	0.06 +/- 0.04
MSA	29.95 +/- 16.21	30.01 +/- 16.41	2.57 +/- 1.5	3.85 +/- 3.33
Cl	43.28 +/- 55.17	37.44 +/- 69.41	35.12 +/- 42.77	65.11 +/- 91.36
NO3	1049.25 +/- 963.99	566.21 +/- 698.62	587.9 +/- 631.32	989.55 +/- 1544.53
SO4	986.51 +/- 349.45	960.7 +/- 359.65	489.17 +/- 204.62	496.54 +/- 234.4
Ox	31.48 +/- 39.01	133.53 +/- 79.81	54.76 +/- 25.79	30.06 +/- 19.49
Na	143.4 +/- 60.28	191.88 +/- 129.69	78.35 +/- 53.12	116.08 +/- 98.66
NH4	413.17 +/- 247.85	221.57 +/- 150.61	208.42 +/- 229.22	309.84 +/- 385.57
K	37.19 +/- 10.02	51.27 +/- 36.69	23.48 +/- 7.79	26.24 +/- 13.88
Mg	21.12 +/- 8.12	34.68 +/- 32.5	12.61 +/- 9.7	12.79 +/- 11.03
Ca	92.76 +/- 43.81	146.52 +/- 133.27	44.6 +/- 41.16	37.26 +/- 32.25
Inositol	3.79 +/- 5.76	1.69 +/- 2.4	0.17 +/- 0.08	0.26 +/- 0.16
Glycerol	3.94 +/- 4.52	0.64 +/- 1.54	0.11 +/- 0.34	0.32 +/- 0.79
Erythriol	1.8 +/- 0.73	1.3 +/- 0.7	0.66 +/- 0.57	0.76 +/- 0.64
Xylitol	0.38 +/- 0.4	1.39 +/- 1.02	0.19 +/- 0.2	0.08 +/- 0.08
Mannitol+arabitol	14.46 +/- 12.66	20.98 +/- 11.71	2.52 +/- 1.62	2.4 +/- 2.12
Sorbitol	4.91 +/- 6.72	8.08 +/- 4.78	0.65 +/- 0.31	0.46 +/- 0.44
Threalose	0.2 +/- 0.67	4.24 +/- 3.24	0.92 +/- 1	0.02 +/- 0.08
Levoglucosan+mannosan	12.04 +/- 9.69	3.22 +/- 2.68	14.18 +/- 10.36	22.43 +/- 14.89
Galactosan	0.33 +/- 0.3	0.11 +/- 0.18	0.54 +/- 0.48	0.87 +/- 0.5
Rhamnose	0.04 +/- 0.07	0.01 +/- 0.03	0 +/- 0	0.04 +/- 0.06
Glucose	18.8 +/- 11.97	13.83 +/- 9.17	2.55 +/- 1.33	3.15 +/- 3.47

Table S8: Output (R^2 and pvalue) of the correlation tests between bacterial genera (a.) or fungal species (b.) in relative abundance and chemical species (log10 of the concentration in ng/m^3). Only the 50 most abundant bacterial genera and fungal species of the dataset have been tested. Spearman correlations have been used. Correlations with an R^2 superior to 0.5 and inferior to -0.5 have been highlighted in red and yellow, respectively.

a.

Bacterial genus	R2														Pvalue												
	OC	EC	MSA	Ox	K	Mg	Ca	xylitol	Mannitol +arabitol	sorbitol	threalose	levo,manno	galactosan	glucose	OC	EC	MSA	Ox	K	Mg	Ca	xylitol	Mannitol arabitol	sorbitol	threalose	levo,manno	galactosan	glucose
Bacillus	0.12	0.15	0.08	-0.16	0.09	0.04	0.08	0.16	0.07	0.16	-0.01	-0.06	-0.04	0.13	4.2E-01	3.0E-01	5.7E-01	2.6E-01	5.2E-01	7.7E-01	5.6E-01	2.6E-01	6.2E-01	2.7E-01	9.3E-01	6.8E-01	7.6E-01	3.6E-01
Hymenobacter	-0.43	-0.38	-0.35	-0.09	-0.31	-0.26	-0.27	-0.38	-0.36	-0.46	-0.18	0.19	0.22	-0.38	1.5E-03	5.8E-03	1.1E-02	5.1E-01	2.8E-02	6.8E-02	5.3E-02	6.0E-03	8.5E-03	6.2E-04	2.0E-01	1.8E-01	1.3E-01	5.7E-03
Sphingomonas	0.43	0.32	0.42	0.44	0.23	0.24	0.27	0.45	0.48	0.49	0.45	-0.48	-0.53	0.30	1.5E-03	2.2E-02	2.2E-03	1.2E-03	1.1E-01	9.5E-02	5.8E-02	9.5E-04	3.8E-04	2.4E-04	1.0E-03	3.5E-04	6.4E-05	3.2E-02
Methylobacterium	-0.38	-0.30	-0.34	0.03	-0.22	-0.16	-0.20	-0.34	-0.34	-0.41	-0.06	0.12	0.20	-0.36	5.7E-03	3.1E-02	1.4E-02	8.4E-01	1.2E-01	2.6E-01	1.6E-01	1.5E-02	1.4E-02	3.0E-03	6.8E-01	3.9E-01	1.7E-01	9.2E-03
Romboutsia	0.26	0.32	0.24	-0.24	0.22	0.05	0.15	0.16	0.22	0.29	-0.13	0.08	0.04	0.32	6.3E-02	2.4E-02	8.7E-02	9.1E-02	1.2E-01	7.2E-01	2.9E-01	2.8E-01	1.2E-01	4.0E-02	3.8E-01	5.6E-01	7.8E-01	2.4E-02
Thermoactinomyces	-0.40	-0.22	-0.45	-0.42	-0.33	-0.42	-0.46	-0.48	-0.47	-0.44	-0.44	0.68	0.68	-0.34	3.4E-03	1.1E-01	9.0E-04	2.1E-03	1.9E-02	2.0E-03	7.4E-04	3.9E-04	5.0E-04	1.4E-03	1.4E-03	5.1E-08	5.5E-08	1.4E-02
Clostridium_sensu_stricto	-0.05	0.05	0.00	-0.38	-0.02	-0.10	-0.09	-0.17	-0.09	-0.07	-0.31	0.34	0.32	0.07	7.3E-01	7.3E-01	9.9E-01	5.7E-03	8.8E-01	4.7E-01	5.5E-01	2.5E-01	5.5E-01	6.4E-01	2.7E-02	1.5E-02	2.3E-02	6.4E-01
Turicibacter	0.09	0.16	0.11	-0.39	0.13	0.03	0.04	-0.01	0.09	0.09	-0.31	0.25	0.23	0.25	5.4E-01	2.5E-01	4.2E-01	4.7E-03	3.8E-01	8.3E-01	7.7E-01	9.2E-01	5.4E-01	5.4E-01	2.5E-02	7.7E-02	1.1E-01	7.3E-02
Sporosarcina	0.17	0.21	0.09	-0.20	0.14	0.02	0.12	0.16	0.14	0.19	-0.01	-0.08	-0.10	0.20	2.4E-01	1.5E-01	5.4E-01	1.5E-01	3.3E-01	8.6E-01	4.1E-01	2.5E-01	3.4E-01	1.7E-01	9.5E-01	5.6E-01	4.9E-01	1.6E-01
Pseudomonas	0.41	0.24	0.49	0.33	0.23	0.33	0.30	0.54	0.54	0.56	0.38	-0.62	-0.61	0.36	3.0E-03	9.0E-02	2.3E-04	1.9E-02	9.8E-02	2.0E-02	3.2E-02	5.0E-05	4.9E-05	2.0E-05	6.2E-03	1.5E-06	1.9E-06	1.0E-02
Lysinibacillus	0.10	0.17	-0.03	-0.13	0.11	0.00	0.08	0.08	0.00	0.07	-0.08	0.01	0.02	0.03	4.7E-01	2.4E-01	8.4E-01	3.7E-01	4.6E-01	9.8E-01	5.6E-01	5.9E-01	9.8E-01	6.4E-01	5.9E-01	9.6E-01	8.6E-01	8.1E-01
Clostridium_XI	0.29	0.34	0.27	-0.26	0.21	0.06	0.14	0.16	0.25	0.33	-0.12	0.05	0.01	0.35	3.8E-02	1.5E-02	5.8E-02	6.5E-02	1.4E-01	6.9E-01	3.2E-01	2.7E-01	7.7E-02	1.8E-02	4.2E-01	7.3E-01	9.2E-01	1.1E-02
Nakamurella	-0.43	-0.29	-0.37	-0.34	-0.27	-0.33	-0.31	-0.54	-0.45	-0.52	-0.38	0.38	0.35	-0.33	1.7E-03	4.0E-02	7.1E-03	1.3E-02	5.9E-02	1.8E-02	2.7E-02	3.7E-05	8.2E-04	1.1E-04	6.3E-03	5.9E-03	1.2E-02	1.8E-02
Acidisphaera	-0.55	-0.40	-0.51	-0.51	-0.31	-0.37	-0.35	-0.65	-0.60	-0.69	-0.59	0.58	0.64	-0.39	3.0E-05	4.0E-03	1.1E-04	1.2E-04	2.6E-02	8.1E-03	1.1E-02	3.0E-07	3.8E-06	1.9E-08	6.0E-06	8.9E-06	5.1E-07	5.3E-03
Tumebacillus	0.25	0.24	0.12	0.01	0.14	0.06	0.12	0.25	0.13	0.22	0.13	-0.06	-0.09	0.13	8.1E-02	8.7E-02	3.9E-01	9.3E-01	3.3E-01	6.9E-01	4.1E-01	8.2E-02	3.6E-01	1.2E-01	3.8E-01	6.8E-01	5.4E-01	3.5E-01
Pedobacter	0.34	0.24	0.34	0.44	0.15	0.21	0.22	0.42	0.41	0.44	0.48	-0.43	-0.50	0.23	1.6E-02	8.4E-02	1.6E-02	1.3E-03	2.9E-01	1.4E-01	1.2E-01	2.0E-03	3.1E-03	1.1E-03	3.3E-04	1.8E-03	1.9E-04	1.1E-01
Anaerobacter	0.02	0.10	0.07	-0.44	0.02	-0.07	-0.02	-0.18	-0.04	-0.05	-0.38	0.36	0.33	0.18	9.0E-01	4.9E-01	6.4E-01	1.1E-03	9.0E-01	6.2E-01	8.8E-01	2.1E-01	7.9E-01	7.4E-01	6.2E-03	9.4E-03	2.0E-02	2.2E-01
Paenibacillus	0.15	0.19	0.13	-0.10	0.07	-0.02	0.01	0.16	0.10	0.19	0.00	-0.04	-0.04	0.11	2.9E-01	1.7E-01	3.8E-01	4.7E-01	6.3E-01	8.8E-01	9.6E-01	2.5E-01	4.8E-01	1.9E-01	9.9E-01	8.1E-01	7.8E-01	4.5E-01
Domibacillus	-0.36	-0.25	-0.37	-0.14	-0.18	-0.18	-0.16	-0.26	-0.34	-0.38	-0.20	0.30	0.36	-0.27	9.3E-03	7.1E-02	8.1E-03	3.3E-01	2.0E-01	2.0E-01	2.6E-01	6.9E-02	1.5E-02	6.7E-03	1.6E-01	3.0E-02	9.4E-03	6.0E-02
Rhodopila	-0.56	-0.47	-0.53	-0.37	-0.37	-0.34	-0.30	-0.60	-0.64	-0.67	-0.47	0.51	0.59	-0.43	1.6E-05	4.8E-04	6.7E-05	7.3E-03	7.0E-03	1.4E-02	3.4E-02	2.7E-06	5.2E-07	6.8E-08	5.1E-04	1.3E-04	6.4E-06	1.5E-03
Terrisporobacter	-0.10	0.00	0.00	-0.37	-0.07	-0.14	-0.17	-0.17	-0.10	-0.07	-0.39	0.39	0.36	0.03	4.9E-01	9.9E-01	9.8E-01	8.2E-03	6.3E-01	3.2E-01	2.3E-01	2.4E-01	5.1E-01	6.0E-01	4.5E-03	4.9E-03	1.0E-02	8.1E-01
Acidicaldus	-0.56	-0.43	-0.51	-0.27	-0.33	-0.27	-0.29	-0.55	-0.58	-0.68	-0.41	0.43	0.48	-0.45	2.1E-05	1.8E-03	1.5E-04	6.0E-02	1.9E-02	5.2E-02	3.8E-02	3.3E-05	7.1E-06	3.0E-08	2.8E-03	1.6E-03	3.8E-04	8.4E-04
Massilia	0.42	0.34	0.39	0.46	0.27	0.30	0.34	0.50	0.49	0.50	0.49	-0.48	-0.54	0.33	2.2E-03	1.3E-02	4.8E-03	7.6E-04	5.4E-02	3.4E-02	1.6E-02	2.1E-04	2.6E-04	2.2E-04	3.0E-04	4.1E-04	4.8E-05	1.6E-02
Neorhizobium	0.37	0.28	0.40	0.49	0.30	0.36	0.36	0.49	0.48	0.49	0.61	-0.48	-0.51	0.25	8.4E-03	4.6E-02	3.5E-03	2.5E-04	3.2E-02	9.0E-03	8.6E-03	2.8E-04	3.8E-04	2.6E-04	2.5E-06	3.6E-04	1.2E-04	7.7E-02
Nocardioides	0.40	0.34	0.29	0.38	0.24	0.27	0.32	0.30	0.36	0.39	0.36	-0.29	-0.37	0.29	3.6E-03	1.6E-02	3.6E-02	6.1E-03	9.7E-02	5.7E-02	2.3E-02	3.1E-02	9.6E-03	5.2E-03	9.4E-03	3.7E-02	7.8E-03	4.0E-02
Rathayibacter	0.48	0.39	0.41	0.46	0.29	0.26	0.29	0.49	0.50	0.54	0.52	-0.49	-0.54	0.28	3.9E-04	5.2E-03	2.6E-03	7.1E-04	4.2E-02	7.0E-02	3.8E-02	2.4E-04	1.7E-04	5.2E-05	8.6E-05	2.2E-04	4.1E-05	4.4E-02
Gpl	0.22	0.19	0.05	-0.03	-0.01	-0.11	0.05	-0.04	0.02	0.10	0.00	-0.03	-0.06	0.16	1.2E-01	1.8E-01	7.3E-01	8.2E-01	9.5E-01	4.4E-01	7.0E-01	7.8E-01	8.7E-01	5.0E-01	9.9E-01	8.2E-01	6.8E-01	2.8E-01
Laceyella	-0.38	-0.19	-0.42	-0.31	-0.28	-0.37	-0.42	-0.42	-0.44	-0.41	-0.33	0.67	0.64	-0.34	6.1E-03	1.7E-01	2.0E-03	2.5E-02	4.5E-02	8.2E-03	2.2E-03	2.1E-03	1.4E-03	2.7E-03	1.7E-02	9.6E-08	4.9E-07	1.4E-02
Phenylobacterium	-0.53	-0.47	-0.47	0.00	-0.52	-0.51	-0.52	-0.45	-0.57	-0.51	-0.13	0.43	0.45	-0.60	6.5E-05	5.5E-04	5.3E-04	9.9E-01	7.9E-05	1.6E-04	9.9E-05	1.0E-03	1.0E-05	1.5E-04	3.6E-01	1.6E-03	8.3E-04	3.8E-06
Rhodococcus	0.20	0.16	0.23	0.44	0.21	0.21	0.24	0.37	0.29	0.28	0.35	-0.37	-0.39	0.09	1.6E-01	2.5E-01	9.9E-02	1.3E-03	1.5E-01	1.4E-01	8.7E-02	7.7E-03	3.6E-02	4.8E-02	1.2E-02	7.7E-03	5.0E-03	5.3E-01
Caryophanon	0.29	0.35	0.10	-0.13	0.23	0.09	0.17	0.21	0.16	0.20	-0.13	-0.10	-0.08	0.24	4.2E-02	1.3E-02	4.9E-01	3.5E-01	9.9E-02	5.4E-01	2.3E-01	1.4E-01	2.6E-01	1.5E-01	3.8E-01	4.9E-01	5.9E-01	9.7E-02

Ammoniphilus	0.06	0.10	-0.02	-0.09	0.08	0.04	0.06	0.14	0.05	0.08	-0.02	0.00	-0.01	0.07	6.6E-01	4.6E-01	9.0E-01	5.1E-01	5.8E-01	8.0E-01	6.7E-01	3.1E-01	7.4E-01	5.9E-01	8.7E-01	9.9E-01	9.2E-01	6.2E-01
Sediminibacterium	-0.42	-0.41	-0.37	0.10	-0.50	-0.48	-0.41	-0.36	-0.50	-0.40	-0.08	0.38	0.40	-0.56	2.0E-03	2.9E-03	7.3E-03	5.1E-01	2.2E-04	3.3E-04	2.5E-03	9.7E-03	2.1E-04	3.6E-03	5.7E-01	5.8E-03	3.9E-03	1.9E-05
Aureimonas	0.32	0.27	0.32	0.43	0.20	0.24	0.24	0.40	0.38	0.39	0.40	-0.41	-0.46	0.21	2.0E-02	6.0E-02	2.1E-02	1.7E-03	1.6E-01	8.5E-02	9.7E-02	4.0E-03	5.5E-03	4.6E-03	3.8E-03	3.1E-03	7.3E-04	1.4E-01
Planococcaceae_incertae_sedis	0.16	0.19	0.01	-0.31	0.13	-0.02	0.06	0.03	0.04	0.09	-0.18	0.13	0.14	0.14	2.6E-01	1.8E-01	9.3E-01	2.6E-02	3.7E-01	8.6E-01	6.7E-01	8.4E-01	7.8E-01	5.1E-01	2.0E-01	3.6E-01	3.4E-01	3.3E-01
Kineococcus	0.46	0.38	0.38	0.42	0.34	0.38	0.35	0.47	0.53	0.51	0.49	-0.51	-0.53	0.34	6.8E-04	5.8E-03	5.5E-03	2.2E-03	1.4E-02	6.6E-03	1.3E-02	4.6E-04	6.9E-05	1.5E-04	2.5E-04	1.2E-04	6.4E-05	1.4E-02
Staphylococcus	-0.04	-0.12	-0.10	0.18	-0.19	-0.07	-0.12	0.01	-0.05	0.05	0.23	-0.07	-0.12	-0.15	7.8E-01	3.9E-01	4.9E-01	2.1E-01	1.9E-01	6.4E-01	4.1E-01	9.4E-01	7.1E-01	7.1E-01	1.1E-01	6.0E-01	4.1E-01	3.1E-01
Sodalis	0.21	0.21	0.09	0.57	0.27	0.31	0.33	0.43	0.24	0.28	0.58	-0.38	-0.35	0.04	1.5E-01	1.3E-01	5.4E-01	1.3E-05	5.4E-02	2.8E-02	1.7E-02	1.4E-03	9.7E-02	4.3E-02	7.1E-06	5.4E-03	1.1E-02	7.7E-01
Devosia	0.37	0.28	0.43	0.42	0.21	0.29	0.30	0.38	0.43	0.41	0.41	-0.39	-0.48	0.30	6.8E-03	4.6E-02	1.7E-03	2.1E-03	1.4E-01	3.8E-02	3.2E-02	5.7E-03	1.7E-03	3.1E-03	2.8E-03	4.5E-03	3.5E-04	3.4E-02
Spirosoma	0.12	0.12	0.14	0.14	-0.01	0.00	0.03	0.04	0.12	0.17	0.18	-0.22	-0.19	0.05	4.1E-01	3.9E-01	3.2E-01	3.4E-01	9.4E-01	1.0E+00	8.5E-01	7.9E-01	4.1E-01	2.4E-01	2.0E-01	1.3E-01	1.7E-01	7.2E-01
Arthrobacter	0.32	0.29	0.25	0.28	0.22	0.21	0.22	0.23	0.31	0.31	0.37	-0.26	-0.34	0.19	2.3E-02	3.9E-02	8.2E-02	4.9E-02	1.3E-01	1.4E-01	1.2E-01	1.1E-01	2.8E-02	2.7E-02	6.9E-03	6.5E-02	1.4E-02	1.8E-01
Flavobacterium	0.40	0.35	0.38	0.24	0.20	0.14	0.21	0.34	0.41	0.47	0.29	-0.36	-0.41	0.29	3.4E-03	1.2E-02	5.5E-03	8.8E-02	1.5E-01	3.1E-01	1.4E-01	1.6E-02	2.6E-03	5.3E-04	3.6E-02	1.0E-02	2.5E-03	3.9E-02
Clavibacter	0.54	0.48	0.41	0.49	0.39	0.35	0.40	0.60	0.58	0.58	0.59	-0.48	-0.56	0.35	4.8E-05	3.2E-04	2.9E-03	2.4E-04	4.8E-03	1.3E-02	3.5E-03	3.7E-06	7.0E-06	8.0E-06	6.1E-06	4.2E-04	2.1E-05	1.1E-02
Gemmatimonas	0.30	0.30	0.11	0.13	0.15	0.12	0.15	0.07	0.25	0.22	0.14	-0.11	-0.18	0.24	3.1E-02	3.3E-02	4.3E-01	3.8E-01	3.0E-01	4.2E-01	2.8E-01	6.0E-01	8.1E-02	1.3E-01	3.2E-01	4.5E-01	2.1E-01	8.8E-02
Gaiella	0.31	0.22	0.29	0.20	0.20	0.22	0.26	0.18	0.32	0.33	0.28	-0.33	-0.40	0.26	2.5E-02	1.3E-01	3.9E-02	1.6E-01	1.7E-01	1.1E-01	6.4E-02	2.1E-01	2.2E-02	1.9E-02	4.9E-02	1.8E-02	3.6E-03	6.7E-02
Bradyrhizobium	-0.19	-0.21	-0.25	0.06	-0.35	-0.35	-0.33	-0.26	-0.30	-0.15	0.04	0.15	0.17	-0.44	1.8E-01	1.5E-01	7.7E-02	6.6E-01	1.3E-02	1.1E-02	1.7E-02	6.8E-02	3.1E-02	3.1E-01	7.6E-01	2.9E-01	2.3E-01	1.1E-03
Solibacillus	0.29	0.25	0.18	-0.13	0.31	0.25	0.32	0.29	0.26	0.27	0.02	-0.15	-0.14	0.36	3.8E-02	7.1E-02	2.1E-01	3.6E-01	2.9E-02	7.3E-02	2.3E-02	3.6E-02	6.5E-02	5.6E-02	8.6E-01	3.1E-01	3.2E-01	8.5E-03
Dyadobacter	0.36	0.31	0.27	0.35	0.17	0.19	0.22	0.37	0.38	0.42	0.52	-0.40	-0.48	0.20	9.0E-03	2.8E-02	5.9E-02	1.2E-02	2.2E-01	1.8E-01	1.1E-01	6.8E-03	6.5E-03	2.4E-03	7.8E-05	3.8E-03	3.8E-04	1.6E-01
Brevundimonas	0.30	0.21	0.31	0.40	0.08	0.18	0.11	0.34	0.37	0.40	0.32	-0.34	-0.41	0.20	3.5E-02	1.4E-01	2.4E-02	3.5E-03	5.6E-01	2.2E-01	4.3E-01	1.4E-02	7.2E-03	3.4E-03	2.0E-02	1.6E-02	2.7E-03	1.6E-01
Beijerinckia	-0.55	-0.51	-0.56	-0.17	-0.38	-0.34	-0.43	-0.42	-0.54	-0.60	-0.41	0.39	0.45	-0.51	3.0E-05	1.3E-04	1.9E-05	2.3E-01	5.4E-03	1.5E-02	1.6E-03	1.9E-03	4.6E-05	3.3E-06	3.1E-03	4.1E-03	9.9E-04	1.5E-04

b.

fungal species	R2																				Pvalue																	
	OC	EC	MSA	Cl	SO4	Ox	Na	K	Mg	Ca	Inositol	Glycerol	Erythriol	xylitol	Mannitol arabitol	sorbitol	threalose	Levo manno	galactosan	glucose	OC	EC	MSA	Cl	SO4	Ox	Na	K	Mg	Ca	Inositol	Glycerol	Erythriol	xylitol	Mannitol arabitol	sorbitol	threalose	Levo manno	galactosan	glucose
Pseudotaenioli na globosa	-0.53	-0.45	-0.54	0.30	-0.39	-0.44	-0.20	-0.35	-0.27	-0.29	-0.44	0.15	-0.43	-0.62	-0.66	-0.70	-0.49	0.52	0.66	-0.47	1.3E-04	1.4E-03	1.0E-04	3.7E-02	6.6E-03	2.1E-03	1.7E-01	1.6E-02	6.3E-02	5.1E-02	1.7E-03	3.3E-01	2.9E-03	2.8E-06	5.6E-07	5.8E-08	5.5E-04	1.5E-04	5.2E-07	9.0E-04
Alternaria sp BMP_2012	0.60	0.53	0.45	-0.11	0.29	0.43	0.33	0.55	0.48	0.47	0.41	-0.15	0.22	0.66	0.73	0.68	0.47	-0.57	-0.64	0.53	7.2E-06	1.2E-04	1.4E-03	4.7E-01	5.2E-02	2.8E-03	2.5E-02	5.3E-05	6.3E-04	7.7E-04	3.9E-03	3.2E-01	1.4E-01	4.1E-07	4.8E-09	1.5E-07	9.5E-04	2.8E-05	1.1E-06	1.2E-04
Cladophialoph ora proteae	-0.54	-0.47	-0.52	0.17	-0.40	-0.50	-0.31	-0.44	-0.38	-0.38	-0.36	0.18	-0.29	-0.68	-0.66	-0.71	-0.54	0.50	0.61	-0.45	1.0E-04	8.8E-04	1.7E-04	2.6E-01	5.4E-03	3.9E-04	3.3E-02	2.2E-03	8.0E-03	8.4E-03	1.4E-02	2.3E-01	5.0E-02	1.9E-07	4.6E-07	2.7E-08	8.3E-05	3.3E-04	5.7E-06	1.4E-03
Botryotinia fuckeliana	0.54	0.42	0.50	-0.07	0.23	0.27	0.35	0.47	0.46	0.44	0.58	0.19	0.36	0.44	0.64	0.59	0.24	-0.47	-0.48	0.59	1.0E-04	3.5E-03	3.9E-04	6.2E-01	1.1E-01	6.8E-02	1.5E-02	8.7E-04	1.2E-03	1.8E-03	2.0E-05	2.0E-01	1.3E-02	2.0E-03	1.2E-06	1.1E-05	1.1E-01	8.1E-04	6.5E-04	1.2E-05
Erysiphe alphitoides	0.58	0.41	0.68	-0.21	0.45	0.20	0.13	0.35	0.25	0.30	0.57	-0.20	0.45	0.61	0.68	0.78	0.30	-0.64	-0.66	0.55	1.9E-05	4.0E-03	1.5E-07	1.6E-01	1.5E-03	1.8E-01	4.0E-01	1.6E-02	9.2E-02	4.1E-02	3.2E-05	1.7E-01	1.6E-03	5.8E-06	1.6E-07	7.5E-11	4.3E-02	1.2E-06	4.1E-07	6.0E-05
Preussia minima	-0.26	-0.27	-0.24	0.24	-0.16	-0.30	0.05	-0.15	-0.07	-0.17	-0.12	0.20	-0.25	-0.21	-0.15	-0.21	-0.09	0.22	0.29	-0.06	7.2E-02	6.3E-02	1.0E-01	9.9E-02	2.8E-01	3.8E-02	7.6E-01	3.1E-01	6.5E-01	2.6E-01	4.2E-01	1.9E-01	9.1E-02	1.6E-01	3.3E-01	1.5E-01	5.7E-01	1.4E-01	5.1E-02	7.0E-01
Naevala minutissima	0.21	0.02	0.38	-0.02	0.10	-0.15	-0.02	0.12	0.05	0.12	0.54	0.24	0.41	0.11	0.24	0.23	-0.29	-0.16	-0.17	0.39	1.6E-01	8.8E-01	8.1E-03	8.7E-01	5.1E-01	3.0E-01	9.0E-01	4.4E-01	7.3E-01	4.2E-01	8.6E-05	1.0E-01	4.7E-03	4.8E-01	1.0E-01	1.3E-01	4.6E-02	2.9E-01	2.5E-01	6.1E-03
Epicoccum sp JJP_2009a	0.32	0.30	0.03	-0.14	0.10	0.47	0.23	0.32	0.30	0.26	0.00	-0.17	-0.01	0.47	0.40	0.39	0.50	-0.28	-0.35	0.16	2.6E-02	3.8E-02	8.2E-01	3.5E-01	5.1E-01	8.1E-04	1.1E-01	2.7E-02	4.1E-02	8.1E-02	9.8E-01	2.6E-01	9.6E-01	9.6E-04	5.4E-03	6.4E-03	3.3E-04	5.5E-02	1.6E-02	2.8E-01

Cladophialophora minutissima	-0.40	-0.32	-0.35	0.26	-0.22	-0.50	-0.19	-0.25	-0.27	-0.21	-0.21	0.28	-0.17	-0.60	-0.57	-0.60	-0.67	0.56	0.60	-0.29
Leptosphaerulina chartarum	0.31	0.30	0.07	0.01	0.10	0.49	0.33	0.39	0.39	0.34	-0.02	-0.16	-0.05	0.44	0.41	0.39	0.54	-0.23	-0.28	0.18
Cryptococcus victoriae	0.49	0.43	0.43	-0.14	0.28	0.47	0.30	0.47	0.40	0.43	0.33	-0.24	0.26	0.64	0.60	0.64	0.52	-0.56	-0.59	0.35
Sarcinomyces petricola	-0.50	-0.43	-0.51	0.17	-0.35	-0.51	-0.28	-0.43	-0.38	-0.38	-0.36	0.18	-0.25	-0.68	-0.67	-0.68	-0.55	0.60	0.67	-0.45
Baeospora myosura	-0.62	-0.49	-0.58	0.19	-0.41	-0.20	-0.12	-0.47	-0.23	-0.38	-0.56	0.04	-0.38	-0.55	-0.63	-0.58	-0.04	0.41	0.54	-0.58
Ustilago bullata	0.31	0.11	0.33	-0.07	0.13	0.09	0.10	0.14	0.16	0.10	0.48	0.02	0.33	0.32	0.35	0.37	-0.22	-0.40	-0.30	0.33
Tremella foliacea	-0.25	-0.22	-0.32	0.24	-0.28	-0.07	-0.21	-0.15	-0.13	-0.07	-0.31	-0.03	-0.35	-0.18	-0.35	-0.36	-0.02	0.03	0.15	-0.32
Podospora pleiospora	0.33	0.25	0.33	0.00	0.38	0.02	0.23	0.34	0.24	0.32	0.35	0.01	0.17	0.41	0.43	0.45	0.16	-0.28	-0.23	0.42
Aureobasidium pullulans	0.50	0.44	0.47	-0.07	0.38	0.34	0.47	0.54	0.45	0.48	0.32	-0.11	0.20	0.55	0.57	0.61	0.40	-0.51	-0.56	0.41
Blumeria graminis	0.47	0.49	0.51	0.17	0.54	-0.23	0.32	0.44	0.31	0.44	0.66	0.42	0.51	0.10	0.39	0.37	-0.24	-0.03	-0.09	0.63
Microdochium majus	0.44	0.51	0.28	-0.09	0.31	0.37	0.21	0.39	0.27	0.37	0.19	0.09	0.27	0.38	0.36	0.39	0.48	-0.27	-0.32	0.25
Epicoccum nigrum	0.32	0.27	0.08	-0.12	0.12	0.55	0.26	0.31	0.31	0.27	-0.01	-0.18	0.02	0.48	0.40	0.41	0.58	-0.29	-0.36	0.17
Aspergillus reptans	-0.21	-0.09	-0.27	0.05	-0.10	-0.42	-0.09	-0.19	-0.25	-0.26	-0.06	0.31	0.06	-0.41	-0.32	-0.30	-0.34	0.49	0.42	-0.11
Pseudozyma shanxiensis	0.41	0.30	0.34	-0.02	0.25	0.03	0.18	0.35	0.31	0.33	0.51	0.12	0.30	0.34	0.47	0.56	0.12	-0.47	-0.40	0.50
Eurotium amstelodami	-0.36	-0.20	-0.46	0.05	-0.22	-0.36	-0.11	-0.32	-0.29	-0.32	-0.29	0.34	-0.11	-0.55	-0.49	-0.49	-0.25	0.55	0.51	-0.30
Handkea utriformis	0.16	0.08	0.31	0.15	0.30	-0.25	0.26	0.15	0.22	0.16	0.44	0.34	0.20	0.07	0.28	0.29	-0.16	-0.11	0.01	0.43
Preussia intermedia	0.00	0.00	0.09	-0.01	0.15	-0.47	0.02	-0.02	-0.04	0.00	0.26	0.51	0.20	-0.24	-0.06	-0.06	-0.42	0.17	0.21	0.19
Psilolechia leprosa	-0.45	-0.38	-0.40	0.51	-0.32	-0.35	-0.06	-0.23	-0.12	-0.14	-0.29	0.25	-0.31	-0.54	-0.54	-0.59	-0.41	0.44	0.57	-0.33
Coprinopsis radiata	0.02	0.05	0.01	0.33	0.04	-0.41	0.10	0.05	0.03	-0.01	0.28	0.40	0.21	-0.26	-0.05	-0.08	-0.39	0.32	0.34	0.23
Mycosphaerella graminicola	0.18	0.25	0.22	-0.04	0.15	0.33	0.10	0.23	0.15	0.21	0.14	-0.23	0.20	0.39	0.29	0.31	0.39	-0.26	-0.31	0.11
Pleospora alfalfae	0.58	0.55	0.47	-0.07	0.41	0.44	0.55	0.61	0.57	0.56	0.34	-0.01	0.20	0.61	0.70	0.64	0.49	-0.47	-0.56	0.52
Dioszegia crocea	0.47	0.36	0.57	0.04	0.41	0.28	0.29	0.50	0.38	0.52	0.42	-0.16	0.30	0.61	0.58	0.64	0.47	-0.55	-0.57	0.45

Cladophialophora minutissima	4.9E-03	3.0E-02	1.6E-02	7.3E-02	1.4E-01	3.5E-04	2.0E-01	8.6E-02	6.8E-02	1.5E-01	1.6E-01	6.0E-02	2.5E-01	9.8E-06	3.3E-05	8.4E-06	3.0E-07	4.0E-05	7.4E-06	4.6E-02
Leptosphaerulina chartarum	3.3E-02	3.8E-02	6.5E-01	9.3E-01	5.1E-01	4.6E-04	2.2E-02	6.1E-03	6.3E-03	1.8E-02	8.7E-01	2.8E-01	7.4E-01	2.0E-03	4.2E-03	7.3E-03	1.0E-04	1.3E-01	6.0E-02	2.4E-01
Cryptococcus victoriae	4.8E-04	2.9E-03	2.6E-03	3.6E-01	5.4E-02	9.8E-04	4.0E-02	9.0E-04	4.9E-03	2.5E-03	2.2E-02	1.1E-01	7.6E-02	1.0E-06	7.9E-06	1.5E-06	1.6E-04	4.0E-05	1.1E-05	1.6E-02
Sarcinomyces petricola	3.4E-04	2.8E-03	2.3E-04	2.4E-01	1.6E-02	2.3E-04	5.8E-02	2.8E-03	7.8E-03	7.7E-03	1.3E-02	2.2E-01	9.4E-02	1.3E-07	3.4E-07	1.2E-07	7.4E-05	9.8E-06	2.2E-07	1.6E-03
Baeospora myosura	2.9E-06	4.1E-04	1.6E-05	2.0E-01	4.0E-03	1.8E-01	4.3E-01	8.6E-04	1.2E-01	7.7E-03	4.8E-05	7.7E-01	9.2E-03	6.4E-05	2.5E-06	2.3E-05	7.8E-01	4.0E-03	9.7E-05	2.2E-05
Ustilago bullata	3.6E-02	4.5E-01	2.5E-02	6.4E-01	3.9E-01	5.7E-01	5.1E-01	3.6E-01	2.7E-01	4.9E-01	5.8E-04	9.0E-01	2.2E-02	2.8E-02	1.6E-02	1.0E-02	1.4E-01	5.7E-03	3.9E-02	2.6E-02
Tremella foliacea	8.9E-02	1.3E-01	2.6E-02	1.0E-01	5.7E-02	6.5E-01	1.6E-01	3.1E-01	3.7E-01	6.4E-01	3.3E-02	8.6E-01	1.6E-02	2.3E-01	1.4E-02	1.2E-02	8.7E-01	8.4E-01	3.1E-01	3.0E-02
Podospora pleiospora	2.2E-02	8.7E-02	2.6E-02	9.7E-01	8.5E-03	8.9E-01	1.3E-01	2.1E-02	9.8E-02	2.8E-02	1.6E-02	9.5E-01	2.7E-01	4.2E-03	2.7E-03	1.6E-03	2.7E-01	5.5E-02	1.3E-01	3.3E-03
Aureobasidium pullulans	3.0E-04	1.8E-03	8.0E-04	6.2E-01	8.9E-03	1.9E-02	7.5E-04	8.6E-05	1.4E-03	6.5E-04	2.8E-02	4.8E-01	1.8E-01	6.5E-05	3.4E-05	6.5E-06	5.9E-03	2.1E-04	3.8E-05	4.1E-03
Blumeria graminis	9.5E-04	5.5E-04	2.9E-04	2.4E-01	9.0E-05	1.2E-01	3.0E-02	1.7E-03	3.7E-02	2.2E-03	5.5E-07	3.7E-03	2.7E-04	4.9E-01	7.1E-03	1.0E-02	1.1E-01	8.4E-01	5.7E-01	1.8E-06
Microdochium majus	2.0E-03	2.2E-04	5.8E-02	5.5E-01	3.6E-02	1.1E-02	1.5E-01	6.6E-03	6.7E-02	1.1E-02	2.1E-01	5.4E-01	6.4E-02	8.6E-03	1.4E-02	6.9E-03	6.1E-04	7.2E-02	2.8E-02	9.3E-02
Epicoccum nigrum	2.9E-02	6.7E-02	5.8E-01	4.2E-01	4.1E-01	6.9E-05	8.1E-02	3.6E-02	3.6E-02	7.0E-02	9.4E-01	2.3E-01	9.1E-01	6.6E-04	4.8E-03	3.8E-03	2.1E-05	4.7E-02	1.2E-02	2.4E-01
Aspergillus reptans	1.6E-01	5.3E-01	6.8E-02	7.6E-01	4.8E-01	3.0E-03	5.3E-01	1.9E-01	9.4E-02	7.7E-02	6.7E-01	3.4E-02	7.0E-01	3.9E-03	3.0E-02	4.0E-02	1.9E-02	5.4E-04	3.1E-03	4.7E-01
Pseudozyma shanxiensis	4.5E-03	4.1E-02	2.0E-02	8.7E-01	8.9E-02	8.3E-01	2.3E-01	1.6E-02	3.6E-02	2.2E-02	2.3E-04	4.4E-01	4.3E-02	2.1E-02	8.9E-04	3.8E-05	4.2E-01	9.0E-04	5.7E-03	3.7E-04
Eurotium amstelodami	1.3E-02	1.8E-01	1.1E-03	7.6E-01	1.3E-01	1.2E-02	4.8E-01	3.0E-02	5.1E-02	2.9E-02	4.7E-02	1.9E-02	4.6E-01	6.7E-05	5.1E-04	5.0E-04	8.8E-02	5.8E-05	2.5E-04	3.7E-02
Handkea utriformis	2.7E-01	5.7E-01	3.1E-02	3.2E-01	3.9E-02	8.7E-02	8.2E-02	3.2E-01	1.5E-01	2.8E-01	2.2E-03	1.9E-02	1.7E-01	6.6E-01	6.1E-02	4.7E-02	2.8E-01	4.8E-01	9.7E-01	2.8E-03
Preussia intermedia	9.8E-01	9.9E-01	5.6E-01	9.4E-01	3.2E-01	7.5E-04	8.8E-01	9.2E-01	8.1E-01	9.8E-01	8.2E-02	2.7E-04	1.7E-01	1.1E-01	6.6E-01	7.0E-01	3.1E-03	2.6E-01	1.7E-01	2.0E-01
Psilolechia leprosa	1.5E-03	8.3E-03	5.8E-03	2.3E-04	2.6E-02	1.6E-02	6.8E-01	1.1E-01	4.3E-01	3.4E-01	4.8E-02	9.4E-02	3.2E-02	8.1E-05	1.0E-04	1.2E-05	4.4E-03	2.2E-03	3.5E-05	2.3E-02
Coprinopsis radiata	9.0E-01	7.2E-01	9.6E-01	2.5E-02	8.1E-01	4.3E-03	5.0E-01	7.2E-01	8.2E-01	9.5E-01	5.9E-02	5.9E-03	1.5E-01	7.8E-02	7.4E-01	5.9E-01	6.4E-03	2.9E-02	2.0E-02	1.2E-01
Mycosphaerella graminicola	2.3E-01	9.7E-02	1.3E-01	8.2E-01	3.3E-01	2.4E-02	5.1E-01	1.2E-01	3.3E-01	1.5E-01	3.5E-01	1.2E-01	1.9E-01	7.4E-03	4.6E-02	3.2E-02	6.7E-03	7.9E-02	3.5E-02	4.4E-01
Pleospora alfalfae	2.0E-05	5.6E-05	9.7E-04	6.2E-01	3.9E-03	2.1E-03	5.5E-05	6.0E-06	2.8E-05	5.1E-05	1.9E-02	9.4E-01	1.7E-01	4.5E-06	5.1E-08	9.9E-07	4.3E-04	9.1E-04	4.4E-05	2.1E-04
Dioszegia crocea	9.4E-04	1.3E-02	3.2E-05	7.8E-01	4.5E-03	6.1E-02	5.1E-02	3.3E-04	8.2E-03	1.6E-04	3.5E-03	2.9E-01	3.9E-02	5.8E-06	2.1E-05	1.2E-06	8.4E-04	5.6E-05	2.9E-05	1.6E-03

Trichocladium asperum	0.13	0.18	-0.02	0.17	0.18	-0.11	0.06	0.05	-0.04	0.00	-0.02	0.00	-0.02	0.00	-0.04	0.07	-0.11	0.28	0.21	0.05	3.8E-01	2.3E-01	9.1E-01	2.5E-01	2.3E-01	4.5E-01	6.6E-01	7.2E-01	8.0E-01	9.8E-01	9.0E-01	9.9E-01	8.7E-01	9.8E-01	7.7E-01	6.6E-01	4.7E-01	6.1E-02	1.6E-01	7.2E-01
Blumeria graminis f sp hordei	0.27	0.27	0.35	0.06	0.28	-0.21	0.17	0.19	0.12	0.20	0.41	0.33	0.43	-0.01	0.25	0.27	-0.06	0.10	0.09	0.45	6.6E-02	6.2E-02	1.6E-02	6.7E-01	6.1E-02	1.7E-01	2.6E-01	1.9E-01	4.3E-01	1.9E-01	3.8E-03	2.3E-02	2.4E-03	9.5E-01	8.7E-02	6.5E-02	7.0E-01	4.9E-01	5.4E-01	1.6E-03
Pyrenochaeta sp CF_2008	0.30	0.25	0.36	-0.19	0.21	-0.47	-0.07	0.08	-0.10	0.02	0.43	0.23	0.29	-0.13	0.17	0.19	-0.35	-0.10	-0.13	0.39	4.1E-02	8.7E-02	1.2E-02	2.0E-01	1.6E-01	8.1E-04	6.4E-01	6.1E-01	5.0E-01	9.0E-01	2.3E-03	1.2E-01	5.2E-02	3.8E-01	2.6E-01	1.9E-01	1.7E-02	5.1E-01	3.9E-01	7.4E-03
Bovista plumbea	0.25	0.21	0.30	0.10	0.30	-0.23	0.25	0.21	0.26	0.15	0.41	0.34	0.21	0.02	0.27	0.30	-0.12	-0.11	-0.01	0.41	9.3E-02	1.6E-01	4.3E-02	5.2E-01	4.3E-02	1.3E-01	9.1E-02	1.5E-01	8.3E-02	3.3E-01	4.4E-03	1.8E-02	1.5E-01	8.8E-01	6.9E-02	3.8E-02	4.3E-01	4.8E-01	9.4E-01	4.5E-03
Strobilurus albipilatus	-0.51	-0.39	-0.50	0.02	-0.40	-0.33	-0.21	-0.44	-0.37	-0.39	-0.35	0.21	-0.12	-0.62	-0.66	-0.65	-0.35	0.54	0.53	-0.49	2.2E-04	7.5E-03	3.8E-04	8.7E-01	5.4E-03	2.4E-02	1.7E-01	2.0E-03	1.2E-02	6.3E-03	1.6E-02	1.6E-01	4.2E-01	3.3E-06	4.5E-07	7.1E-07	1.5E-02	8.4E-05	1.1E-04	5.2E-04
Phaeosphaeria herpotrichoides	0.18	0.15	0.14	-0.28	0.12	0.15	0.13	0.20	0.16	0.19	0.29	0.29	0.32	0.16	0.20	0.28	0.12	-0.21	-0.22	0.24	2.3E-01	3.1E-01	3.6E-01	5.7E-02	4.4E-01	3.3E-01	3.8E-01	1.7E-01	2.7E-01	1.9E-01	4.7E-02	4.9E-02	2.8E-02	2.9E-01	1.8E-01	5.7E-02	4.3E-01	1.6E-01	1.4E-01	1.0E-01
Humicola nigrescens	-0.01	0.09	-0.06	0.16	0.13	-0.16	0.01	-0.04	-0.11	-0.05	-0.11	0.18	-0.06	-0.04	-0.15	-0.06	-0.11	0.23	0.17	-0.05	9.6E-01	5.4E-01	6.9E-01	2.8E-01	3.8E-01	3.0E-01	9.5E-01	7.7E-01	4.7E-01	7.3E-01	4.4E-01	2.2E-01	6.9E-01	7.8E-01	3.2E-01	6.8E-01	4.8E-01	1.3E-01	2.5E-01	7.2E-01
Sporidiobolus sp JPS_2007a	0.44	0.31	0.39	0.04	0.37	0.39	0.26	0.43	0.35	0.32	0.33	-0.29	0.15	0.61	0.53	0.59	0.30	-0.48	-0.47	0.33	1.9E-03	3.5E-02	6.1E-03	7.9E-01	1.0E-02	6.7E-03	7.2E-02	2.5E-03	1.7E-02	2.9E-02	2.3E-02	4.5E-02	3.3E-01	4.4E-06	1.5E-04	1.5E-05	4.2E-02	6.8E-04	9.0E-04	2.5E-02
Stereum annosum	-0.24	-0.25	-0.24	0.22	-0.12	0.05	-0.04	-0.09	-0.01	0.01	-0.14	0.12	-0.27	-0.03	-0.19	-0.15	-0.07	0.05	0.20	-0.18	1.0E-01	9.0E-02	1.1E-01	1.4E-01	4.2E-01	7.3E-01	8.0E-01	5.6E-01	9.4E-01	9.4E-01	3.3E-01	4.1E-01	6.6E-02	8.6E-01	1.9E-01	3.1E-01	6.3E-01	7.1E-01	1.8E-01	2.2E-01
Phoma paspali	0.41	0.29	0.43	-0.13	0.21	0.46	0.21	0.41	0.40	0.42	0.31	-0.22	0.13	0.64	0.60	0.62	0.55	-0.77	-0.75	0.35	4.2E-03	4.5E-02	2.7E-03	3.7E-01	1.6E-01	1.0E-03	1.5E-01	3.9E-03	5.8E-03	3.2E-03	3.4E-02	1.4E-01	3.8E-01	1.4E-06	8.4E-06	4.1E-06	5.3E-05	2.5E-10	1.5E-09	1.5E-02
Ustilago hordei	0.32	0.13	0.42	-0.14	0.25	-0.10	-0.04	0.10	0.01	0.08	0.55	0.00	0.45	0.22	0.34	0.39	-0.29	-0.26	-0.25	0.37	2.8E-02	3.8E-01	3.5E-03	3.5E-01	9.3E-02	4.9E-01	7.9E-01	5.1E-01	9.6E-01	6.0E-01	5.9E-05	1.0E+00	1.5E-03	1.3E-01	1.8E-02	7.3E-03	4.4E-02	7.3E-02	8.7E-02	1.0E-02
Schizopora radula	-0.20	-0.22	-0.18	0.24	-0.19	0.00	0.14	0.03	0.15	0.09	-0.05	0.18	-0.29	-0.10	-0.13	-0.15	-0.11	0.08	0.24	-0.11	1.7E-01	1.4E-01	2.2E-01	1.0E-01	2.0E-01	1.0E+00	3.4E-01	8.6E-01	3.1E-01	5.3E-01	7.5E-01	2.2E-01	4.7E-02	4.9E-01	3.8E-01	3.2E-01	4.7E-01	5.8E-01	9.8E-02	4.4E-01
Podospora tetraspora	0.16	0.19	0.13	0.15	0.31	-0.32	0.14	0.16	0.03	0.13	0.26	0.20	0.19	0.07	0.12	0.20	-0.09	-0.04	-0.02	0.27	2.8E-01	1.9E-01	3.7E-01	3.2E-01	3.5E-02	2.7E-02	3.5E-01	2.9E-01	8.2E-01	3.8E-01	7.4E-02	1.7E-01	1.9E-01	6.6E-01	4.4E-01	1.7E-01	5.5E-01	7.7E-01	9.1E-01	6.6E-02
Torula caligans	0.33	0.38	0.15	-0.01	0.09	0.24	0.32	0.37	0.41	0.37	0.22	0.13	0.20	0.21	0.35	0.28	0.24	-0.14	-0.22	0.30	2.4E-02	8.7E-03	3.2E-01	9.5E-01	5.3E-01	1.1E-01	3.0E-02	1.0E-02	4.2E-03	9.9E-03	1.4E-01	3.8E-01	1.9E-01	1.5E-01	1.6E-02	5.8E-02	1.0E-01	3.4E-01	1.4E-01	3.9E-02
Alternaria brassicae	0.29	0.27	0.26	-0.06	0.01	0.43	0.29	0.32	0.36	0.37	0.13	-0.20	-0.09	0.49	0.46	0.38	0.50	-0.67	-0.70	0.21	4.7E-02	7.1E-02	7.8E-02	6.8E-01	9.6E-01	2.5E-03	5.1E-02	2.8E-02	1.2E-02	1.0E-02	3.8E-01	1.8E-01	5.5E-01	4.2E-04	1.2E-03	8.7E-03	3.4E-04	2.1E-07	3.9E-08	1.5E-01
Phlebia radiata	-0.29	-0.26	-0.41	0.15	-0.29	-0.01	0.07	-0.16	0.04	-0.11	-0.30	0.11	-0.42	-0.31	-0.29	-0.29	-0.02	0.10	0.30	-0.32	4.5E-02	7.4E-02	4.4E-03	3.0E-01	4.7E-02	9.6E-01	6.2E-01	2.8E-01	7.8E-01	4.7E-01	4.2E-02	4.6E-01	3.7E-03	3.5E-02	4.7E-02	4.7E-02	8.7E-01	4.8E-01	4.0E-02	3.1E-02
Hyphodontia alutaria	-0.26	-0.32	-0.37	-0.02	-0.35	0.19	-0.16	-0.23	-0.18	-0.20	-0.32	-0.08	-0.31	-0.13	-0.28	-0.21	0.03	0.17	0.28	-0.39	7.6E-02	3.1E-02	1.1E-02	8.8E-01	1.8E-02	2.0E-01	2.9E-01	1.2E-01	2.3E-01	1.8E-01	3.1E-02	6.1E-01	3.2E-02	3.9E-01	5.8E-02	1.5E-01	8.4E-01	2.5E-01	6.1E-02	7.2E-03
Melampsora epitea	0.49	0.46	0.31	-0.02	0.37	0.49	0.37	0.45	0.40	0.42	0.12	-0.15	0.10	0.57	0.46	0.47	0.46	-0.36	-0.45	0.34	4.2E-04	1.0E-03	3.1E-02	9.0E-01	9.8E-03	4.4E-04	1.0E-02	1.4E-03	4.8E-03	2.9E-03	4.1E-01	3.2E-01	4.9E-01	2.6E-05	1.3E-03	8.9E-04	1.3E-03	1.3E-02	1.6E-03	1.9E-02
Cryptococcus chernovii	0.56	0.53	0.41	-0.11	0.37	0.43	0.37	0.53	0.43	0.46	0.33	-0.21	0.35	0.63	0.62	0.65	0.56	-0.47	-0.52	0.39	4.9E-05	1.4E-04	3.8E-03	4.6E-01	1.1E-02	2.5E-03	9.7E-03	1.3E-04	2.4E-03	1.3E-03	2.3E-02	1.5E-01	1.7E-02	2.3E-06	3.6E-06	6.7E-07	4.4E-05	8.4E-04	1.5E-04	7.4E-03
Sarcomyxa serotina	-0.43	-0.34	-0.62	-0.13	-0.41	-0.11	-0.27	-0.38	-0.34	-0.46	-0.48	-0.13	-0.20	-0.39	-0.53	-0.50	-0.12	0.35	0.42	-0.54	2.6E-03	2.0E-02	3.7E-06	3.8E-01	3.8E-03	4.5E-01	6.9E-02	7.7E-03	1.9E-02	1.2E-03	6.4E-04	3.8E-01	1.7E-01	6.9E-03	1.4E-04	3.7E-04	4.3E-01	1.5E-02	3.4E-03	8.2E-05

Table S9: Output (R^2 and pvalue) of the correlation tests between bacterial and fungal estimated richness (Chao1 value) or concentration (qPCR based) and meteorological parameters. Spearman correlations have been used. Correlations with an R^2 superior to 0.5 (or inferior to -0.5) have been highlighted in red.

Meteorological parameter	Bacterial estimated richness: Chao1 value (number of estimated bacterial genera)		Fungal estimated richness: Chao1 value (number of estimated fungal species)		Bacterial concentration: 16S rRNA gene qPCR value (number of gene copies per cubic meter of air)		Fungal concentration: 18S rRNA gene qPCR value (number of gene copies per cubic meter of air)	
	R2	**Pvalue**	**R2**	**Pvalue**	**R2**	**Pvalue**	**R2**	**Pvalue**
Relative Humidity_mean (%)	-0.41	2.50E-03	0.10	5.14E-01	-0.20	1.6E-01	-0.16	2.8E-01
Temperature_max (°C)	0.52	7.75E-05	0.22	1.38E-01	0.23	9.8E-02	0.30	3.2E-02
Temperature_mean (°C)	0.47	5.53E-04	0.22	1.44E-01	0.21	1.4E-01	0.33	1.7E-02
Temperature_min (°C)	0.40	3.96E-03	0.21	1.57E-01	0.19	1.7E-01	0.38	5.6E-03
UV radiation_max	0.28	4.79E-02	0.36	1.30E-02	0.38	5.4E-03	0.27	5.5E-02
Wind direction_mean (°)	0.06	6.76E-01	0.51	2.91E-04	0.10	4.7E-01	0.22	1.3E-01
Wind speed_max (m/s)	-0.24	8.60E-02	0.21	1.64E-01	0.19	1.8E-01	0.24	8.4E-02
Wind speed_mean (m/s)	-0.16	2.50E-01	0.36	1.24E-02	-0.01	9.2E-01	0.04	7.6E-01

Supplementary figures

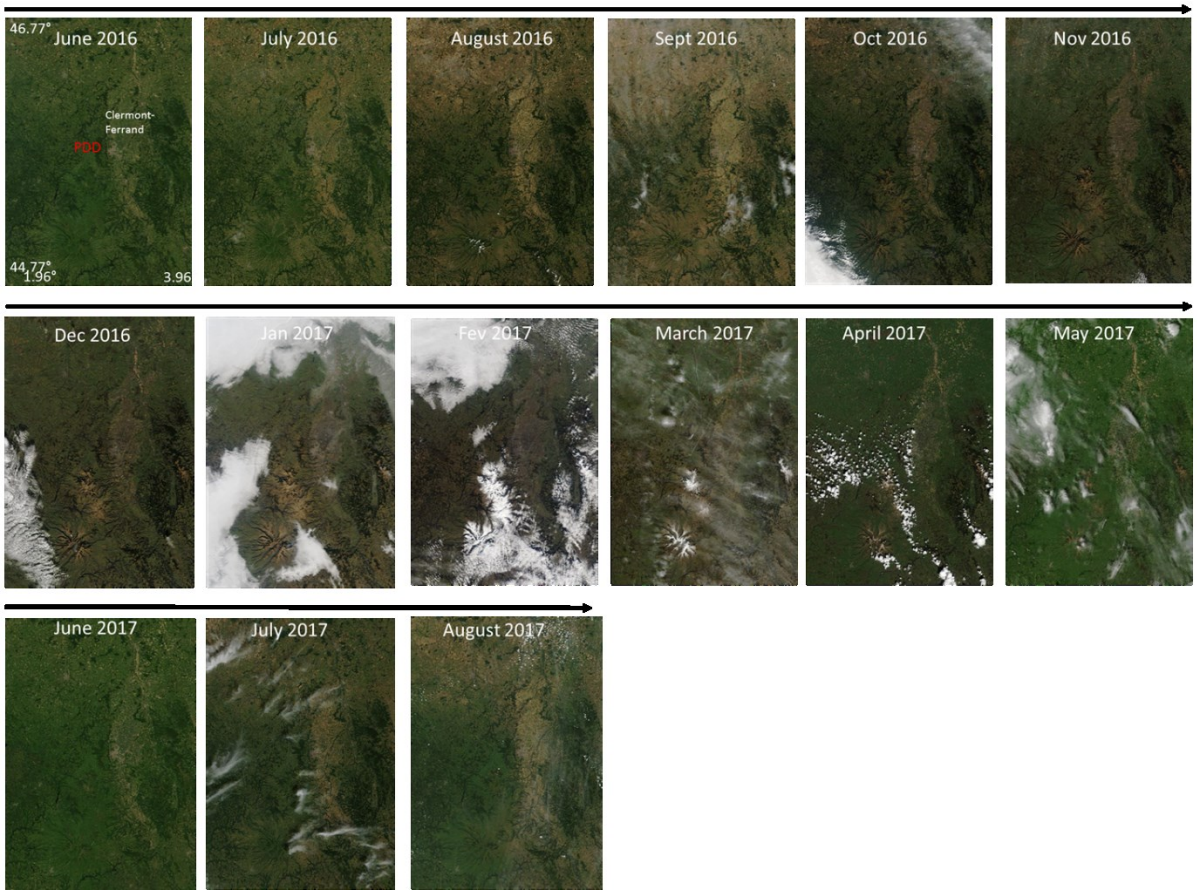


Figure S1: Monthly NASA satellite images of Puy-de-Dôme surrounding surfaces.

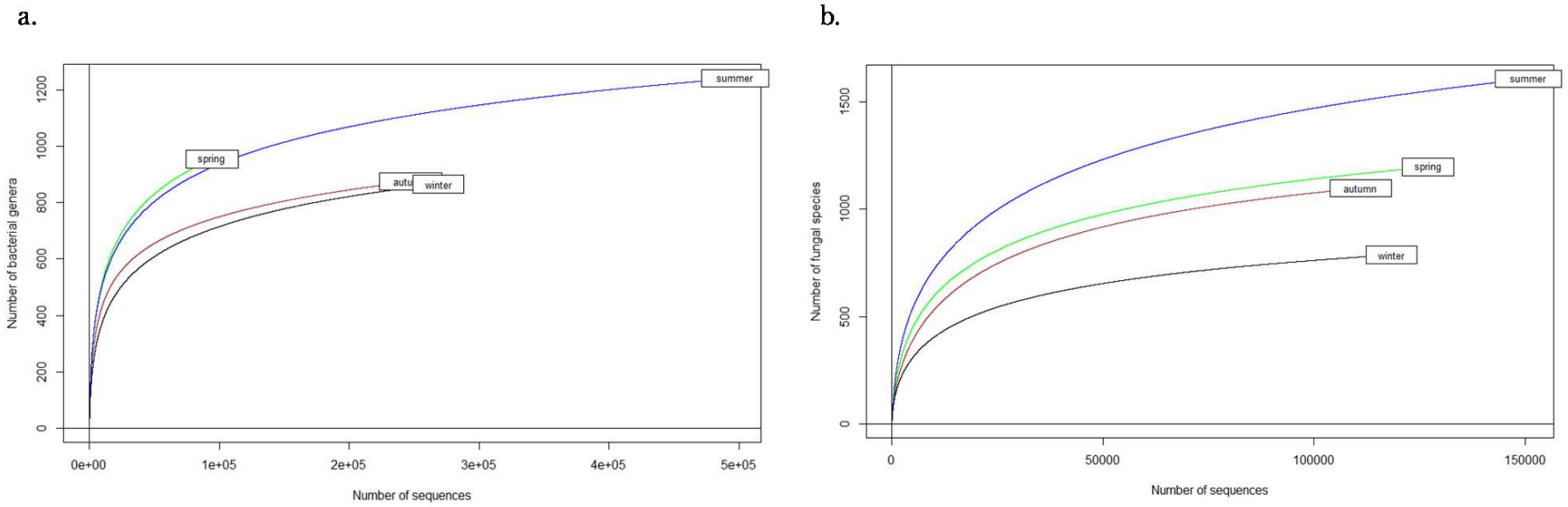
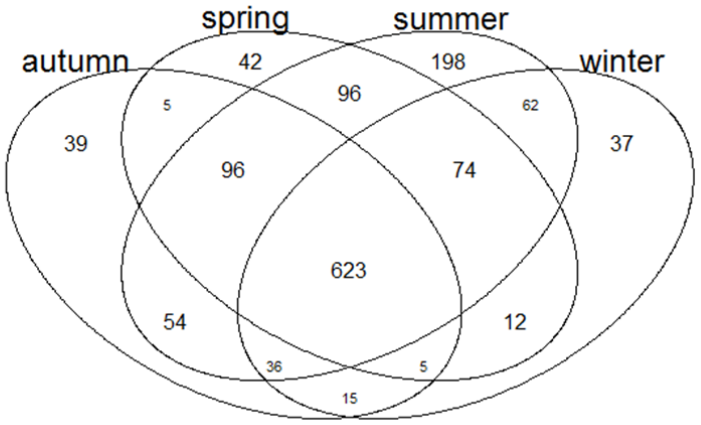


Figure S2: Rarefaction curves of the number of bacterial genera (a.) and fungal species (b.) per season.

a.

Percentage of unique bacterial genera per season:
-Autumn: 4.5%
-Spring: 4.4%
-Summer: 16.0%
-Winter: 4.3%



b.

Percentage of unique fungal species per season:
-Autumn: 17.4%
-Spring: 14.4%
-Summer: 28.4%
-Winter: 8.0%

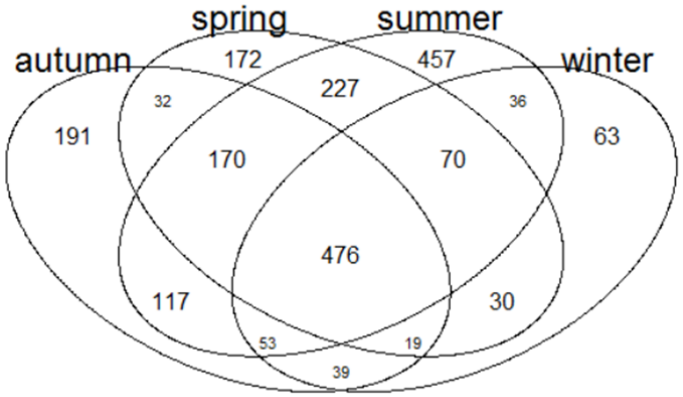


Figure S3: Venn diagrams showing the number of shared and unique bacterial genera (a.) and fungal species (b.) between the samples of each season after rarefaction. The percentage represented by the numbers of unique bacterial genera and fungal species for each season are summarized beside the figures.

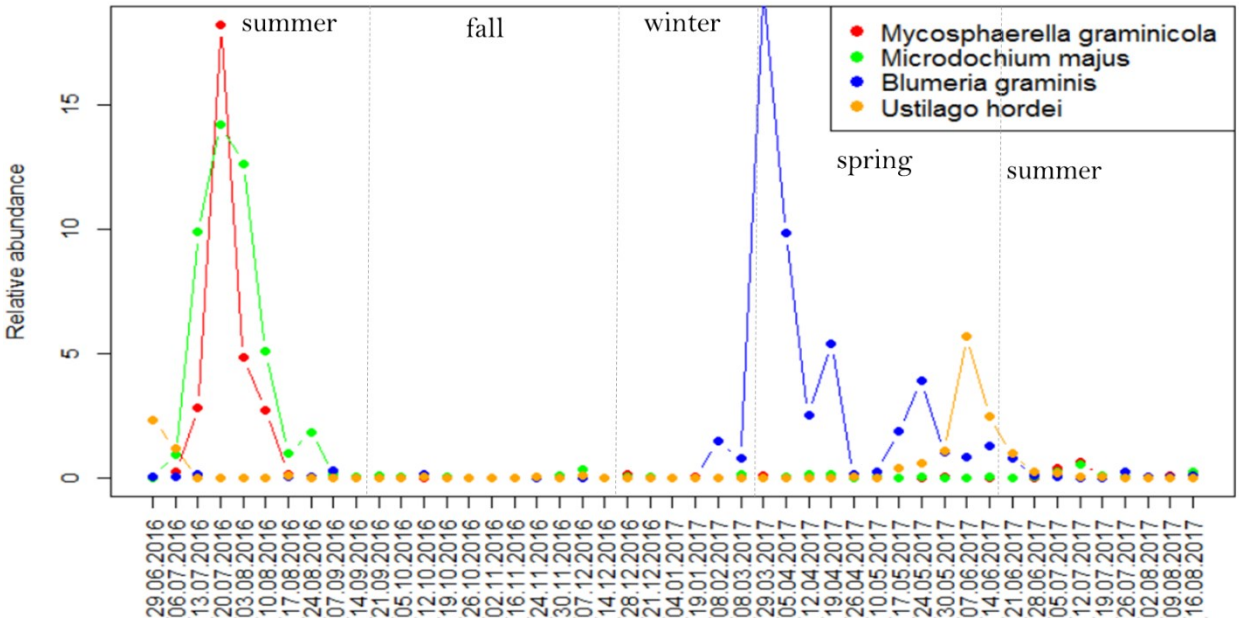


Figure S4: Temporal evolution of the relative abundance of four fungal phytopathogens over the year.

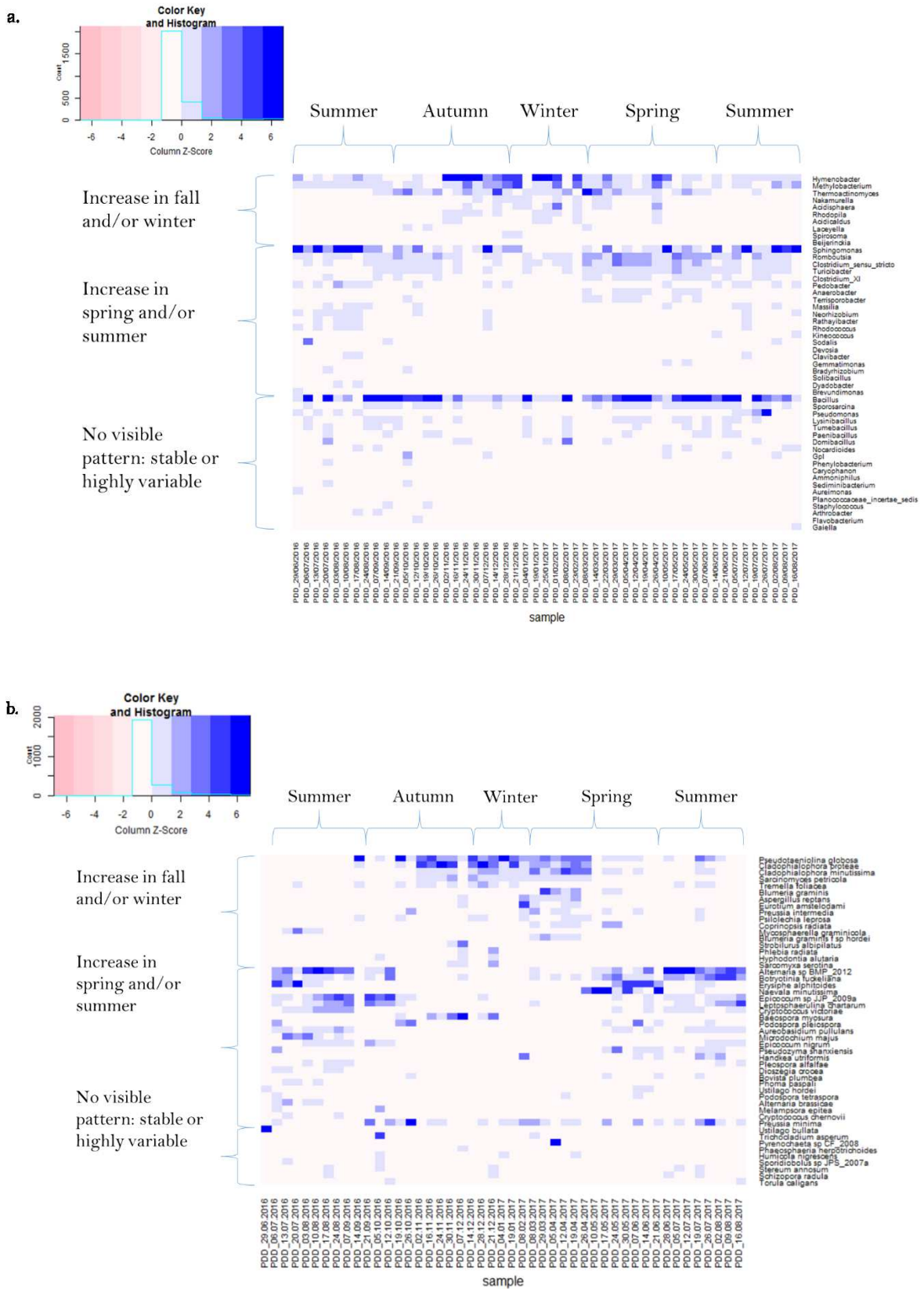


Figure S5: Heatmap showing the temporal evolution of the relative abundances (the relative abundances are centered and scaled) of the fifty most abundant bacterial genera (a.) and fungal species (b.) in the dataset. The microbial taxa are grouped based on the three general trends.

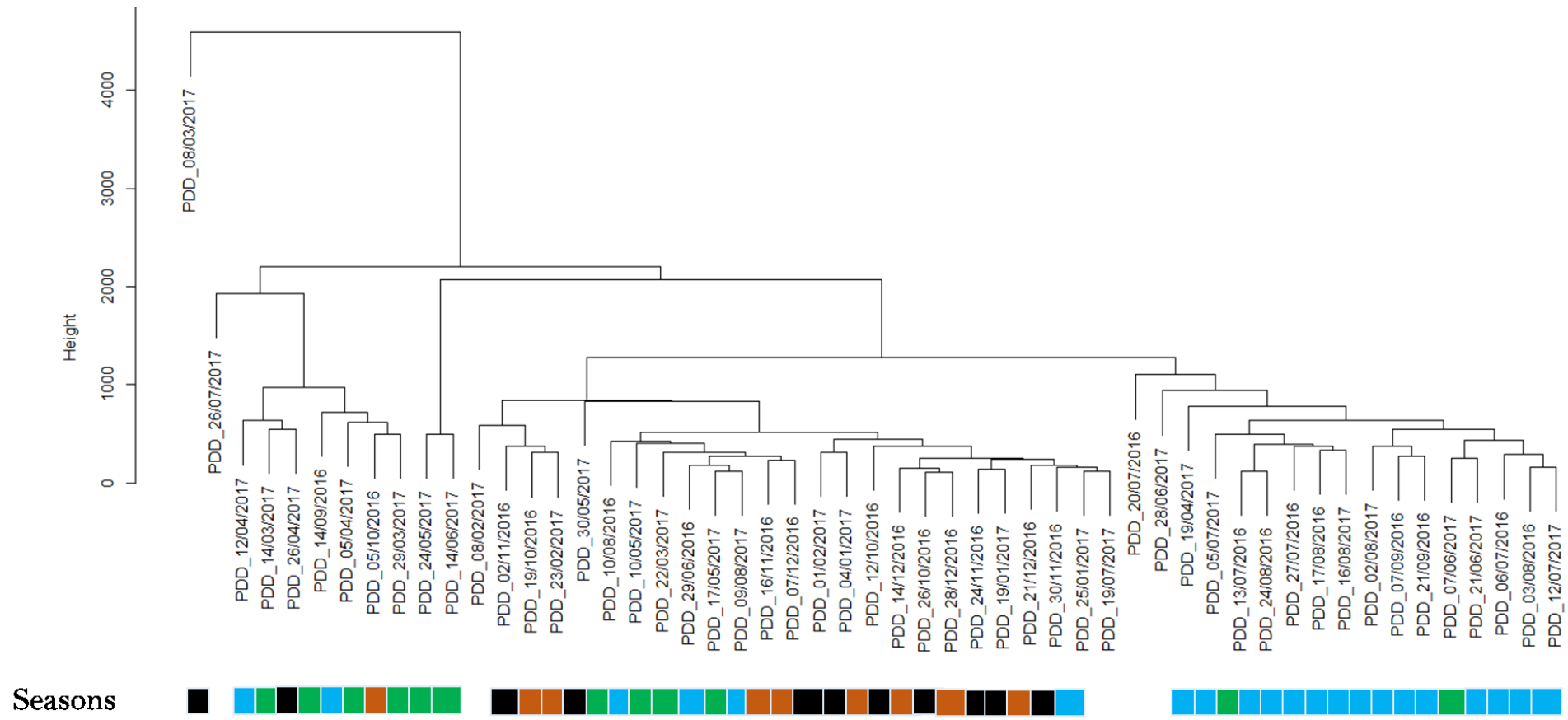


Figure S6: Hierarchical cluster analysis (average method) on the particulate matter chemical concentrations based on the Euclidean distance matrix. The season to which each sample belongs (black=winter, blue=summer, green=spring and brown=autumn) is indicated under each site in a frieze.

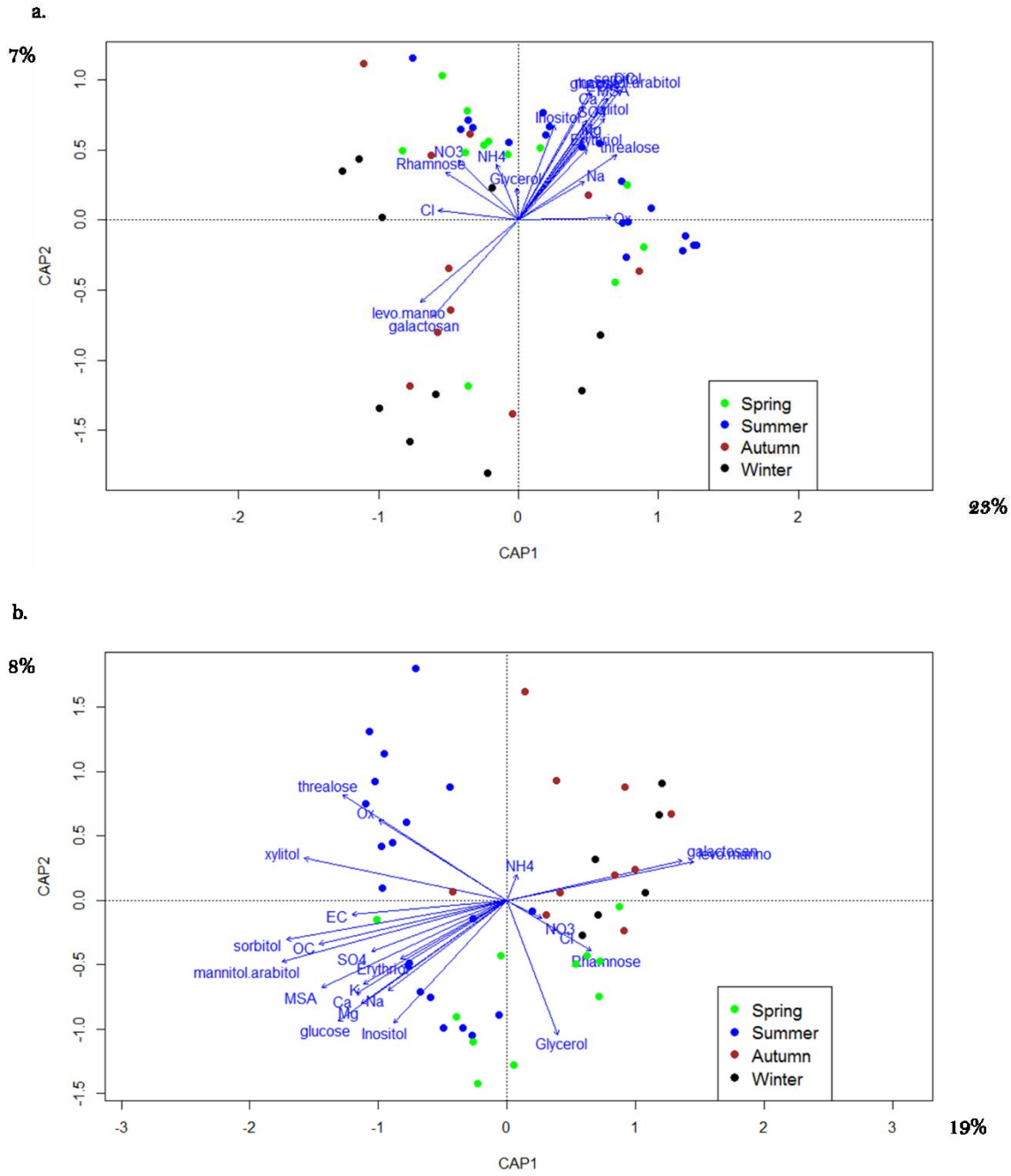


Figure S7: Constrained Analysis of Principal Coordinates. Part of variation in the distribution of both the bacterial (a.) and fungal (b.) communities explained by the chemical variables on the first 2 axes. The samples are colored depending on the season.

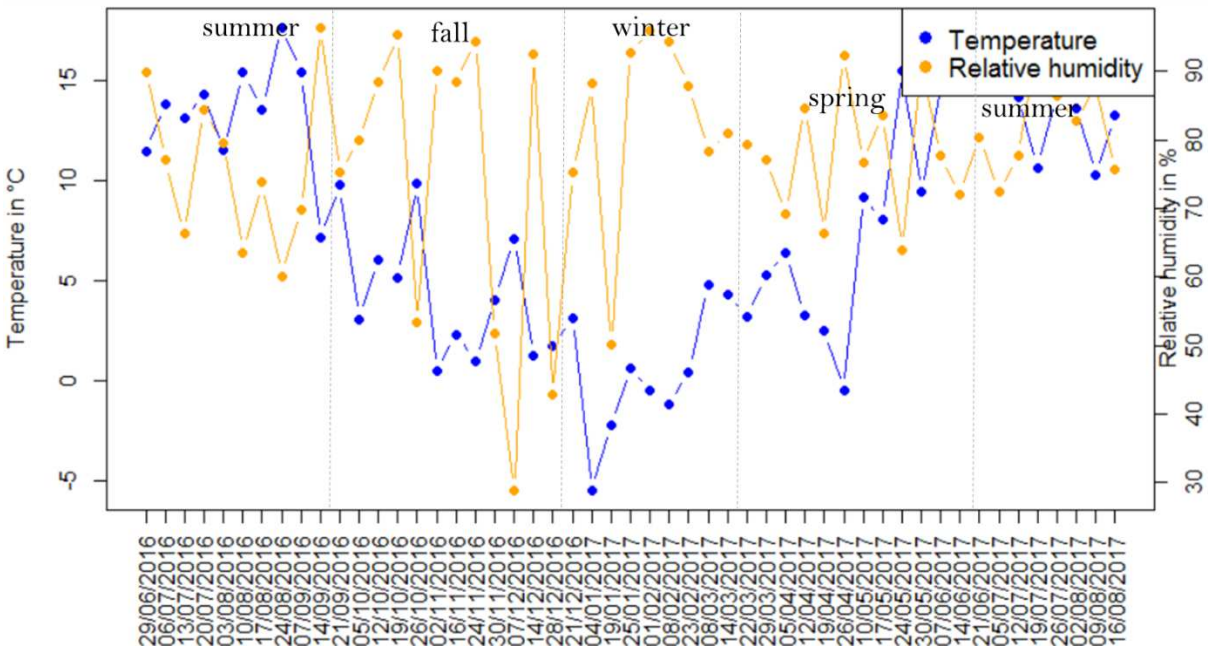
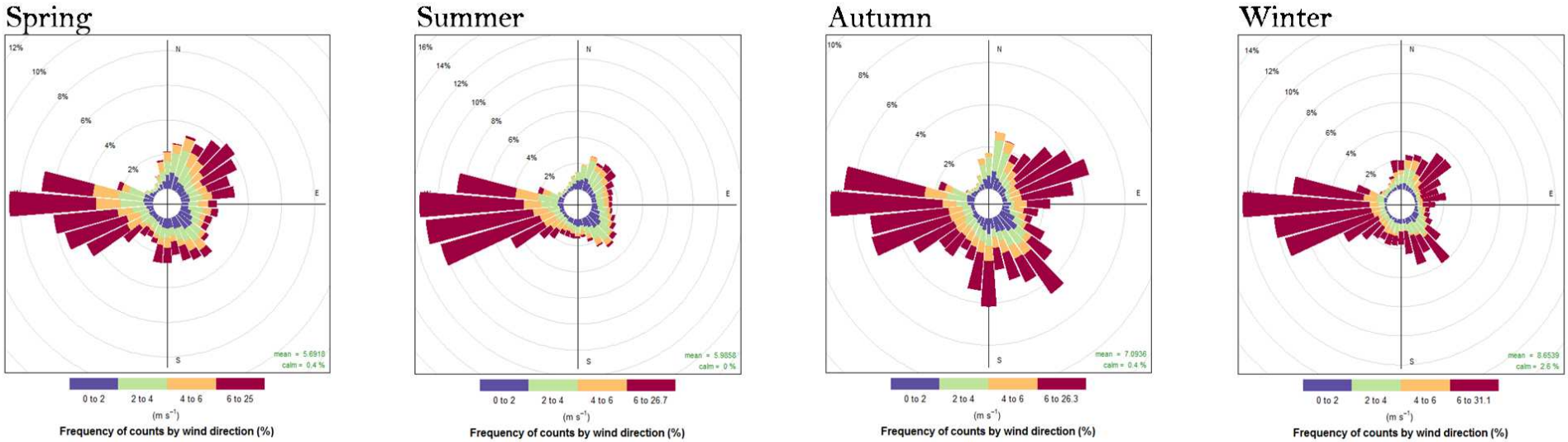


Figure S8: Temporal evolution of the temperature (degree celsius) and relative humidity (percentage)(averaged over the week) over the year.

a.



b.

Meteorological parameters	spring	summer	autumn	winter
Wind speed (m/s)	5.74 +/- 1.1	5.98 +/- 1.7	6.87 +/- 2.74	8.65 +/- 2.06
Wind direction (°)	174.02 +/- 25.6	194.59 +/- 34	164.66 +/- 42.21	180.04 +/- 38.16
Relative humidity (%)	79.1 +/- 8.3	79.22 +/- 9.74	76.32 +/- 23.17	78.67 +/- 18.35
Temperature (°C)	7.09 +/- 5.22	13.02 +/- 2.56	4.01 +/- 3.03	0.53 +/- 3.14
UV	34.75 +/- 3.9	34 +/- 4.21	15.76 +/- 4.45	17.89 +/- 5.43

Figure S9: Meteorological conditions per season. Wind roses covering the sampling time at each season (a) and meteorological parameters averaged per season (b).

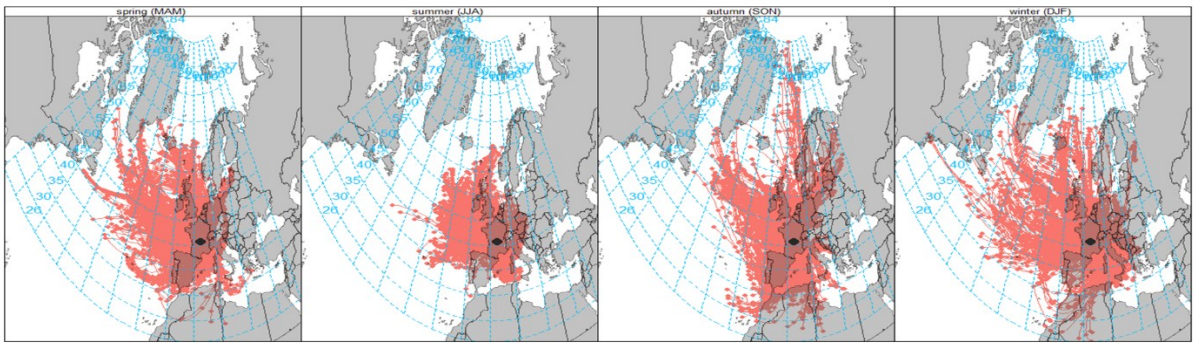


Figure S10: Air mass backtrajectories over 3 days at each season.

Chapter 5: Microbial functional signature in the atmospheric boundary layer

In preparation for submission in Environmental Microbiology

Romie Tignat-Perrier[1,2], Aurélien Dommergue[1], Alban Thollot[1], Olivier Magand[1], Timothy M. Vogel[2], Catherine Larose[2]

[1]Institut des Géosciences de l'Environnement, Université Grenoble Alpes, CNRS, IRD, Grenoble INP, Grenoble, France

[2]Environmental Microbial Genomics, Laboratoire Ampère, École Centrale de Lyon, Université de Lyon, Écully, France

Abstract

Microorganisms are ubiquitous in the atmosphere and can reach concentrations up to 10^6 microbial cells per cubic meter of air. Some airborne microbial cells were shown to be particularly resistant to atmospheric physical and chemical conditions (*e.g.,* UV radiation, desiccation, presence of radicals). In addition to surviving, some microorganisms of airborne origin were shown to be able to grow on atmospheric chemicals in laboratory experiments. Metagenomic investigations have been used to identify specific signatures of microbial functional potential in different ecosystems (soil, seawater, feces etc.). These studies have reported coherence between both microbial adaptation (*e.g.*, genetic modifications) and physical selection (*i.e.* the death of non-resistant cells) and their physical and chemical conditions. We conducted a comparative metagenomic study on the overall microbial functional potential and specific metabolic and stress-related microbial functions of atmospheric microorganisms in order to determine whether airborne microbial communities possess a specific functional potential signature. Air samples collected at nine sites around the world and samples from different ecosystems (soil, sediment, snow, feces, surface seawater etc.) were compared. While air samples grouped with their underlying ecosystems, they were characterized by a relatively high proportion of fungi. Fungal cells and spores are innately resistant entities and might likely resist and survive atmospheric physical stress better than bacterial cells. The high concentrations of fungi in air resulted in the higher proportion of

sequences annotated as genes involved in stress-related functions (*i.e.* functions related to the response to desiccation, UV radiation, oxidative stress etc.).

Introduction

Microorganisms are ubiquitous in the atmosphere and reach concentrations of up to 10^6 microbial cells per cubic meter of air[1]. Due to their important roles in public health and meteorological processes, understanding how airborne microbial communities are distributed over time and space is critical. While the dynamics of the concentration and taxonomic diversity of airborne microbial communities in the planetary boundary layer has been recently described[2–5] (**Chapter 3**), the functional distribution of airborne microbial communities remains unknown. Most studies have investigated functions of microbial strains of airborne origin after cultivation in the laboratory[6–11], which limits their analyses to the roughly 1% of easily cultivable microorganisms. Culture-independent techniques and especially metagenomic studies applied to air microbiology have the potential to provide additional information on the selection and genetic adaptation of airborne microorganisms to the atmospheric environment. However, to our knowledge, only three short-scale (*i.e.* a few samples) metagenomic studies on airborne microbial communities exist[12–14]. Metagenomic investigations of the complex microbial communities of many ecosystems have provided evidence that microorganism functional signatures reflect the abiotic conditions of their environment[15–18]. These comparative metagenomic studies showed specific microbial functional signatures in several ecosystems (for example, soil, seawater, lakes, feces, sludge) due to different relative abundances of specific microbial functions. This observed potential adaptation of microbial communities to their environment could have resulted from genetic modifications (microbial adaptation)[14,19–22] as well as the death of sensitive cells and survival of resistant cells (physical selection) that usually occurs when microorganisms are exposed to adverse conditions.

The presence of a potential microbial functional signature in the atmosphere has not been investigated yet. Microbial strains of airborne origin have been shown to survive and develop under conditions typically found in the atmosphere (high concentrations of H_2O_2, typical cloud carbonaceous sources, UV radiation etc.)[6–11,23]. While atmospheric chemicals, such as high concentrations of radicals, might lead to microbial adaptation, physical and unfavorable

conditions characteristic of the atmosphere such as UV radiation, low water content and cold temperatures might select which microorganisms can survive in the atmosphere. From the pool of microbial cells being aerosolized from Earth surfaces, these adverse conditions might act as a filter in selecting cells already resistant to unfavorable physical conditions. Fungal cells and especially fungal spores might be particularly adapted to survive in the atmosphere due to their innate resistance[24] and might behave differently than bacterial cells. Still, the proportion and nature (*i.e.* fungi versus bacteria) of microbial cells that are adapted to atmospheric conditions within airborne microbial communities are unknown.

Our objective was to identify whether airborne microbial communities in the planetary boundary layer possess a specific functional signature as compared to other environments as this might indicate a potential functional adaptation (microbial adaptation and/or physical selection) of airborne microorganisms. We used a metagenomic approach to compare the differences and similarities of both the overall functional potential and specific microbial functions (metabolic and stress-related functions) between microbial communities from air and other ecosystems (soil, sediment, surface seawater, river water, snow, human feces, phyllosphere, and hydrothermal vent). We sampled airborne microbial communities at nine different locations around the world during several weeks to get a global-scale view as well as to capture the between and within-site variability in airborne microbial functional potential.

Material and Methods

Sites and sampling

Air samples were collected at nine sites in 2016 and 2017. Sites were characterized by different latitudes (from the Arctic to the mid-latitude of the southern Hemisphere; **Fig. 1**), different elevations from sea level to high altitude (5380 m; **Fig. 1**) and different environment types (isolated marine site for Amsterdam-Island, coastal site for Cape-Point, polar site for Station-Nord and terrestrial sites for Grenoble, Chacaltaya, Puy-de-Dôme, Pic-du-Midi, Storm-Peak and Namco (**Table S1** in supplementary data). The number of samples collected per site varied from seven to sixteen (**Table S1** in supplementary data). The sampling starting date and collected air volume per air sample are presented in **Table S2** in supplementary data. We collected particulate matter smaller than 10 µm (PM10) on pre-treated quartz fiber filters using high volume air samplers (TISCH, DIGITEL, home-made). To avoid contamination, quartz

fiber filters as well as all the material in contact with the filters (*i.e.* filter holders, aluminium foils and plastic bags in which the filters were transported) were sterilized using strong heating (500°C for 8 hours) and UV radiation, respectively as detailed in Dommergue *et al.* (2019)[25] (Chapter 2). The collection time per sample lasted one week, and the collected volumes ranged from 2000 m^3 to 10000 m^3 after standardization using SATP standards (Standard Ambient Pressure and Temperature). Detailed sampling protocols are presented in Dommergue *et al.* (2019)[25]. MODIS (Moderate resolution imaging spectroradiometer) land cover approach (5'x5' resolution)[26,27] was used to quantify landscapes in the 50 km diameter area of our nine sampling sites (**Fig. S1** in supplementary data).

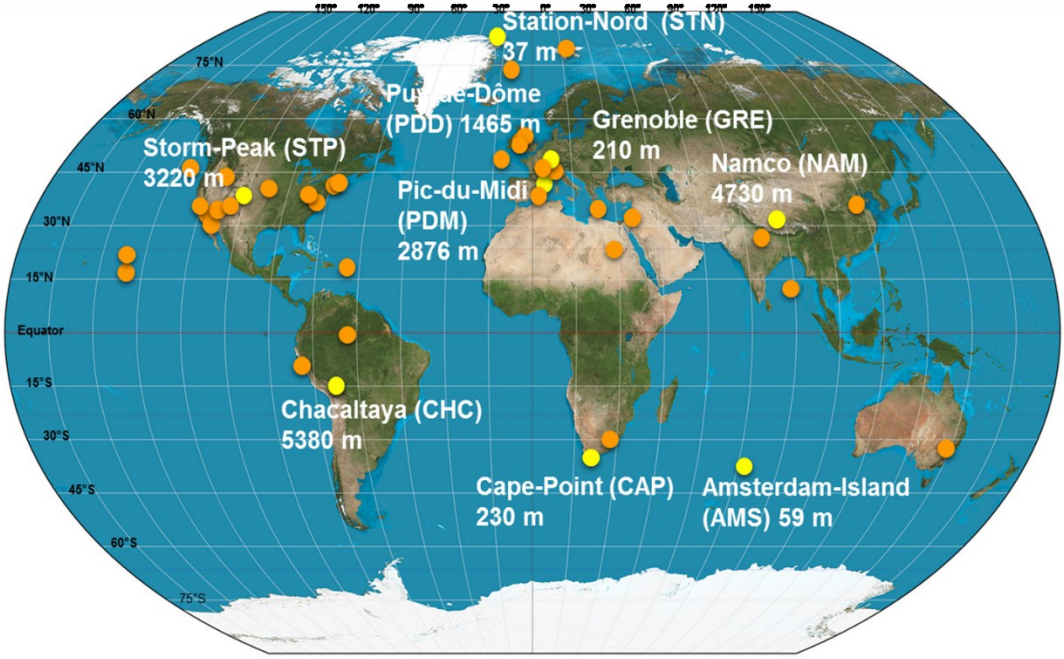


Figure 1: Map showing the geographical location and elevation from sea level of our nine sampling sites (in yellow), and the geographical position of whose public metagenomes come from (in orange). Abbreviations of our nine sampling sites are indicated in brackets.

Molecular biology analyses

DNA extraction

DNA was extracted from three circular pieces (punches) from the quartz fiber filters (diameter of one punch: 38 mm) using the DNeasy PowerWater kit with some modifications as detailed in Dommergue *et al.* 2019[25] (Chapter 2). During the cell lysis, the PowerBead tube containing the three punches and the pre-heated lysis solution were heated at 65°C during one hour after a 10-min vortex treatment at maximum speed. We then separated the filter debris from the

lysate after centrifugation of the mixture at 1000 rcf during 4 min. From this step, we continued the DNA extraction following the DNeasy PowerWater protocol. DNA concentration was measured using the High Sensitive Qubit Fluorometric Quantification (Thermo Fisher Scientific). DNA was stored at -20°C.

Real-Time qPCR analyses

16S rRNA gene qPCR. The bacterial cell concentration was approximated by the number of 16S rRNA gene copies per cubic meter of air. The V3 region of the 16S rRNA gene was amplified using the SensiFast SYBR No-Rox kit (Bioline) and the following primers sequences: Eub 338f 5'-ACTCCTACGGGAGGCAGCAG-3' as the forward primer and Eub 518r 5'-ATTACCGCGGCTGCTGG-3' as the reverse primer[28] on a Rotorgene 3000 machine (Qiagen). The reaction mixture of 20µL contained 10µL of SYBR master mix, 2µL of DNA and RNAse-free water to complete the final 20µL volume. The qPCR 2-steps program consisted of an initial step at 95°C for 2min for enzyme activation, then 35 cycles of 5sec at 95°C and 20sec at 60°C hybridization and elongation. A final step was added to obtain a denaturation from 55°C to 95°C with increments of $1°C.s^{-1}$. The amplicon length was around 200 bp. PCR products obtained from DNA from a pure culture of *Escherichia coli* were cloned in a plasmid (pCR™2.1-TOPO® vector, Invitrogen) and used as standard after quantification with the Broad-Range Qubit Fluorometric Quantification (Thermo Fisher Scientific).

18S rRNA gene qPCR. The fungal cell concentration was estimated by the number of 18S rRNA gene copies per cubic meter of air. The region located at the end of the SSU 18S rRNA gene, near the ITS1 region, was quantified using the SensiFast SYBR No-Rox kit (Bioline) and the following primers sequences: FR1 5'-AICCATTCAATCGGTAIT-3' as the forward primer and FF390 5'-CGATAACGAACGAGACCT-3' as the reverse primer[29] on a Rotorgene 3000 machine (Qiagen). The reaction mixture of 20µL contained 10µL of SYBR master mix, 2µL of DNA and RNAse-free water to complete the final 20µL volume. The qPCR 2-steps program consisted of an initial step at 95°C for 5min for enzyme activation, then 35 cycles of 15sec at 95°C and 30sec at 60°C hybridization and elongation. A final step was added to obtain a denaturation from 55°C to 95°C with increments of $1°C.s^{-1}$. The amplicon length was around 390 bp. PCR products obtained from DNA from a soil sample were cloned in a plasmid (pCR™2.1-TOPO® vector,

Invitrogen) and used as standard after quantification with the Broad-Range Qubit Fluorometric Quantification (Thermo Fisher Scientific).

MiSeq Illumina metagenomic sequencing

Metagenomic library preparation. Metagenomic libraries were prepared from 1 ng of DNA using the Nextera XT Library Prep Kit and indexes following the protocol in Illumina's "Nextera XT DNA Library Prep Kit" reference guide with some modifications for samples whose DNA concentration was < 1 ng as follows. The tagmented DNA was amplified over 13 PCR cycles instead of 12 PCR cycles, and the libraries (after indexing) were resuspended in 30 µL of RBS buffer instead of 52.5 µL. Metagenomic sequencing was performed using the MiSeq and V2 technology of Illumina with 2 x 250 cycles. At the end of the sequencing, the adapter sequences were removed by internal Illumina software.

Reads quality filtering. Reads 1 and reads 2 per sample were not paired but merge in a common file before filtering them based on read quality using the tool FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) using a minimum read quality of Q20, minimum read length of 120 bp and one maximum number of N per read. Samples with less than 6000 filtered sequences were removed of the dataset.

Downloading of public metagenomes

Public metagenomes were downloaded from the MGRAST and SRA (NCBI) databases as quality filtered read-containing fasta files and raw read containing fastq files, respectively. The fastq files containing raw reads underwent the same quality filtering as our metagenomes (as discussed above). The list of the metagenomes, type of ecosystem, number of sequences and sequencing technology are summarized in **Table S1** in supplementary data. The sampling sites are positioned on the map in **Figure 1**.

Data analyses

All graphical and multivariate statistical analyses were carried out using the vegan[30], ggplot2[31] and reshape2[32] packages in the R environment.

Annotation of the metagenomic reads. Firstly, to access the overall functional potential of each sample, the filtered sequences per sample were functionally annotated using Diamond,

then the gene-annotated sequences were grouped in the different SEED functional classes (around 7000 functional classes, referred simply to as functions) using MEGAN6[33]. Functional classes which were present ≤ 2 times in a sample were removed of this sample. In parallel, the Kraken software[34] was used to retrieve the bacterial and fungal sequences separately from the filtered sequences using the Kraken bacterial database and FindFungi[35] fungal database (both databases included complete genomes), respectively. Separately, both the bacteria and fungi-associated sequences were also functionally annotated using Diamond and MEGAN6 (number of sequences functionally annotated in **Table S3** in supplementary data).

Secondly, for specific metabolic and stress-related functions, we annotated the sequences using eggNOG-Mapper[36] (Diamond option), then examined specific GO (Gene Ontology) terms. The different GO terms used were the following: GO:0042744 (hydrogen peroxide catabolic activity), GO:0015049 (methane monooxygenase activity) as specific metabolic functions and GO:0043934 (sporulation), GO:0009650 (response to UV), GO:0034599 (cell response to oxidative stress), GO:0008219 (cell death), GO:0009269 (response to desiccation) as stress-related functions. The number of hits of each GO term was normalized per 10000 annotated sequences and calculated from all sequences, bacteria-associated sequences and fungi-associated sequences for each sample. The number of sequences annotated by eggNOG-Mapper[36] was also evaluated (**Table S1** in supplementary data). The putative concentration of a specific function or functional class in the samples is determined as the concentration of sequences annotated as one of the functional genes associated to this function (or functional class).

Statistics. Observed functional richness and evenness were calculated per sample after rarefaction on all sequences (rarefaction at 2000 sequences), bacteria-associated sequences (rarefaction at 500 sequences) and fungi-associated sequences (rarefaction at 500 sequences). The distribution of the samples was analyzed based on the SEED functional classes. PCoA and hierarchical clustering analyses (average method) were carried out on the Bray-Curtis dissimilarity matrix based on the Hellinger-transformed relative abundances of the different SEED functional classes. SIMPER analyses were used to identify the functions responsible for the clustering of samples in groups. ANOVA analyses were used to test the difference between the percentage of fungi-associated sequences as well as the number of

hits of each Gene Ontology term (normalized per 10000 annotated sequences) among the different sites and the different ecosystems.

Results

Percentage of fungi-associated sequences

The percentage of sequences annotated as belonging to fungal genomes (or fungi-associated sequences, as opposed to bacteria-associated sequences) was on average higher in air samples compared to soil ($P<10^{-7}$), snow ($P=0.007$), seawater ($P=0.045$) and sediment samples ($P=0.047$; **Fig. 2** and **Table S1** in supplementary data). Within the air samples, NAM (19%), STN (24%) and CHC (27%) showed on average the lowest percentages of fungi-associated sequences while STP (88%), GRE (79%), AMS (71%) and PDD (62%) showed the highest percentages. For the ecosystems that were only represented by one sample and therefore not integrated in the ANOVA test, we observed average percentages of fungi-associated reads of 3% in feces, 9% in hydrothermal vents, 19% in river water samples and 37% in the phyllosphere. Some samples from soil, sediments and seawater such as French agricultural soil (61%), Peru sediments (53%) and Celtic seawater (53%) had relatively high percentages of fungi-associated sequences while other samples had less than 50%. The number of fungal and bacterial cells was estimated using 16S rRNA and 18S rRNA gene copy numbers per cubic meter of air, respectively. qPCR results on air samples are available in Chapter 3[5]. Air samples showed ratios between bacterial cell and fungal cell concentrations from around 4.5 times up to 160 times lower than soil samples (**Table S4** in supplementary data).

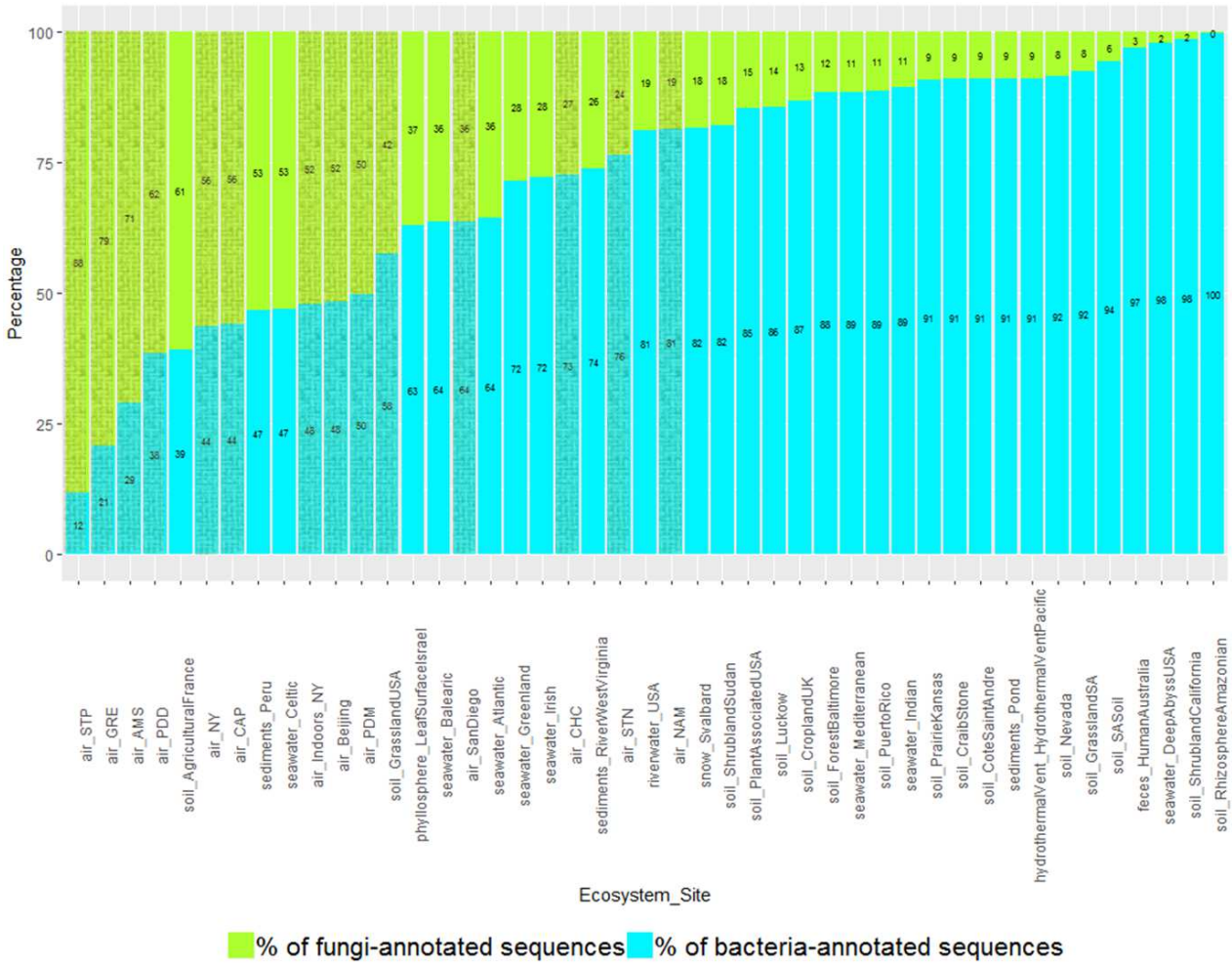


Figure 2: Percentage of fungi-associated sequences and bacteria-associated sequences over the total number of sequences annotated as belonging to fungal and bacterial genomes in the metagenomes. The mean was calculated for the sampling site including several metagenomes. Air sites are distinguished by hatching lines.

Airborne microbial functional profiles

The fifty most abundant SEED functional classes represented in air samples are listed in **Table S5** in supplementary data. The 5-FCL-like protein, the long chain fatty acid CoA ligase and the TonB-dependent receptor were the top three functions observed in air samples when including all the sequences (**Table S4** in supplementary data). The airborne microbial functional profiles based on the SEED functions were compared between samples from the different weeks of sampling and between different locations. The profiles were graphed using PCoA multivariate analysis to visualize differences and similarities. The different air samples (sampled during sequential weeks) from the same site did not cluster tightly together on the PCoA multivariate analysis (**Fig. S2** in supplementary data). In order to incorporate weekly

variances when comparing sites, we used the microbial functional profile averaged per site in the subsequent multivariate analyses done with the samples of the other ecosystems (**Fig. 3**). The PCoA multivariate analysis showed that terrestrial air sites (GRE, NAM, STP, PDD, PDM, CHC, New York) grouped with the soil, sediment and snow samples while the marine and coastal air sites (AMS, CAP, San Diego) were situated between the soil samples and the seawater and river water samples (**Fig. 3**). The polar air site STN did not group with the other air samples. The functions responsible for the differentiation of the groups (Group 1: terrestrial air samples, soil, sediment and snow samples; Group 2: marine, polar and arctic air samples; Group 3: seawater samples) are presented in **Table S6** in supplementary data. When considering only the bacteria-associated sequences (not including fungal sequences), the distribution of the air terrestrial sites did not change, while the marine AMS, coastal CAP and polar STN air sites were further from the seawater and river water samples than when the fungi-associated sequences were included (**Fig. S3** in supplementary data). The distribution of the different samples underwent further changes when considering only the fungal sequences. We observed an absence of a clear separation between soil and seawater samples which (for the majority) grouped closely together, and air terrestrial samples did not group with soil, sediment or snow samples (**Fig. S3** in supplementary data).

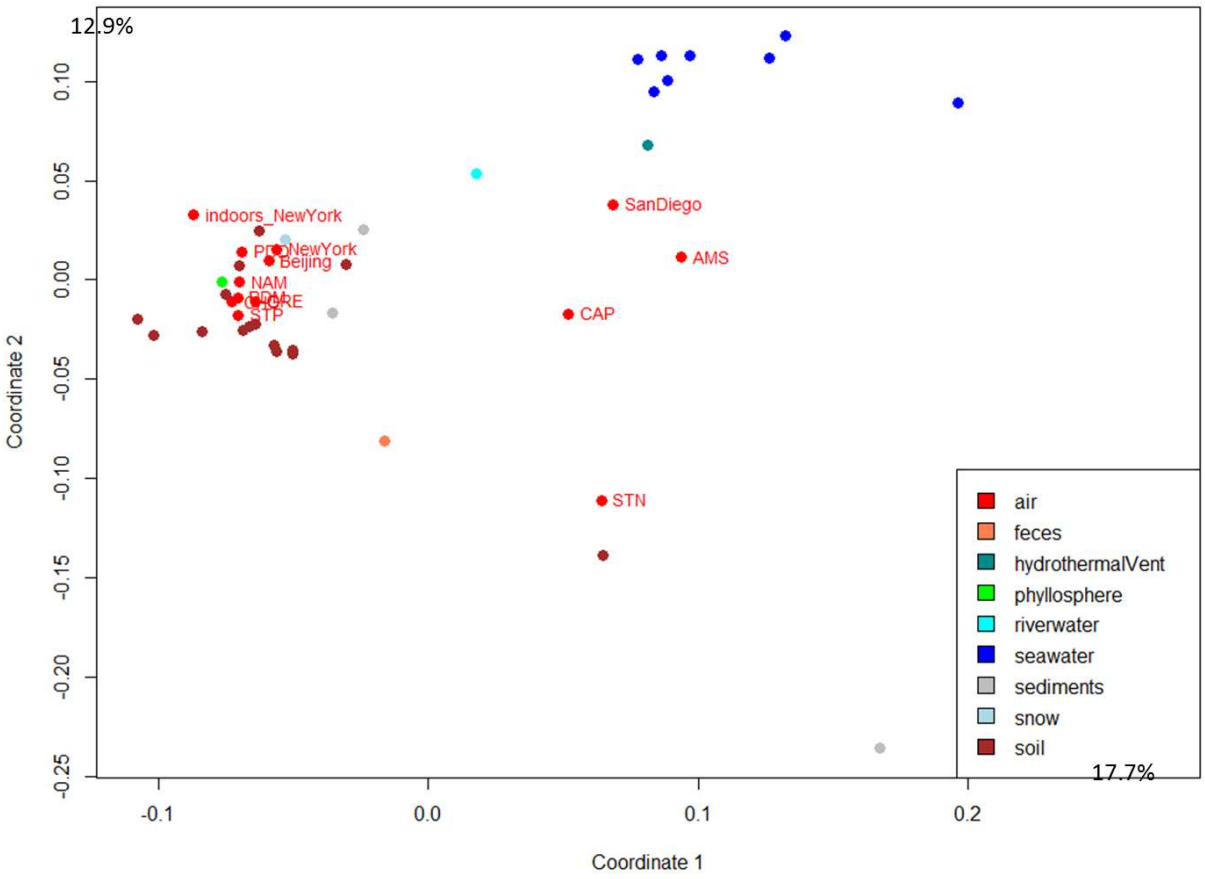


Figure 3: PCoA analysis of the Bray-Curtis dissimilarity matrix based on the functional potential structure of each site. For the site including several metagenomes, the average profile was calculated. Colors indicate the ecosystems in which the sites belong to.

Airborne microbial functional richness and evenness

Functional richness and evenness were evaluated using the relative abundance of sequences in the different SEED categories. The average richness in SEED functional classes (or functions) in air was lower than the average functional class richness in soil, surface seawater, hydrothermal vents, river water, phyllosphere, and feces (P<0.05) (**Table S3** and **Fig. S4** in supplementary data). Among the air samples, the functional class richness was highest in Beijing (4060 +/- 112 functional classes) and New York indoor air samples (3302 +/- 299 functional classes) (P<0.05), and lowest in STN (956 +/- 547 functional classes). When looking at the bacteria-annotated sequences, almost the same trend was observed, *i.e.* the functional class richness in air was lower than in soil, hydrothermal vents, river water, phyllosphere, and feces, and not different from the other ecosystems (P<0.05 and >0.05, respectively) (**Table S3** in supplementary data). The functional class richness was higher in Beijing (2835 +/- 59

functional classes) and New-York indoor air samples (2183 +/- 387 functional classes) compared to the other air samples whose values ranged between 270 +/- 197 functional classes in AMS and 1142 +/- 461 functional classes in CHC. For fungi-associated sequences, the functional class richness in air was lower than the functional class richness in soil, surface seawater, feces, hydrothermal vents, river water and phyllosphere (P<0.05) (**Table S3** in supplementary data). Within air samples, the functional class richness based on fungi-associated sequences was higher in Beijing (1129 +/- 92 functional classes) and New York indoor air samples (687 +/- 206 functional classes) than in the other air sites (P=10^{-9}) whose values ranged from 66 +/- 58 functional classes in AMS and 392 +/- 131 functional classes in STP (**Table S3** in supplementary data). The functional class evenness in air was on average higher than in soil (P=0.03), and not different to the functional class evenness observed in the other ecosystems (sediment, seawater, snow). When looking at the bacteria and fungi-associated sequences separately, the functional class evenness in air was on average higher than in soil, feces, phyllosphere and riverwater (P<0.05) (**Table S3** in supplementary data).

Concentration of specific microbial functions which might have a role under atmospheric conditions

Two metabolic functions associated with atmospheric abundant chemicals (H_2O_2 and CH_4) were examined, hydrogen catabolism and methane monoxygenase activity. The concentration of hydrogen peroxide catabolic related functions per 10000 sequences varied between air sites (P=$1.7x10^{-9}$) with highest values for AMS (27 +/- 1) and GRE (27 +/- 1) (**Fig. S4** in supplementary data). However, there was no difference between all the air samples averaged and the other ecosystems (P=0.78). The French agricultural soil showed the highest relative abundance (133 +/- 4). When considering the fungi and bacteria-associated sequences separately, the concentration in air was on average lower compared to soil (P=$2x10^{-4}$) and not different from all the other ecosystems (P=0.82), respectively (**Fig. S4** in supplementary data). The number of methane monooxygenase-related functions per 10000 sequences was only detectable when considering all the sequences (*i.e.* bacteria and fungi-associated sequences). The number of methane monooxygenase-related functions did not vary between air sites (P=0.59) while we observed a high variability between sampling periods within sites (**Fig. S5** in supplementary data), but on average it was not different from the ecosystems (P=0.76).

Different stress response functions (sporulation, UV response, oxidative stress cell response, desiccation response, cell death related functions, chromosome plasmid partitioning protein ParA and lipoate synthase) were examined. The concentration of sporulation-related functions per 10000 annotated sequences largely varied between air sites (P=2.6x10^{-9}), with the lowest values observed for STN (7 +/- 9), San Diego (9 +/- 6), NAM (17 +/- 15) and CHC (26 +/-13), and the highest values observed for STP (120 +/- 18), Beijing (126 +/- 22), GRE (131 +/- 21) and New York (141 +/- 98) (**Fig. 4**). It was on average higher in air compared to soil (P=10^{-7}), sediments (P=8x10^{-3}) and seawater (P=0.04) although the Celtic seawater sample presented a very high concentration (127). Snow showed a relatively high average concentration (36) which was not different from air concentration (P>0.05). For the ecosystems including one value (*i.e.* one sample, so not integrated in the ANOVA tests), feces showed a relatively high concentration of sporulation-related functions (41) while hydrothermal vent, phyllosphere and river water showed relatively low concentrations compared to air (<10). When considering the fungi-associated sequences separately from the bacteria-associated sequences, the same trend was observed, *i.e.* the concentration of sporulation-related functions in air was on average higher compared to soil (P=10^{-7}), sediments (P=10^{-7}), seawater (P=4x10^{-7}) as well as phyllosphere, hydrothermal vent and river water. The concentration was relatively high in the Celtic seawater (186) and the snow samples (163 +/- 47). We also observed a large variability within air sites (P=3.6x10^{-7}). When considering the bacteria-associated sequences only, this concentration was not different between air and the other ecosystems (P>0.05) and showed a smaller variability between air sites. Two samples, the phyllosphere (35) and the shrubland soil from Sudan (32) showed high numbers of sporulation-related functions per 10000 annotated sequences.

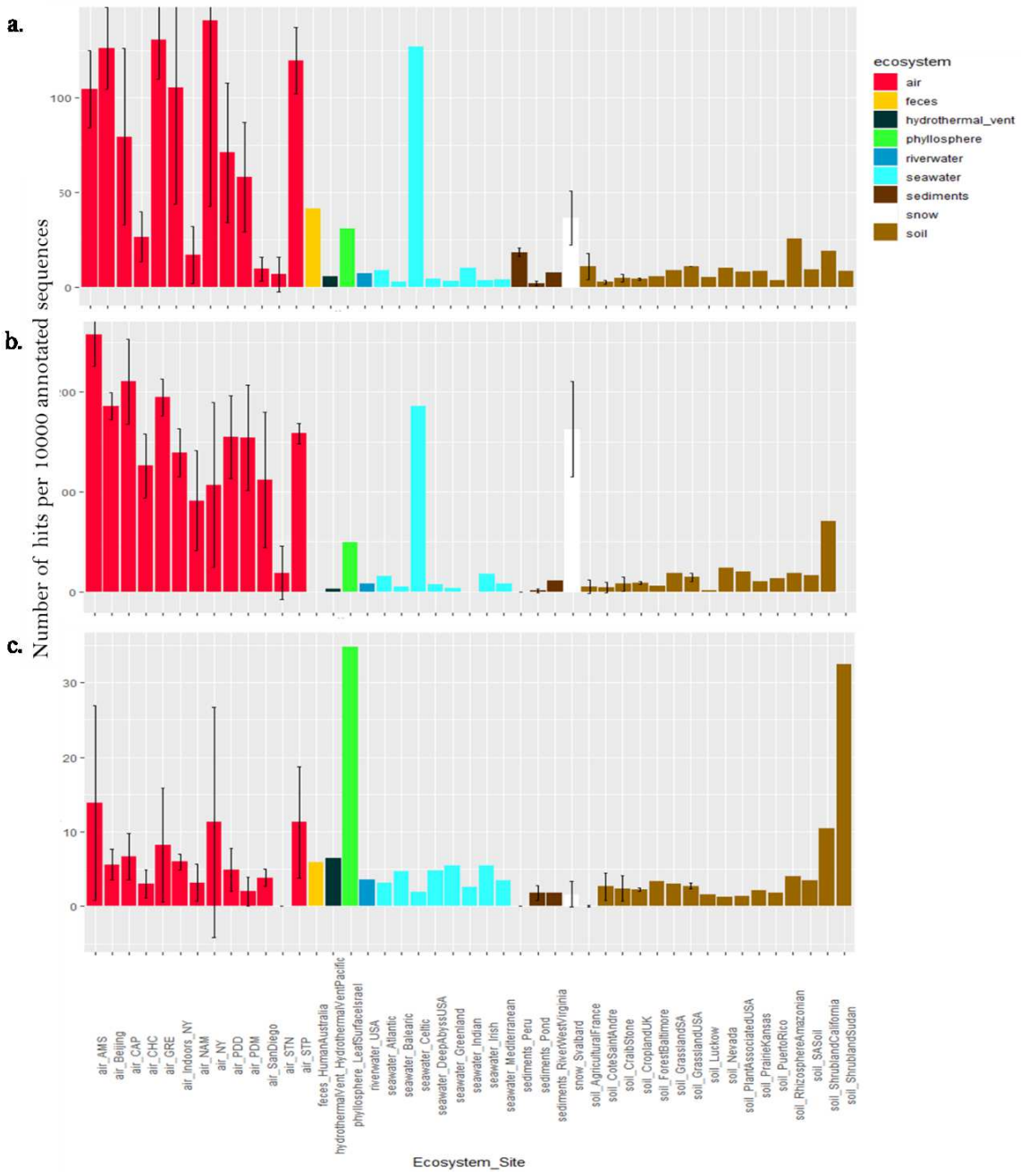


Figure 4: Average number of hits of sporulation related functions per 10000 annotated sequences from (a) all sequences, (b) fungi-associated sequences and (c) bacteria-associated sequences per site. Colors indicate the ecosystems in which the sites belong to. For the sites including several metagenomes, the standard deviation was added.

The concentration of UV response related functions per 10000 annotated sequences varied between air sites (P=9x10^{-5}), with values ranging from 16 +/- 2 in NAM and 19 +/- 4 in STN to 29 +/- 3 in STP and 36 +/- 6 in AMS (**Fig. 5**). The concentration was on average higher in air compared to sediments (P=3x10^{-3}), and comparable to soil, snow and seawater (P>0.05). The

other ecosystems showed lower ratios (feces, phyllosphere) or comparable concentrations (hydrothermal vent, river water) compared to air. Within the soil samples, the French agricultural soil samples showed a high average concentration (56 +/- 8), which increased the average ratio observed in soil samples. When considering both fungi and bacteria-associated sequences separately, this ratio was not different between air and the other ecosystems (*P*>0.05).

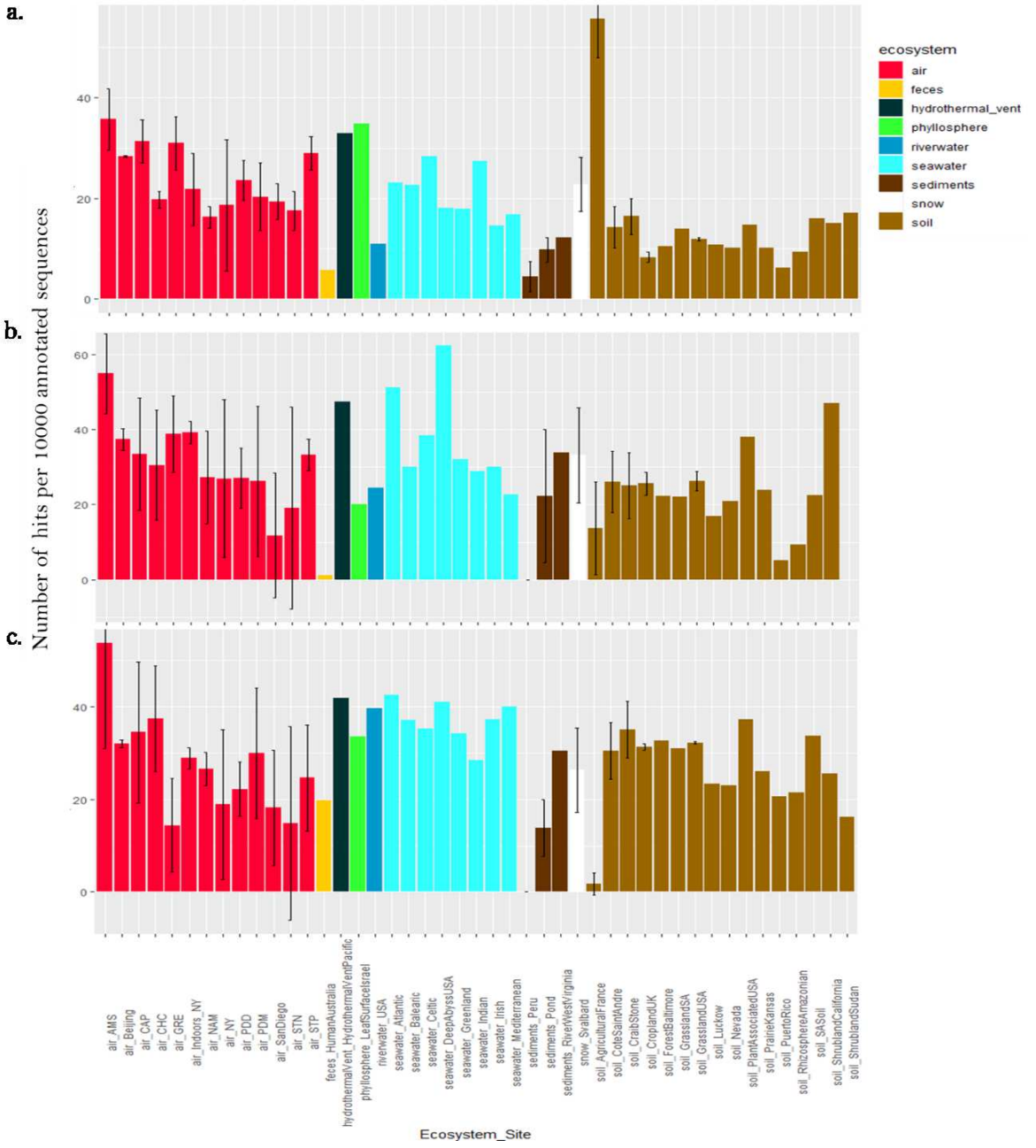


Figure 5: Average number of hits of UV protection related functions per 10000 annotated sequences from (a) all sequences, (b) fungi-associated sequences and (c) bacteria-associated sequences per site. Colors indicate the ecosystems in which the sites belong to. For the sites including several metagenomes, the standard deviation was added.

The concentration of oxidative stress cell response related functions per 10000 annotated sequences varied largely between air sites (P=3x10^{-12}), with the lowest values observed for STN (23 +/- 5), San Diego (11 +/- 3) and NAM (28 +/- 10), and the highest values observed for STP (105 +/- 16), AMS (108 +/- 16) and GRE (119 +/- 19) (**Fig. 6**). The concentration was on average higher in air compared to soil (P=10^{-7}), sediments (P=6x10^{-4}) and seawater (P=0.01). Snow showed a relatively high average value (46 +/- 11), not different from air (P>0.05). The other ecosystems (feces, river water, hydrothermal vent, phyllosphere) showed lower ratios compared to air. When considering fungi-associated sequences separately, the concentration of oxidative stress related functions per 10000 sequences was on average higher in air compared to soil (P=4x10^{-6}) and sediments (P=10^{-4}). Feces showed a very high average value (2237). When considering bacteria-associated sequences separately, this concentration was not different between air and the other ecosystems (P>0.05). When considering both fungi and bacteria-associated separately, the variability in the concentration oxidative stress cell response related functions between air sites diminished and their difference was not detected anymore (P>0.05).

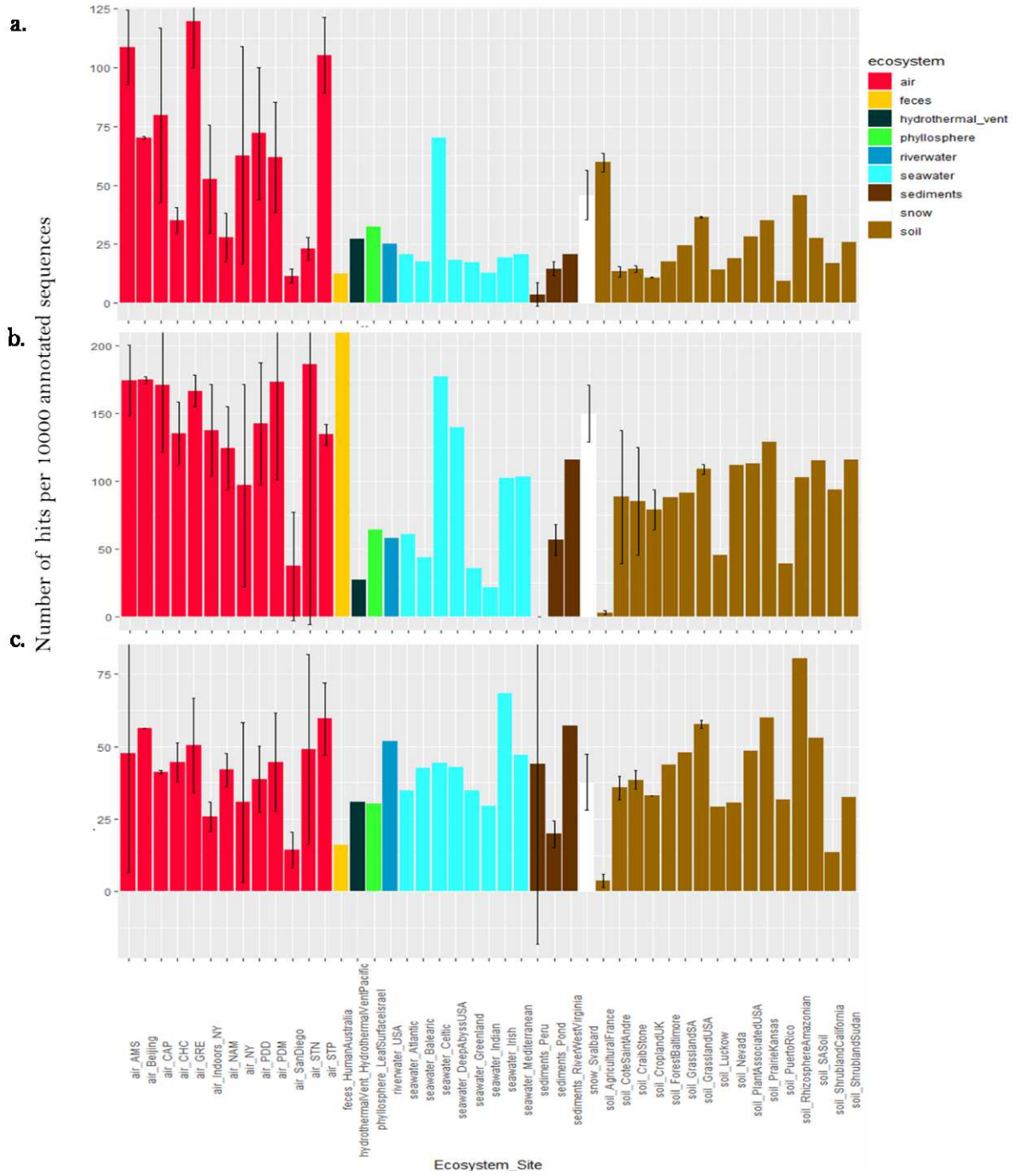


Figure 6: Average number of hits of oxidative stress cell response related functions per 10000 annotated sequences from (a) all sequences, (b) fungi-associated sequences and (c) bacteria-associated sequences per site. Colors indicate the ecosystems in which the sites belong to. For the sites including several metagenomes, the standard deviation was added.

The concentration of cell death related functions per 10000 annotated sequences varied between air sites (P=4x10^{-9}) with values ranging from 6 +/- 7 in NAM and 8 +/- 9 in STN to 32 +/- 26 in PDD and 88 +/- 7 in San Diego (**Fig. 7**). The concentration was not different between

air and the other ecosystems (P=0.76). When considering fungi-associated sequences separately, the concentration of cell death related functions per 10000 sequences was on average higher in air compared to sediments (P=8x10^{-3}) and soil (P=2x10^{-3}), and comparable to the other ecosystems. Within air samples, the concentration of cell death related functions was not different between air sites (P=0.076). When considering bacteria-associated sequences separately, the concentration of cell death related functions was different between air sites (P=10^{-3}) but was not different between air and the other ecosystems (P=0.35).

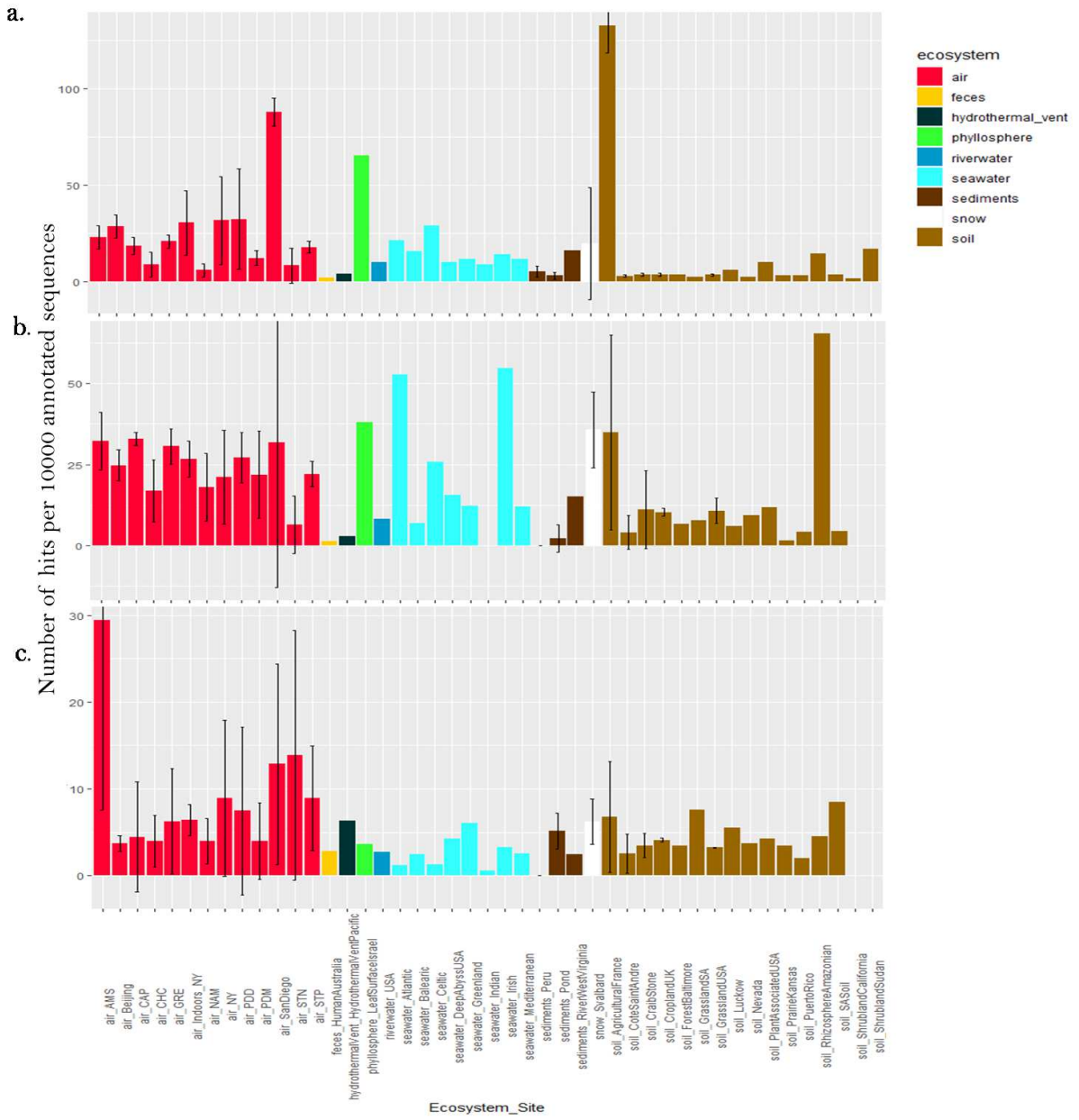


Figure 7: Average number of hits of cell death related functions per 10000 annotated sequences from (a) all sequences, (b) fungi-associated sequences and (c) bacteria-associated sequences per site. Colors indicate the ecosystems in which the sites belong to. For the sites including several metagenomes, the standard deviation was added.

The concentration of desiccation response related functions per 10000 sequences varied between air sites (P=2x10^{-5}), with the highest values in GRE (4 +/- 1), STP (4 +/- 1) and AMS (3 +/- 3), and the lowest values in STN (0.5 +/- 1) and San Diego (0.1 +/- 0.1) (**Fig. 8**). It was on average higher in air compared to the other ecosystems (P=2x10^{-5}). Still Svalbard snow and French agricultural soil showed high values (2 +/- 1 and 3 +/- 1, respectively) (**Fig. 8**). When considering fungi-associated sequences only, concentration of desiccation response related functions in air was higher compared to soil (P=5x10^{-4}) and Svalbard snow ratio was high (10 +/- 10). No difference between the ecosystems was observed when considering bacteria-associated sequences separately (P=0.97).

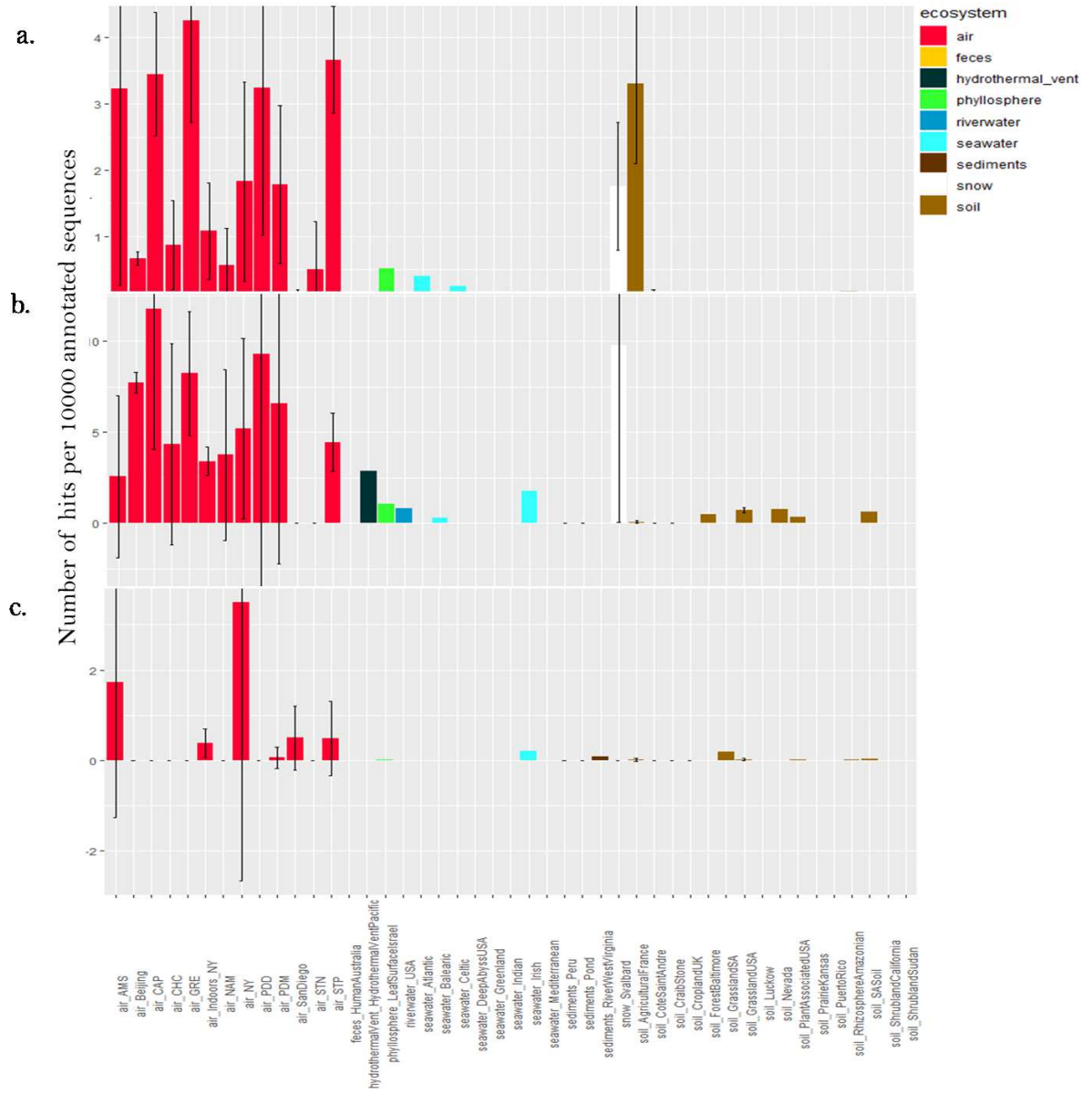


Figure 8: Average number of hits of desiccation response related functions per 10000 annotated sequences from (a) all sequences, (b) fungi-associated sequences and (c) bacteria-associated sequences per site. Colors indicate the ecosystems in which the sites belong to. For the sites including several metagenomes, the standard deviation was added.

Two proteins (lipoate synthase and chromosome plasmid partitioning protein ParA) related to stress response showed high relative concentrations in bacteria-associated sequences of a few air samples compared to the other ecosystems (**Fig. S7** in supplementary data), although the number of sequences related to these proteins was on average in air was not higher than other ecosystems (*P*=0.33 and 0.48 for lipoate synthase and chromosome plasmid partitioning protein ParA, respectively).

Discussion

Air metagenomes were characterized by a relatively high percentage of fungi-associated sequences. This percentage varied across the different sites with a higher percentage at terrestrial sites whose surrounding landscapes were vegetated like Grenoble (GRE), Puy-de-Dôme (PDD) and Pic-du-midi (PDM) (surrounding landscapes in **Fig. S1** in supplementary data). This percentage was also relatively high at the marine site Amsterdam-Island (AMS), where fungi might come from the ocean and/or the vegetated surfaces of the small island. A high percentage of fungi-associated sequences was also observed in air samples from Beijing, New York and San Diego retrieved from other aeromicrobiology studies, suggesting that our particular DNA extraction method set-up specifically for quartz fiber filter[25] was not causing a larger lysis of fungal cells and spores. Similarly, the sequencing technology (Illumina MiSeq) could not have been responsible for the larger percentage of fungi-associated sequences observed in our samples as the Beijing as well as the New York and San Diego air samples originate from an Illumina HiSeq and 454 sequencing technology, respectively. qPCR results on the 16S rRNA gene (bacterial cell concentration estimation) and on the 18S rRNA gene (fungal cell concentration estimation) on our air samples in comparison to soil samples (Côte Saint André, France) showed that the ratio between fungal and bacterial cell number was much higher (from 4.5 to 160 times higher for the most vegetated site Grenoble) in air than in soil (**Table S6** in supplementary data). The ratio between fungal and bacterial cell number might be higher in air than in other environments like soil[37], which might explain the relatively higher percentage of fungi-associated sequences observed in air metagenomes.

Fungi are expected to be found mostly as fungal spores in air, although the relative concentration of fungal spores and fungal hyphae fragments in air is unknown. Our results showed that the number of sporulation-related functions was higher in air than the other ecosystems (with the exception of snow, phyllosphere and Celtic seawater). While fungal hyphae are not expected to be particularly resistant to extreme conditions such as UV radiation, fungal spores are specifically produced to resist and survive overall adverse atmospheric conditions[24]. Their thick membrane and dehydrated nature make them particularly resistant to abiotic atmospheric conditions such as UV radiation, oxidative stress, desiccation as well as osmotic stress. **Figure 9** presents a conceptual model that could explain the higher ratio between fungi and bacteria observed in air. During aerosolization and aerial

transport, bacteria and fungi might be under stress and might undergo a physical selection with the survival of the most resistant cells to the adverse atmospheric conditions (*i.e.* UV radiation, desiccation etc.) and the death of non-resistant cells. As fungi (and especially fungal spores) might be naturally more resistant and adapted to atmospheric conditions than bacteria, we expect a larger decline of bacterial cells compared to fungal cells and spores in air. This might have as a consequence an increase in the ratio between fungi and bacteria from their non-atmospheric origins (*i.e.* the surrounding ecosystems) (**Fig. 9**).

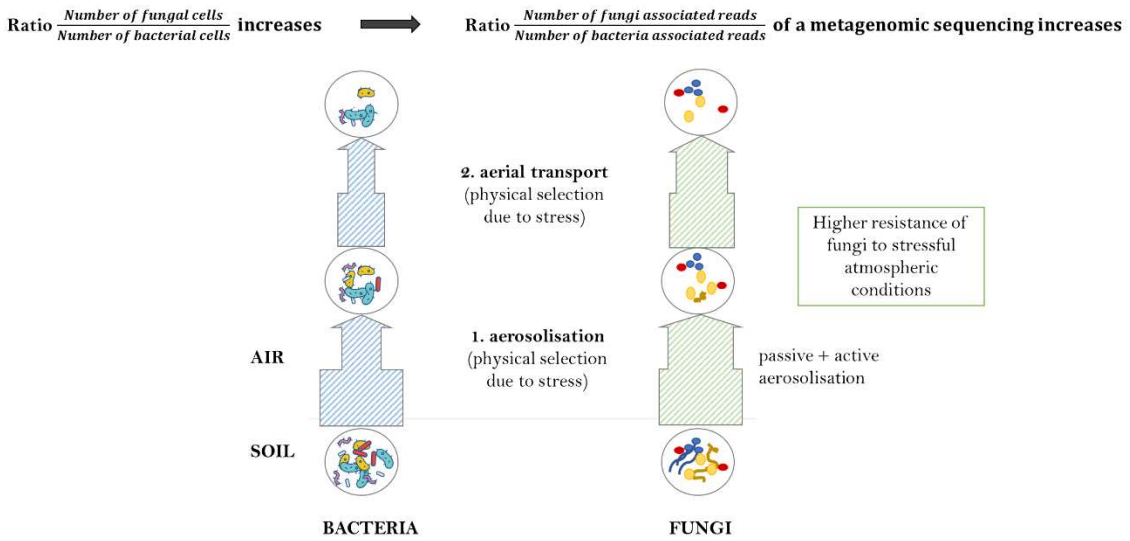


Figure 9: Conceptual model that might explain the higher ratio between fungi and bacteria in air relatively to airborne microbial cell sources such as soil and ocean water. The thickness of the arrows represents the impact of the physical selection on airborne microbial cells, that is the more microbial cells survive the physical selection (during aerosolization and aerial transport), the thicker is the arrow.

Metagenomic investigations of different ecosystems revealed a specific functional potential signature of their associated microbial communities[15,18]. These specific signatures are thought to result from microbial adaptation and/or physical selection to the environmental abiotic conditions[17,21,22] and are a reflection of the high relative abundances of genes coding for specific functions essential for microorganisms to survive and develop in these environments. For example, microbial metagenomes of human feces were characterized by high relative abundances of sequences annotated as beta-glucosidases that are associated with high

intestinal concentrations of complex glycosides; and microbial metagenomes of oceans were enriched in sequences annotated as enzymes catalyzing DMSP (dimethylsulfoniopropionate), which is an organosulfur compound produced by phytoplankton[15]. Our results showed a clear separation between seawater, river water, human feces and almost all the soil samples (which grouped with the sediment and snow samples at the scale used here) on the PCoA analysis based on the microbial functional potential (**Fig. 3**). For air microbiomes, the PCoA analyses showed that the individual air samples did not group for each site (**Fig. S2** in supplementary data) and that they did not form a cluster separated from the other ecosystems based on the overall microbial functional potential averaged per site (**Fig. 3**). Air samples seemed to group with their underlying ecosystems. While terrestrial air samples (GRE, NAM, CHC, STP, PDD, PDM) grouped with snow, soil and sediment samples, the marine (AMS), coastal (CAP) and arctic (STN) air samples were closer to surface seawater and river water samples. This result suggests that airborne microbial functional potential (and especially metabolic functional potential as SEED functional classes included mainly metabolic functions and few stress response related functions) might be dependent on the ecosystems from which microorganisms are aerosolized. Moreover, it seems that bacteria-associated sequences are mainly responsible for the distribution of the samples on the PCoA analysis (as we can see when we do the PCoA only with the fungi-associated sequences) although they were in smaller numbers compared to fungi-associated sequences for many of the air samples (*i.e.* STP, GRE, AMS, PDD, CAP, Beijing etc.). The low statistical weight of fungi-associated sequences relative to the overall sequences might be related to their very low richness nt in terms of functional genes that might have created the spreading of the samples on the PCoA based on the fungi-associated sequences (**Fig. S3** in supplementary data).

The high variability between the air sites and between air samples of the same site could be explained by the variability in the inputs from the different surrounding landscapes. Our previous paper showed that local inputs were the main sources of planetary boundary layer microorganisms and that local meteorology (especially the wind direction) had a major impact on the temporal variability of airborne microbial communities by affecting which of the different local sources were upwind[5] (Chapter 3). Our results did not show a specific (metabolic) functional potential signature for the atmosphere, which was rather mainly driven by the surrounding landscapes. Our results are consistent with both a pre-metabolic

adaptation of airborne microorganisms to the chemicals of the sources (*i.e.* surrounding landscapes) and a potential metabolic adaptation to these chemicals in the atmosphere.

Atmospheric chemistry is dependent on the underlying ecosystem chemistry since the main sources of atmospheric chemicals are Earth surface emissions. Yet, the oxidizing conditions of the atmosphere[38] might lead to rapid transformations of atmospheric chemicals by photochemical reactions. These specific atmospheric chemical reactions (*i.e.* photochemical) produce species which, with the gases like CH_4, characterize the atmosphere (O_3, H_2O_2, OH etc.). Although some strains from cloud water origin have been shown to metabolize and grow on culture medium in the presence of H_2O_2[10], radical species and their precursors are reactive compounds and might not easily serve as energy and carbon sources for microorganisms[39,40]. Our results on specific metabolic related functions showed that functions related to methane monooxygenase activity (CH_4 degradation) and hydrogen peroxide catabolism (H_2O_2 degradation) were present in air but not in higher proportion than in other ecosystems (**Fig. 4** and **Fig. S6** in supplementary data). Reactive compounds can cause oxidative stress to airborne microorganisms. In association to adverse physical conditions like UV radiation and desiccation, oxidative compounds might create more of a physical stress than provide a new metabolic source for airborne microorganisms. Laboratory investigations of cultivable microorganisms of an airborne origin showed the presence of particularly resistant strains under stressful conditions similar to the atmospheric ones (*i.e.* similar UV radiation levels; different oxidative conditions)[23,41]. However, no study has shown whether these apparently adapted cells represented the majority of airborne microorganisms. Since the overall SEED functional classes included mainly metabolic functions, specific stress related functions using GO (Gene Ontology) terms were also evaluated. We observed that on average, air showed more stress-related functions (UV response, desiccation, cell death and oxidative stress response related functions) than the other ecosystems. However, when the annotated sequences were separated to sequences belonging to fungal and bacterial genomes, the bacteria and fungi-associated sequences from air samples did not show necessarily a significant higher concentration of stress-related functions compared to the samples coming from other ecosystems (**Fig. 4** to **9**). Yet, stress related functions remained present in air and when looking at more specific proteins, they were occasionally found in relatively high concentration in air samples such as lipoate synthase and chromosome plasmid partitioning

protein ParA which might play a role in oxidative stress[42,43] and more generally in stress resistance and adaptability of microorganisms[44,45], respectively (**Fig. S7** in supplementary data). The detection of metagenomic sequences annotated as specific proteins in air samples remains difficult because of the low microbial biomass recovered. That is why we examined the presence and concentration of global functions (*i.e*. UV protection related functions, oxidative stress response related functions etc.) rather than specific functional genes.

The constant and large inputs of microbial cells to the planetary boundary layer as well as their relatively short residence time (a few hours to a few days) in the air might have hindered the observation of the potential adaptation (physical selection and/or microbial adaptation) of airborne microorganisms to the stressful atmospheric conditions as well as to the atmospheric chemicals as discussed above. This issue might be addressed by investigating microbial functional potential in the free troposphere (preferentially high enough above the ground so as not to be influenced by the surface) where the microbial fluxes are smaller than in the planetary boundary layer and where microbial airborne residence time might last much longer than in the boundary layer. This troposphere approach might help in determining the role of stress in the atmosphere and validate our conceptual model on the physical stress of microbial cells taking place during aerosolization and aerial transport selecting the resistant cells (**Fig. 10**). Another explanation might be due to the metagenomic approach that samples both living and dead cells. Aerosolization has been shown to be particularly stressful and even lethal for microorganisms[46–48]. Our results on death-related functions would be consistent with air harboring a higher percentage of both bacterial and fungal dead cells compared to other ecosystems (**Fig. 8**). The functional potential from the dead cells in air might have a greater weight on the overall functional potential observed and lead to the dilution of the functional potential of the actual living cells that have adapted to atmospheric conditions. This might apply for both the overall functional potential discussed previously and the stress-related functions.

Conclusion

We conducted the first global comparative metagenomic analysis to characterize the microbial functional potential signature in the planetary boundary layer. Air samples showed no specific signature of microbial functional potential which was mainly correlated to the

surrounding landscapes. However, air samples were characterized by a relatively high percentage of fungi-associated sequences compared to the source ecosystems (soil, surface seawater etc.). The relatively higher concentrations of fungi in air drove the higher proportions of stress-related functions observed in air metagenomes. Fungal cells and specifically fungal spores are innately resistant entities well adapted to atmospheric conditions and which might survive better aerosolization and aerial transport than bacterial cells. Stress related functions were present in airborne bacteria but rarely in higher concentrations compared to the bacterial communities in other ecosystems. However, the constant flux of microbial cells to the planetary boundary layer might have complicated the determination of a physical selection and/or microbial adaptation of airborne microorganisms, especially bacterial communities, and further investigations are needed.

References

1. Zhen, Q. *et al.* Meteorological factors had more impact on airborne bacterial communities than air pollutants. *Sci. Total Environ.* **601–602**, 703–712 (2017).
2. Cáliz, J., Triadó-Margarit, X., Camarero, L. & Casamayor, E. O. A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12229–12234 (2018).
3. Els, N., Baumann-Stanzer, K., Larose, C., Vogel, T. M. & Sattler, B. Beyond the planetary boundary layer: Bacterial and fungal vertical biogeography at Mount Sonnblick, Austria. *Geo: Geography and Environment* **6**, e00069 (2019).
4. Innocente, E. *et al.* Influence of seasonality, air mass origin and particulate matter chemical composition on airborne bacterial community structure in the Po Valley, Italy. *Sci. Total Environ.* **593–594**, 677–687 (2017).
5. Tignat-Perrier, R. *et al.* Global airborne microbial communities controlled by surrounding landscapes and wind conditions. *Scientific Reports* (under review).
6. Ariya, P. A., Nepotchatykh, O., Ignatova, O. & Amyot, M. Microbiological degradation of atmospheric organic compounds. *Geophysical Research Letters* **29**, 34-1-34–4 (2002).
7. Amato, P. *et al.* Microbial population in cloud water at the Puy de Dôme: Implications for the chemistry of clouds. *Atmospheric Environment* **39**, 4143–4153 (2005).
8. Amato, P. *et al.* A fate for organic acids, formaldehyde and methanol in cloud water: their biotransformation by micro-organisms. *Atmospheric Chemistry and Physics* **7**, 4159–4169 (2007).
9. Vaïtilingom, M. *et al.* Contribution of Microbial Activity to Carbon Chemistry in Clouds. *Appl Environ Microbiol* **76**, 23–29 (2010).
10. Vaïtilingom, M. *et al.* Potential impact of microbial activity on the oxidant capacity and organic carbon budget in clouds. *PNAS* **110**, 559–564 (2013).
11. Hill, K. A. *et al.* Processing of atmospheric nitrogen by clouds above a forest environment. *Journal of Geophysical Research: Atmospheres* **112**, (2007).

12. Amato, P. *et al.* Metatranscriptomic exploration of microbial functioning in clouds. *Sci Rep* **9**, 1–12 (2019).
13. Cao, C. *et al.* Inhalable Microorganisms in Beijing's PM2.5 and PM10 Pollutants during a Severe Smog Event. *Environ. Sci. Technol.* **48**, 1499–1507 (2014).
14. Yooseph, S. *et al.* Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468**, 60–66 (2010).
15. Delmont, T. O. *et al.* Metagenomic mining for microbiologists. *ISME J* **5**, 1837–1843 (2011).
16. Xie, W. *et al.* Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *The ISME Journal* **5**, 414–426 (2011).
17. Li, Y., Zheng, L., Zhang, Y., Liu, H. & Jing, H. Comparative metagenomics study reveals pollution induced changes of microbial genes in mangrove sediments. *Scientific Reports* **9**, 5739 (2019).
18. Tringe, S. G. *et al.* Comparative metagenomics of microbial communities. *Science* **308**, 554–557 (2005).
19. Brune, A., Frenzel, P. & Cypionka, H. Life at the oxic–anoxic interface: microbial activities and adaptations. *FEMS Microbiol Rev* **24**, 691–710 (2000).
20. Palenik, B. *et al.* Genome sequence of Synechococcus CC9311: Insights into adaptation to a coastal environment. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13555–13559 (2006).
21. Hindré, T., Knibbe, C., Beslon, G. & Schneider, D. New insights into bacterial adaptation through *in vivo* and *in silico* experimental evolution. *Nature Reviews Microbiology* **10**, 352–365 (2012).
22. Rey, O., Danchin, E., Mirouze, M., Loot, C. & Blanchet, S. Adaptation to Global Change: A Transposable Element–Epigenetics Perspective. *Trends in Ecology & Evolution* **31**, 514–526 (2016).
23. Joly, M. *et al.* Survival of microbial isolates from clouds toward simulated atmospheric stress factors. *Atmospheric Environment* **117**, 92–98 (2015).
24. Huang, M. & Hull, C. M. Sporulation: How to survive on planet Earth (and beyond). *Curr Genet* **63**, 831–838 (2017).
25. Dommergue, A. *et al.* Methods to investigate the global atmospheric microbiome. *Front. Microbiol.* **10**, (2019).
26. Shannan, S., Collins, K. & Emanuel, W. R. Global mosaics of the standard MODIS land cover type data. (2014).
27. Friedl, M. A. *et al.* Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment* **83**, 287–302 (2002).
28. Fierer, N., Jackson, J. A., Vilgalys, R. & Jackson, R. B. Assessment of Soil Microbial Community Structure by Use of Taxon-Specific Quantitative PCR Assays. *Appl. Environ. Microbiol.* **71**, 4117–4120 (2005).
29. Chemidlin Prévost-Bouré, N. *et al.* Validation and application of a PCR primer set to quantify fungal communities in the soil environment by real-time quantitative PCR. *PLoS ONE* **6**, e24166 (2011).
30. Oksanen, J. *et al.* Community Ecology Package. (2019).
31. Hadley, W. & Winston, C. Create Elegant Data Visualisations Using the Grammar of Graphics. (2019).
32. Wickham, H. Flexibly Reshape Data: A Reboot of the Reshape Packa. (2017).

33. Huson, D. H., Richter, D. C., Mitra, S., Auch, A. F. & Schuster, S. C. Methods for comparative metagenomics. *BMC Bioinformatics* **10 Suppl 1**, S12 (2009).
34. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* **15**, R46 (2014).
35. Donovan, P. D., Gonzalez, G., Higgins, D. G., Butler, G. & Ito, K. Identification of fungi in shotgun metagenomics datasets. *PLOS ONE* **13**, e0192898 (2018).
36. Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
37. Malik, A. A. *et al.* Soil Fungal:Bacterial Ratios Are Linked to Altered Carbon Cycling. *Front. Microbiol.* **7**, (2016).
38. Seinfeld, J. H. & Pandis, S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, *3rd Edition | Environmental Chemistry |Chemistry | Subjects | Wiley*. (Wiley-Interscience, 1998).
39. Imlay, J. A. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nat Rev Microbiol* **11**, 443–454 (2013).
40. Sheng, H., Nakamura, K., Kanno, T., Sasaki, K. & Niwano, Y. Microbicidal Activity of Artificially Generated Hydroxyl Radicals. in *Interface Oral Health Science 2014* (eds. Sasaki, K., Suzuki, O. & Takahashi, N.) 203–215 (Springer Japan, 2015).
41. Yang, Y., Yokobori, S. & Yamagishi, A. UV-resistant bacteria isolated from upper troposphere and lower stratosphere. *Biol.Sci.Space* **22**, (2008).
42. Bunik, V. I. 2-Oxo acid dehydrogenase complexes in redox regulation. *Eur. J. Biochem.* **270**, 1036–1042 (2003).
43. Allary, M., Lu, J. Z., Zhu, L. & Prigge, S. T. Scavenging of the cofactor lipoate is essential for the survival of the malaria parasite Plasmodium falciparum. *Mol Microbiol* **63**, 1331–1344 (2007).
44. Shoeb, E. *et al.* Horizontal gene transfer of stress resistance genes through plasmid transport. *World J. Microbiol. Biotechnol.* **28**, 1021–1025 (2012).
45. Zhang, H. *et al.* Stress resistance, motility and biofilm formation mediated by a 25kb plasmid pLMSZ08 in Listeria monocytogenes. *Food Control* **94**, 345–352 (2018).
46. Alsved, M. *et al.* Effect of Aerosolization and Drying on the Viability of Pseudomonas syringae Cells. *Front Microbiol* **9**, 3086 (2018).
47. Thomas, R. J. *et al.* The Cell Membrane as a Major Site of Damage during Aerosolization of Escherichia coli. *Appl. Environ. Microbiol.* **77**, 920–925 (2011).
48. Zhen, H., Han, T., Fennell, D. E. & Mainelis, G. Release of Free DNA by Membrane-Impaired Bacterial Aerosols Due to Aerosolization and Air Sampling. *Appl. Environ. Microbiol.* **79**, 7780–7789 (2013).

Chapter 5: Microbial functional signature in the atmospheric boundary layer

Romie Tignat-Perrier[1,2], Aurélien Dommergue[1], Alban Thollot[1], Olivier Magand[1], Timothy M. Vogel[2], Catherine Larose[2]

[1]Institut des Géosciences de l'Environnement, Université Grenoble Alpes, CNRS, IRD, Grenoble INP, Grenoble, France

[2]Environmental Microbial Genomics, Laboratoire Ampère, École Centrale de Lyon, Université de Lyon, Écully, France

Supplementary Information

Supplementary tables

Table S1: Number of samples, ecosystem, sequencing technology, database and accession number, number of sequences per sample (mean + standard deviation), percentage of fungi and bacteria-associated sequences per site and percentage of annotated sequences (mean + standard deviation) per site.

Site	Country/ Ocean	Information on the site	Number of samples	Ecosystem	Sequencing technology	Database and reference numbers or study	Total sequence number	Annotated sequence number by eggNOG-Mapper	Fungi-associated sequences	Percentage of fungi-associated sequences over total read number	Percentage of fungi-associated sequences over fungi-and bacteria-associated sequence number	Percentage of fungi-associated sequences annotated by eggNOG-Mapper	Bacteria-associated sequence number	Percentage of bacteria-associated sequences over total sequence number	Percentage of bacteria-associated sequences over fungi-and bacteria-associated sequence number	Percentage of bacteria-associated sequences annotated by eggNOG- Mapper
Air Amsterdam-Island (AMS)	Sub-Antarctica	marine, remote	3	air	MiSeq	our study	97881 +- 93551	17676 +- 15935	4152 +- 4089	4.1 +- 0.2	71 +- 4	60 +- 2	1670 +- 1549	1.7 +- 0.4	29 +- 4	57 +- 2
Air Beijing	China	urban	2	air	HiSeq	mgm4516366.3, mgm4516459.3	2248590 +- 177298	141849 +- 9576	250843 +- 91619	11 +- 3.2	52 +- 10	17 +- 1	226290 +- 8208	10.1 +- 1.2	48 +- 10	39 +- 1
Air Cape-Point (CAP)	South-Africa	coastal	2	air	MiSeq	our study	90043 +- 6341	20479 +- 1447	7227 +- 5972	7.8 +- 6.1	56 +- 29	50 +- 12	4530 +- 1286	5.1 +- 1.8	44 +- 29	61 +- 1
Air Chacaltaya (CHC)	Bolivia	high-altitude mountain peak	9	air	MiSeq	our study	103239 +- 54187	32699 +- 18131	3479 +- 2580	3.5 +- 1.2	27 +- 24	61 +- 7	11113 +- 6411	10.2 +- 2	73 +- 24	67 +- 2
Air Grenoble (GRE)	France	urban	9	air	MiSeq	our study	248064 +- 158109	42853 +- 30690	24234 +- 15561	9.7 +- 0.5	79 +- 10	48 +- 3	7082 +- 8061	2.7 +- 1.6	21 +- 10	59 +- 5
Air New York indoors (indoors_NY)	USA	indoors	4	air	454	SRR1000232, SRR1000254, SRR999213, SRR999215	400997 +- 49680	126245 +- 12742	36858 +- 16604	9.2 +- 4.1	52 +- 17	47 +- 8	32035 +- 10923	8.3 +- 3.7	48 +- 17	70 +- 9
Air Namco (NAM)	China	high-altitude plateau, semi-arid	9	air	MiSeq	our study	149952 +- 92976	48012 +- 36340	2958 +- 1910	2.1 +- 1.1	19 +- 12	68 +- 9	15901 +- 13188	10 +- 2.6	81 +- 12	69 +- 2
Air New York (NY)	USA	urban, coastal	6	air	454	SRR1000260, SRR1000269, SRR999217, SRR999218, SRR999219, SRR999220	521791 +- 277049	99566 +- 51023	85350 +- 41529	20 +- 11.6	56 +- 8	46 +- 41	69161 +- 38301	18.1 +- 14.5	44 +- 8	37 +- 47
Air Puy-de-Dôme (PDD)	France	continental, moutain peak	10	air	MiSeq	our study	396666 +- 364681	65304 +- 68592	25029 +- 32432	5.8 +- 2.5	62 +- 16	50 +- 6	13112 +- 12079	3.9 +- 3	38 +- 16	56 +- 4
Air Pic-du-Midi (PDM)	France	high-altitude moutain peak	13	air	MiSeq	our study	186766 +- 197396	33016 +- 31653	8676 +- 8233	5 +- 2.1	50 +- 10	54 +- 4	8115 +- 7687	5.5 +- 2.6	50 +- 10	63 +- 5
Air San-Diego	USA	urban coastal	2	air	454	SRR999211, SRR999212	781206 +- 65608	229544 +- 3651	5960 +- 4678	0.8 +- 0.7	36 +- 1	44 +- 3	10318 +- 7888	1.4 +- 1.1	64 +- 1	54 +- 13
Air Station-Nord (STN)	Greenland	polar	2	air	MiSeq	our study	23463 +- 24385	5935 +- 5528	1276 +- 1702	3.6 +- 3.5	24 +- 18	59 +- 38	2460 +- 2606	10.2 +- 0.5	76 +- 18	65 +- 3
Air Storm-Peak (STP)	USA	high-altitude mountain peak	6	air	MiSeq	our study	469168 +- 242715	111530 +- 58582	99110 +- 56113	20.7 +- 2.6	88 +- 4	53 +- 2	12559 +- 7185	2.8 +- 1.1	12 +- 4	59 +- 5
Human feces Sydney	Australia	human feces	1	feces	HiSeq	mgm4675774.3	2111825 +- NA	503328 +- NA	28357 +- NA	1.3 +- NA	3 +- NA	29 +- NA	929682 +- NA	44 +- NA	97 +- NA	35 +- NA
Hydrothermal vent Pacific	Pacific Ocean	deep sea water	1	hydrothermal vent	454	mgm4481541.3	758485 +- NA	409294 +- NA	12501 +- NA	1.6 +- NA	9 +- NA	56 +- NA	128294 +- NA	16.9 +- NA	91 +- NA	84 +- NA
Leaf surface Israel	Israel	leaf surface	1	phyllosphere	HiSeq	mgm4534773.3	12272440 +- NA	3274840 +- NA	460612 +- NA	3.8 +- NA	37 +- NA	23 +- NA	784441 +- NA	6.4 +- NA	63 +- NA	70 +- NA

Seawater Atlantic sea	North Atlantic Ocean	surface sea water	1	seawater	MiSeq	mgm4719942.3	1198007 +- NA	301519 +- NA	54904 +- NA	4.6 +- NA	36 +- NA	11 +- NA	99510 +- NA	8.3 +- NA	64 +- NA	51 +- NA
Seawater Balearic sea	Balearic Sea	surface sea water	1	seawater	MiSeq	mgm4719938.3	878884 +- NA	531765 +- NA	50995 +- NA	5.8 +- NA	36 +- NA	60 +- NA	89238 +- NA	10.2 +- NA	64 +- NA	88 +- NA
Seawater Celtic sea	Celtic Sea	surface sea water	1	seawater	MiSeq	mgm4719941.3	1702779 +- NA	283691 +- NA	510855 +- NA	11.4 +- NA	53 +- NA	18 +- NA	112212 +- NA	6.6 +- NA	47 +- NA	55 +- NA
Deep abyss USA	USA	deep sea water	1	seawater	454	mgm4668304.3	2016153 +- NA	330380 +- NA	1895 +- NA	0.1 +- NA	2 +- NA	68 +- NA	90930 +- NA	4.5 +- NA	98 +- NA	75 +- NA
Seawater Greenland sea	Greenland Sea	surface sea water	1	seawater	MiSeq	mgm4719947.3	1358477 +- NA	687315 +- NA	79494 +- NA	5.9 +- NA	28 +- NA	51 +- NA	200245 +- NA	14.7 +- NA	72 +- NA	77 +- NA
Seawater Indian sea	Indian Sea	surface sea water	1	seawater	MiSeq	mgm4719994.3	126564 +- NA	56086 +- NA	2527 +- NA	2 +- NA	11 +- NA	55 +- NA	21431 +- NA	16.9 +- NA	89 +- NA	89 +- NA
Seawater Irish sea	Irish Sea	surface sea water	1	seawater	MiSeq	mgm4719940.3	1362228 +- NA	538286 +- NA	27051 +- NA	2 +- NA	28 +- NA	21 +- NA	69952 +- NA	5.1 +- NA	72 +- NA	70 +- NA
Seawater Mediterranean sea	Mediterranean Sea, Eastern bassin	surface sea water	1	seawater	MiSeq	mgm4719936.3	1241549 +- NA	686395 +- NA	17384 +- NA	1.4 +- NA	11 +- NA	43 +- NA	133889 +- NA	10.8 +- NA	89 +- NA	87 +- NA
River water Colorado	USA	river water	1	riverwater	HiSeq	mgm4628878.3	1392059 +- NA	575616 +- NA	39553 +- NA	2.8 +- NA	19 +- NA	31 +- NA	169972 +- NA	12.2 +- NA	81 +- NA	76 +- NA
River sediments West Virginia	USA	river sediments	1	sediments	HiSeq	mgm4589537.3	2072338 +- NA	569631 +- NA	58672 +- NA	2.8 +- NA	26 +- NA	28 +- NA	166079 +- NA	8 +- NA	74 +- NA	71 +- NA
Sediments Peru	Peru	seafloor sediments	2	sediments	454	mgm4440960.3, mgm4459940.3	126239 +- 35030	3606 +- 829	1780 +- 2323	1.2 +- 1.5	53 +- 47	13 +- 16	539 +- 18	0.4 +- 0.1	47 +- 47	33 +- 11
Sediments Pond	France	shallow pond sediments	4	sediments	MiSeq	our lab; Sanchez-Cid *et al.*, submitted	44236 +- 17409	15568 +- 7046	1450 +- 991	3.2 +- 1.2	9 +- 3	80 +- 5	16704 +- 12966	36.7 +- 18.1	91 +- 3	56 +- 8
Snow Svalbard	Norway	fresh snow, artic	7	snow	MiSeq	our lab; Bergk-Pinto *et al.*, under review	226368 +- 70313	49428 +- 23023	3788 +- 1542	1.7 +- 0.6	18 +- 6	61 +- 12	18469 +- 8699	8.5 +- 3.2	82 +- 6	54 +- 5
Agricultural soil France	France	agricultural soil	3	soil	MiSeq	mgm4705012.3, mgm4697958.3, mgm4697957.3	8209393 +- 2836681	5913635 +- 1326693	194398 +- 108165	2.3 +- 0.5	61 +- 3	60 +- 13	129460 +- 84783	1.5 +- 0.5	39 +- 3	65 +- 11
Soil Cote Saint Andre	France	agricultural soil	6	soil	MiSeq	our lab; Sanchez-Cid *et al.*, under review	174898 +- 80968	54841 +- 22225	1638 +- 640	1 +- 0.1	9 +- 1	82 +- 2	16806 +- 6846	9.8 +- 0.6	91 +- 1	67 +- 1
Soil CraibStone	Scottland	agricultural soil	5	soil	MiSeq	our lab; Sanchez-Cid *et al.*, submitted	128815 +- 82837	41175 +- 22413	1452 +- 953	1.1 +- 0.1	9 +- 1	78 +- 4	15472 +- 11906	11.5 +- 1.2	91 +- 1	65 +- 1
Cropland UK	United Kingdom	cropland	2	soil	MiSeq	mgm4781436.3, mgm4781437.3	485163 +- 163475	304642 +- 104503	9970 +- 3861	2 +- 0.1	13 +- 0	88 +- NA	66040 +- 25119	13.5 +- 0.6	87 +- 0	95 +- 0
Forest soil Baltimore	USA	temperate deciduous broadleaf forest soil	1	soil	MiSeq	mgm4819073.3	4600481 +- NA	959764 +- NA	34207 +- NA	0.7 +- NA	12 +- NA	62 +- NA	260826 +- NA	5.7 +- NA	88 +- NA	79 +- NA
Grassland USA	South-Africa	tropical grassland	1	soil	MiSeq	mgm4819072.3	2519738 +- NA	638149 +- NA	22974 +- NA	0.9 +- NA	8 +- NA	63 +- NA	279551 +- NA	11.1 +- NA	92 +- NA	79 +- NA
Grassland USA	USA	temperate grassland	2	soil	MiSeq	mgm4623641.3, mgm4623640.3	12195227 +- 1436683	697967 +- 122932	1753640 +- 1903	14.5 +- 1.7	42 +- 1	2 +- 0	2381163 +- 99911	19.6 +- 1.5	58 +- 1	13 +- 2
Soil Lucknow India	India	soil	1	soil	454	mgm4461840.3	1187505 +- NA	658023 +- NA	53911 +- NA	4.5 +- NA	14 +- NA	91 +- NA	322160 +- NA	27.1 +- NA	86 +- NA	91 +- NA
Soil Nevada	USA	soil	1	soil	454	mgm4451106.3	1248623 +- NA	725892 +- NA	29880 +- NA	2.4 +- NA	8 +- NA	84 +- NA	326929 +- NA	26.2 +- NA	92 +- NA	91 +- NA
Plant soil USA	USA	soil	1	soil	HiSeq	mgm4767414.3	17632266 +- NA	1425603 +- NA	253827 +- NA	1.4 +- NA	15 +- NA	33 +- NA	1473019 +- NA	8.4 +- NA	85 +- NA	52 +- NA
Prairie Kansas	USA	prairie soil	1	soil	MiSeq	mgm4477804.3	5348832 +- NA	343702 +- NA	27412 +- NA	0.5 +- NA	9 +- NA	51 +- NA	270464 +- NA	5.1 +- NA	91 +- NA	56 +- NA
Soil PuertoRico	Puerto Rico	subtropical forest	1	soil	454	mgm4446153.3	725275 +- NA	452063 +- NA	10926 +- NA	1.5 +- NA	11 +- NA	88 +- NA	85868 +- NA	11.8 +- NA	89 +- NA	95 +- NA
Rhizosphere Amazonia	Brazil	tropical broadleaf forest	1	soil	HiSeq	mgm4723911.3	8884491 +- NA	1415017 +- NA	2075 +- NA	0 +- NA	0 +- NA	52 +- NA	1027815 +- NA	11.6 +- NA	100 +- NA	60 +- NA
Soil South-Africa	South-Africa	tropical grassland	1	soil	MiSeq	mgm4819068.3	2757834 +- NA	759329 +- NA	25123 +- NA	0.9 +- NA	6 +- NA	64 +- NA	414056 +- NA	15 +- NA	94 +- NA	79 +- NA
Shrubland California	USA	shrubland	1	soil	MiSeq	mgm4806895.3	2213724 +- NA	47591 +- NA	3742 +- NA	0.2 +- NA	1 +- NA	11 +- NA	243528 +- NA	11 +- NA	98 +- NA	7 +- NA
Shrubland Sudan	Sudan	shrubland	1	soil	MiSeq	mgm4806896.3	185966 +- NA	1169 +- NA	1181 +- NA	0.6 +- NA	18 +- NA	7 +- NA	5381 +- NA	2.9 +- NA	82 +- NA	11 +- NA

Table S2: Standardized collected air volume and sampling starting date of each air sample that we collected for this study.

site	Sample name	Standardized collected air volume (m^3)	Sampling starting date (ending date 7 days after, same hour) (month/day/year)
AMS	AMS_10/09/2016	5232	10/09/16
AMS	AMS_08/10/2016	4449	08/10/16
AMS	AMS_21/10/2016	5059	10/21/16
CAP	CAP_21/10/2016	4679	10/21/16
CAP	CAP_28/10/2016	545	10/28/16
CHC	CHC_23/09/2016	1148	09/23/16
CHC	CHC_21/10/2016	1154	10/21/16
CHC	CHC_28/10/2016	1220	10/28/16
CHC	CHC_01/07/2016	1284	01/07/16
CHC	CHC_08/07/2016	1284	08/07/16
CHC	CHC_15/07/2016	1289	07/15/16
CHC	CHC_12/08/2016	1160	12/08/16
CHC	CHC_19/08/2016	1158	08/19/16
CHC	CHC_02/09/2016	1156	02/09/16
GRE	GRE_03/07/2017	4688	03/07/17
GRE	GRE_10/07/2017	4717	10/07/17
GRE	GRE_17/07/2017	4677	07/17/17
GRE	GRE_24/07/2017	4718	07/24/17
GRE	GRE_31/07/2017	4665	07/31/17
GRE	GRE_07/08/2017	4762	07/08/17
GRE	GRE_14/08/2017	4729	08/14/17
GRE	GRE_21/08/2017	4707	08/21/17
GRE	GRE_04/09/2017	4742	04/09/17
NAM	NAM_17/05/2017	5511	05/17/17

NAM	NAM_25/05/2017	5503	05/25/17
NAM	NAM_02/06/2017	5513	02/06/17
NAM	NAM_13/06/2017	4218	06/13/17
NAM	NAM_20/06/2017	5418	06/20/17
NAM	NAM_29/06/2017	5415	06/29/17
NAM	NAM_07/07/2017	5483	07/07/17
NAM	NAM_14/07/2017	5413	07/14/17
NAM	NAM_21/07/2017	5465	07/21/17
PDD	PDD_07/06/2017	8761	06/07/17
PDD	PDD_14/06/2017	8360	06/14/17
PDD	PDD_21/06/2017	8672	06/21/17
PDD	PDD_28/06/2017	9012	06/28/17
PDD	PDD_02/08/2017	7399	08/02/17
PDD	PDD_09/08/2017	9926	08/09/17
PDD	PDD_30/05/2017	10232	05/30/17
PDD	PDD_12/07/2017	8713	07/12/17
PDD	PDD_19/07/2017	8620	07/19/17
PDD	PDD_26/07/2017	8664	07/26/17
PDM	PDM_20/06/2016	9664	06/20/16
PDM	PDM_29/06/2016	6803	06/29/16
PDM	PDM_12/07/2016	7550	12/07/16
PDM	PDM_19/07/2016	8040	07/19/16
PDM	PDM_26/07/2016	7794	07/26/16
PDM	PDM_02/08/2016	8103	02/08/16
PDM	PDM_09/08/2016	7747	08/09/16
PDM	PDM_16/08/2016	8100	08/16/16
PDM	PDM_23/08/2016	7956	08/23/16
PDM	PDM_13/09/2016	7931	09/13/16
PDM	PDM_20/09/2016	7853	09/20/16

PDM	PDM_06/09/2016	7867	06/09/16
PDM	PDM_16/08/2016	8100	08/16/16
STN	STN_27/03/2017	5153	03/27/17
STN	STN_15/05/2017	5246	05/15/17
STP	STP_14/07/2017	11213	07/14/17
STP	STP_21/07/2017	9333	07/21/17
STP	STP_28/07/2017	5702	07/28/17
STP	STP_11/08/2017	5702	08/11/17
STP	STP_18/08/2017	5702	08/18/17
STP	STP_25/08/2017	5702	08/25/17

Table S3: Functional richness and evenness after rarefaction per site, based on the SEED functional classes. For site including several samples, the mean and standard deviation have been calculated.

Site	Ecosystem	ALL sequences				FUNGI-ASSOCIATED sequences				BACTERIA-ASSOCIATED sequences			
		Number of annotated sequences using Diamond and MEGAN6	Rarefaction	Functional richness after rarefaction	Functional evenness after rarefaction	Number of annotated sequences using Diamond and MEGAN6	Rarefaction	Functional richness after rarefaction	Functional evenness after rarefaction	Number of annotated sequences using Diamond and MEGAN6	Rarefaction	Functional richness after rarefaction	Functional evenness after rarefaction
air Amsterdam-Island (AMS)	air	3927 +- 3321	1737 +- 456	1087 +- 554	0.94 +- 0.02	81 +- 80	81 +- 80	66 +- 58	0.99 +- 0.01	554 +- 480	360 +- 193	270 +- 197	0.96 +- 0.01
air Beijing	air	180196 +- 11408	2000 +- 0	4060 +- 112	0.86 +- 0	5960 +- 214	500 +- 0	1129 +- 92	0.89 +- 0	82004 +- 5643	500 +- 0	2835 +- 58	0.87 +- 0
air Cape-Point (CAP)	air	8176 +- 4856	2000 +- 0	1634 +- 337	0.93 +- 0.02	211 +- 15	211 +- 15	162 +- 6	0.97 +- 0	1890 +- 726	500 +- 0	739 +- 192	0.95 +- 0.01
air Chacaltaya (CHC)	air	15853 +- 8907	1848 +- 456	2062 +- 714	0.92 +- 0.02	380 +- 219	346 +- 175	223 +- 109	0.96 +- 0.02	5268 +- 3052	467 +- 99	1142 +- 461	0.93 +- 0.02
air Grenoble (GRE)	air	5765 +- 6870	1802 +- 297	1256 +- 700	0.94 +- 0.02	412 +- 382	308 +- 156	235 +- 153	0.97 +- 0.01	2193 +- 2949	445 +- 86	658 +- 528	0.96 +- 0.02
air indoors New York (indoors_NY)	air	32135 +- 11235	2000 +- 0	3302 +- 299	0.91 +- 0.01	1546 +- 802	500 +- 0	697 +- 206	0.95 +- 0.01	10067 +- 4782	500 +- 0	2183 +- 387	0.93 +- 0.01
air Namco (NAM)	air	23081 +- 19276	2000 +- 0	2280 +- 478	0.91 +- 0.01	596 +- 495	381 +- 114	287 +- 136	0.95 +- 0.02	7600 +- 6515	500 +- 0	1300 +- 372	0.92 +- 0.02
air New York (NY)	air	5481 +- 4324	1639 +- 561	1384 +- 849	0.89 +- 0.04	286 +- 231	275 +- 217	150 +- 109	0.91 +- 0.07	769 +- 622	362 +- 205	446 +- 335	0.94 +- 0.02
air Puy-de-Dôme (PDD)	air	11053 +- 9757	1976 +- 75	1700 +- 775	0.93 +- 0.02	656 +- 748	300 +- 198	297 +- 239	0.96 +- 0.03	4277 +- 4138	500 +- 0	989 +- 617	0.94 +- 0.03
air Pic-du-Midi (PDM)	air	9422 +- 8988	1769 +- 490	1575 +- 778	0.94 +- 0.02	363 +- 354	267 +- 185	218 +- 163	0.98 +- 0.02	3252 +- 3366	460 +- 101	832 +- 511	0.95 +- 0.02
air San Diego	air	14573 +- 8176	2000 +- 0	2021 +- 81	0.9 +- 0	184 +- 191	184 +- 191	91 +- 66	0.96 +- 0.04	1737 +- 1841	468 +- 46	628 +- 429	0.95 +- 0.01
air Station-Nord (STN)	air	2863 +- 2408	1580 +- 594	956 +- 547	0.95 +- 0.02	111 +- 111	111 +- 111	88 +- 81	0.98 +- 0.01	1089 +- 1085	411 +- 127	486 +- 400	0.96 +- 0.01
air Storm-Peak (STP)	air	11763 +- 7684	2000 +- 0	1865 +- 519	0.92 +- 0.01	973 +- 537	476 +- 58	392 +- 131	0.94 +- 0.02	3757 +- 3006	500 +- 0	971 +- 409	0.94 +- 0.02
human feces Sydney	feces	560641 +- NA	2000 +- NA	2802 +- NA	0.82 +- NA	3591 +- NA	500 +- NA	317 +- NA	0.61 +- NA	304338 +- NA	500 +- NA	2557 +- NA	0.87 +- NA
leaf surface Israel	phyllosphere	1042866 +- NA	2000 +- NA	4292 +- NA	0.87 +- NA	10644 +- NA	500 +- NA	1336 +- NA	0.85 +- NA	247373 +- NA	500 +- NA	3165 +- NA	0.88 +- NA
river water USA	river water	295902 +- NA	2000 +- NA	3550 +- NA	0.87 +- NA	6142 +- NA	500 +- NA	1001 +- NA	0.89 +- NA	76857 +- NA	500 +- NA	2497 +- NA	0.87 +- NA
seawater Balearic sea	seawater	340618 +- NA	2000 +- NA	2957 +- NA	0.87 +- NA	17554 +- NA	500 +- NA	1825 +- NA	0.9 +- NA	55185 +- NA	500 +- NA	1823 +- NA	0.86 +- NA
seawater Celtic sea	seawater	335790 +- NA	2000 +- NA	3325 +- NA	0.87 +- NA	11736 +- NA	500 +- NA	1831 +- NA	0.91 +- NA	41271 +- NA	500 +- NA	1964 +- NA	0.87 +- NA
deep abyss USA	seawater	333284 +- NA	2000 +- NA	3649 +- NA	0.87 +- NA	1006 +- NA	500 +- NA	390 +- NA	0.92 +- NA	43055 +- NA	500 +- NA	2590 +- NA	0.9 +- NA
seawater Greenland sea	seawater	417826 +- NA	2000 +- NA	3223 +- NA	0.88 +- NA	21407 +- NA	500 +- NA	2164 +- NA	0.91 +- NA	97826 +- NA	500 +- NA	2376 +- NA	0.87 +- NA
hydrothermal Vent Pacific Ocean	Hydrothermal vent	217796 +- NA	2000 +- NA	3621 +- NA	0.87 +- NA	3855 +- NA	500 +- NA	950 +- NA	0.9 +- NA	62974 +- NA	500 +- NA	2533 +- NA	0.89 +- NA
seawater Indian sea	seawater	40507 +- NA	2000 +- NA	2178 +- NA	0.9 +- NA	799 +- NA	500 +- NA	476 +- NA	0.96 +- NA	12001 +- NA	500 +- NA	1183 +- NA	0.93 +- NA
seawater Irish sea	seawater	287629 +- NA	2000 +- NA	3283 +- NA	0.88 +- NA	2394 +- NA	500 +- NA	662 +- NA	0.89 +- NA	32848 +- NA	500 +- NA	1843 +- NA	0.86 +- NA
seawater Mediterranean sea	seawater	381180 +- NA	2000 +- NA	3375 +- NA	0.87 +- NA	3999 +- NA	500 +- NA	898 +- NA	0.87 +- NA	74727 +- NA	500 +- NA	2112 +- NA	0.88 +- NA
seawater North Atlantic Ocean	seawater	206085 +- NA	2000 +- NA	3143 +- NA	0.87 +- NA	2956 +- NA	500 +- NA	771 +- NA	0.9 +- NA	35702 +- NA	500 +- NA	1663 +- NA	0.89 +- NA
sediments Peru	sediments	6348 +- 1251	2000 +- 0	1138 +- 7	0.92 +- 0	13 +- 4	13 +- 4	13 +- 4	1 +- 0	133 +- 66	133 +- 66	90 +- 33	0.97 +- 0.02
sediments Pond	sediments	10569 +- 5084	2000 +- 0	1791 +- 252	0.92 +- 0.01	647 +- 422	441 +- 106	367 +- 149	0.97 +- 0.01	7038 +- 5091	500 +- 0	1364 +- 392	0.93 +- 0.01
river sediments West Virginia	sediments	315551 +- NA	2000 +- NA	3869 +- NA	0.87 +- NA	10660 +- NA	500 +- NA	1223 +- NA	0.88 +- NA	74774 +- NA	500 +- NA	2381 +- NA	0.88 +- NA
snow Svalbard	snow	26069 +- 15249	2000 +- 0	2243 +- 498	0.91 +- 0.01	648 +- 360	418 +- 152	329 +- 143	0.96 +- 0.02	8702 +- 4544	500 +- 0	1317 +- 356	0.92 +- 0.02

agricultural soil France	soil	907295 +- 258019	2000 +- 0	764 +- 47	0.72 +- 0.02	8707 +- 6045	442 +- 116	129 +- 74	0.68 +- 0.1	7044 +- 2095	500 +- 0	118 +- 55	0.49 +- 0.05
soil Cote Saint Andre	soil	34947 +- 16517	2000 +- 0	2552 +- 263	0.9 +- 0.01	783 +- 355	491 +- 14	350 +- 92	0.94 +- 0.01	8680 +- 3716	500 +- 0	1418 +- 191	0.91 +- 0.01
soil CraibStone	soil	27629 +- 18784	2000 +- 0	2406 +- 327	0.91 +- 0.01	668 +- 473	465 +- 46	336 +- 127	0.95 +- 0.02	7980 +- 6183	500 +- 0	1346 +- 289	0.92 +- 0.01
cropland UK	soil	122625 +- 44684	2000 +- 0	3490 +- 132	0.87 +- 0	5106 +- 1863	500 +- 0	1001 +- 112	0.89 +- 0.02	34986 +- 13329	500 +- 0	2254 +- 103	0.89 +- 0.01
forest soil Baltimore	soil	606468 +- NA	2000 +- NA	3998 +- NA	0.86 +- NA	33130 +- 27346	500 +- 0	1438 +- 259	0.84 +- 0.02	334568 +- 288964	500 +- 0	2518 +- 4	0.86 +- 0
soil Puerto Rico	soil	170277 +- NA	2000 +- NA	3607 +- NA	0.86 +- NA	6380 +- NA	500 +- NA	971 +- NA	0.88 +- NA	45321 +- NA	500 +- NA	2229 +- NA	0.87 +- NA
grassland SA	soil	191711 +- 271120	1000 +- 1414	2093 +- 2959	0.43 +- 0.61	14252 +- 9262	500 +- 0	1216 +- 199	0.85 +- 0.01	305911 +- 260034	500 +- 0	3147 +- 93	0.9 +- 0.02
grassland USA	soil	926515 +- 144620	2000 +- 0	3496 +- 45	0.86 +- 0	23892 +- 4105	500 +- 0	1539 +- 35	0.84 +- 0	242586 +- 44860	500 +- 0	2332 +- 42	0.86 +- 0
soil lucknow India	soil	303263 +- NA	2000 +- NA	3971 +- NA	0.88 +- NA	19077 +- NA	500 +- NA	1357 +- NA	0.87 +- NA	145062 +- NA	500 +- NA	2707 +- NA	0.89 +- NA
soil Nevada	soil	305066 +- NA	2000 +- NA	3528 +- NA	0.87 +- NA	12931 +- NA	500 +- NA	1324 +- NA	0.85 +- NA	147232 +- NA	500 +- NA	2465 +- NA	0.89 +- NA
plant soil USA	soil	1839508 +- NA	2000 +- NA	3610 +- NA	0.87 +- NA	65330 +- NA	500 +- NA	1663 +- NA	0.84 +- NA	612710 +- NA	500 +- NA	2491 +- NA	0.87 +- NA
prairie Kansas	soil	475235 +- NA	2000 +- NA	3393 +- NA	0.86 +- NA	11087 +- NA	500 +- NA	1155 +- NA	0.85 +- NA	119659 +- NA	500 +- NA	2282 +- NA	0.87 +- NA
rhizosphere Amazonia	soil	1061420 +- NA	2000 +- NA	3330 +- NA	0.81 +- NA	337 +- NA	337 +- NA	157 +- NA	0.91 +- NA	412216 +- NA	500 +- NA	1912 +- NA	0.79 +- NA
soil South-Africa	soil	464748 +- NA	2000 +- NA	4239 +- NA	0.88 +- NA	8883 +- NA	500 +- NA	1092 +- NA	0.85 +- NA	187173 +- NA	500 +- NA	3252 +- NA	0.91 +- NA
shrubland California	soil	65837 +- NA	2000 +- NA	2693 +- NA	0.87 +- NA	298 +- NA	298 +- NA	126 +- NA	0.92 +- NA	14541 +- NA	500 +- NA	1757 +- NA	0.89 +- NA
shrubland Sudan	soil	1751 +- NA	1751 +- NA	864 +- NA	0.96 +- NA	77 +- NA	77 +- NA	68 +- NA	0.99 +- NA	727 +- NA	500 +- NA	437 +- NA	0.97 +- NA

Table S4: qPCR on the 16s rRNA gene and on the 18S rRNA gene on air and soil samples, and ratio between these qPCR. Means and standard deviations were calculated on three (Cote Saint André), nine (Amsterdam-Island and Namco) and ten (Grenoble) samples. qPCR results for the air samples have already been presented in Tignat-Perrier *et al.* (under review) (Chapter 3).

	qPCR 18S rRNA gene number	**qPCR 16S rRNA gene number**	**Ratio qPCR16S/qPCR18S**
AIR SAMPLES			
NAM (Namco)	$4.97x10^{3} \pm 3.44x10^{3}$	$3.56x10^{6} \pm 3.01x10^{6}$	716
GRE (Grenoble)	$5.28x10^{4} \pm 3.61x10^{4}$	$1.20x10^{6} \pm 9.38x10^{5}$	23
AMS (Amsterdam-Island)	$7.51x10^{3} \pm 6.96x10^{3}$	$1.49x10^{5} \pm 9.17x10^{4}$	20
SOIL SAMPLES			
Côte Saint André	$1.13x10^{3} \pm 2.9x10^{2}$	$3.70x10^{6} \pm 1.9x10^{6}$	3265

Table S5: Top 50 of the SEED functions observed in the air samples (mean +/- standard deviation) considering all the sequences (*i.e.* bacteria and fungi-associated sequences).

Function	air_AMS	air_Beijing	air_CAP	air_CHC	air_GRE	air_IndoorsNY	air_NAM	air_NY	air_PDD	air_PDM	air_SanDiego	air_STN	air_STP
\5-FCL-like protein\""	2.01 +/- 0.66	2.04 +/- 0.12	1.6 +/- 0.13	1.9 +/- 0.13	1.84 +/- 0.33	1.62 +/- 0.08	1.97 +/- 0.15	1.53 +/- 0.54	1.68 +/- 0.32	1.78 +/- 0.37	1.41 +/- 0.12	1.47 +/- 0.24	1.81 +/- 0.31
\Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)\""	2.3 +/- 0.09	1.35 +/- 0.3	1.29 +/- 0.47	1.46 +/- 0.25	1.62 +/- 0.27	1.17 +/- 0.2	1.51 +/- 0.12	1.12 +/- 0.53	1.79 +/- 1.2	1.58 +/- 0.22	0.91 +/- 0.02	2.01 +/- 0.45	1.86 +/- 0.42
\TonB-dependent receptor\""	1.3 +/- 0.44	0.36 +/- 0.03	1.02 +/- 0.4	0.94 +/- 0.14	1.07 +/- 0.46	0.91 +/- 0.15	0.98 +/- 0.12	0.72 +/- 0.12	0.99 +/- 0.51	1.03 +/- 0.28	0.45 +/- 0.04	1.06 +/- 0.52	0.83 +/- 0.27
\3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)\""	0.65 +/- 0.28	0.55 +/- 0.02	0.75 +/- 0.01	0.9 +/- 0.11	0.8 +/- 0.26	0.62 +/- 0.06	1.04 +/- 0.11	0.56 +/- 0.17	0.75 +/- 0.25	0.79 +/- 0.24	0.45 +/- 0.05	0.84 +/- 0.09	0.91 +/- 0.21
\COG2363\""	0.53 +/- 0.3	0.5 +/- 0.02	0.58 +/- 0.14	0.51 +/- 0.08	0.58 +/- 0.16	0.56 +/- 0.07	0.49 +/- 0.11	0.26 +/- 0.2	0.54 +/- 0.13	0.55 +/- 0.13	0.56 +/- 0.26	0.48 +/- 0.07	0.45 +/- 0.05
\Aldehyde dehydrogenase (EC 1.2.1.3)\""	0.43 +/- 0.21	0.4 +/- 0.03	0.44 +/- 0.1	0.31 +/- 0.04	0.58 +/- 0.13	0.29 +/- 0.07	0.39 +/- 0.06	0.68 +/- 1.11	0.42 +/- 0.2	0.35 +/- 0.17	0.29 +/- 0.04	0.25 +/- 0.23	0.47 +/- 0.09
\Adenylate cyclase (EC 4.6.1.1)\""	0.26 +/- 0.18	0.17 +/- 0.02	0.18 +/- 0.11	0.36 +/- 0.15	0.27 +/- 0.18	0.29 +/- 0.06	0.79 +/- 0.07	0.25 +/- 0.24	0.41 +/- 0.23	0.46 +/- 0.14	0.23 +/- 0.06	0.53 +/- 0.02	0.32 +/- 0.1
\Beta-galactosidase (EC 3.2.1.23)\""	0.15 +/- 0.06	0.23 +/- 0.05	0.34 +/- 0.03	0.34 +/- 0.15	0.25 +/- 0.14	0.22 +/- 0.05	0.3 +/- 0.06	1.41 +/- 2.13	0.28 +/- 0.16	0.3 +/- 0.21	0.22 +/- 0.08	0.18 +/- 0.25	0.35 +/- 0.07
\DNA-directed RNA polymerase beta' subunit (EC 2.7.7.6)\""	0.58 +/- 0.13	0.53 +/- 0.03	0.39 +/- 0.07	0.38 +/- 0.07	0.3 +/- 0.13	0.15 +/- 0.04	0.38 +/- 0.04	0.09 +/- 0.08	0.51 +/- 0.42	0.39 +/- 0.11	0.32 +/- 0.11	0.29 +/- 0.17	0.39 +/- 0.07
\Aspartate aminotransferase (EC 2.6.1.1)\""	0.53 +/- 0.08	0.39 +/- 0.02	0.35 +/- 0.08	0.38 +/- 0.07	0.35 +/- 0.13	0.43 +/- 0.07	0.33 +/- 0.05	0.4 +/- 0.32	0.35 +/- 0.14	0.34 +/- 0.09	0.29 +/- 0.01	0.37 +/- 0.28	0.33 +/- 0.03
\Cobalt-zinc-cadmium resistance protein CzcA\""	0.17 +/- 0.1	0.3 +/- 0.08	0.2 +/- 0.01	0.36 +/- 0.06	0.26 +/- 0.13	0.49 +/- 0.14	0.26 +/- 0.04	0.72 +/- 0.24	0.4 +/- 0.18	0.38 +/- 0.15	0.24 +/- 0.01	0.63 +/- 0.41	0.22 +/- 0.04
\DNA topoisomerase I (EC 5.99.1.2)\""	0.07 +/- 0.07	0.17 +/- 0.01	0.08 +/- 0.05	0.12 +/- 0.05	0.08 +/- 0.03	0.12 +/- 0.02	0.16 +/- 0.04	3.08 +/- 4.05	0.14 +/- 0.11	0.1 +/- 0.06	0.16 +/- 0.01	0.27 +/- 0.1	0.1 +/- 0.03
\High-affnity carbon uptake protein Hat/HatR\""	0.29 +/- 0.04	0.16 +/- 0.09	0.09 +/- 0.05	0.19 +/- 0.09	1.27 +/- 0.94	0.17 +/- 0.06	0.35 +/- 0.18	0.31 +/- 0.18	0.23 +/- 0.15	0.14 +/- 0.09	0.08 +/- 0.04	0.12 +/- 0.17	0.32 +/- 0.19
\Beta-lactamase\""	0.47 +/- 0.18	0.15 +/- 0.01	0.28 +/- 0.04	0.29 +/- 0.12	0.23 +/- 0.12	0.26 +/- 0.06	0.41 +/- 0.09	0.62 +/- 0.64	0.24 +/- 0.1	0.41 +/- 0.11	0.26 +/- 0	0.2 +/- 0.16	0.3 +/- 0.05
\DNA-directed RNA polymerase beta subunit (EC 2.7.7.6)\""	0.39 +/- 0.19	0.49 +/- 0.03	0.35 +/- 0.1	0.35 +/- 0.07	0.24 +/- 0.11	0.15 +/- 0.06	0.34 +/- 0.09	0.04 +/- 0.04	0.44 +/- 0.48	0.3 +/- 0.09	0.31 +/- 0.12	0.51 +/- 0.26	0.3 +/- 0.04
\Butyryl-CoA dehydrogenase (EC 1.3.99.2)\""	0.34 +/- 0.22	0.29 +/- 0.04	0.31 +/- 0.15	0.3 +/- 0.13	0.36 +/- 0.16	0.19 +/- 0.05	0.33 +/- 0.04	0.21 +/- 0.1	0.3 +/- 0.17	0.3 +/- 0.09	0.19 +/- 0.02	0.36 +/- 0.23	0.35 +/- 0.09
\Aspartyl-tRNA(Asn) amidotransferase subunit A (EC 6.3.5.6)\""	0.16 +/- 0.15	0.08 +/- 0	0.04 +/- 0.06	0.1 +/- 0.14	0.04 +/- 0.05	0.04 +/- 0.03	0.06 +/- 0.03	3.02 +/- 4.05	0.04 +/- 0.03	0.06 +/- 0.05	0.11 +/- 0.02	0.11 +/- 0.09	0.07 +/- 0.03
\FIG039061: hypothetical protein related to heme utilization\""	0.28 +/- 0.19	0.34 +/- 0.02	0.36 +/- 0.12	0.29 +/- 0.12	0.32 +/- 0.09	0.33 +/- 0.06	0.26 +/- 0.05	0.17 +/- 0.14	0.31 +/- 0.13	0.3 +/- 0.11	0.37 +/- 0.04	0.2 +/- 0.09	0.34 +/- 0.07
\DNA polymerase III alpha subunit (EC 2.7.7.7)\""	0.32 +/- 0.11	0.39 +/- 0	0.47 +/- 0.15	0.34 +/- 0.06	0.27 +/- 0.13	0.24 +/- 0.02	0.36 +/- 0.03	0.11 +/- 0.09	0.27 +/- 0.09	0.32 +/- 0.2	0.29 +/- 0.06	0.17 +/- 0.12	0.24 +/- 0.09
\D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)\""	0.48 +/- 0.09	0.26 +/- 0	0.27 +/- 0.06	0.29 +/- 0.08	0.3 +/- 0.13	0.29 +/- 0.02	0.3 +/- 0.06	0.1 +/- 0.08	0.29 +/- 0.12	0.3 +/- 0.12	0.27 +/- 0.14	0.49 +/- 0.41	0.25 +/- 0.08
\3-ketoacyl-CoA thiolase (EC 2.3.1.16)\""	0.43 +/- 0.06	0.41 +/- 0.03	0.24 +/- 0.1	0.28 +/- 0.04	0.31 +/- 0.17	0.19 +/- 0.02	0.3 +/- 0.04	0.16 +/- 0.1	0.25 +/- 0.11	0.29 +/- 0.14	0.23 +/- 0.12	0.37 +/- 0.16	0.34 +/- 0.1
\Arylsulfatase (EC 3.1.6.1)\""	0.61 +/- 0.09	0.11 +/- 0.03	1.17 +/- 0.93	0.17 +/- 0.03	0.16 +/- 0.09	0.08 +/- 0.02	0.28 +/- 0.09	0.09 +/- 0.05	0.15 +/- 0.11	0.19 +/- 0.11	2.98 +/- 0.08	0.2 +/- 0.09	0.12 +/- 0.05
\Enoyl-CoA hydratase (EC 4.2.1.17)\""	0.31 +/- 0.1	0.24 +/- 0	0.24 +/- 0.05	0.31 +/- 0.12	0.27 +/- 0.11	0.26 +/- 0.07	0.34 +/- 0.07	0.16 +/- 0.06	0.24 +/- 0.06	0.3 +/- 0.11	0.21 +/- 0.04	0.46 +/- 0.04	0.27 +/- 0.04
\Acetyl-coenzyme A synthetase (EC 6.2.1.1)\""	0.15 +/- 0.07	0.38 +/- 0.01	0.36 +/- 0.21	0.35 +/- 0.09	0.29 +/- 0.09	0.21 +/- 0.04	0.27 +/- 0.04	0.27 +/- 0.16	0.24 +/- 0.11	0.28 +/- 0.07	0.2 +/- 0.12	0.23 +/- 0.04	0.31 +/- 0.03
\Copper-translocating P-type ATPase (EC 3.6.3.4)\""	0.12 +/- 0.05	0.35 +/- 0.05	0.18 +/- 0.04	0.2 +/- 0.09	0.24 +/- 0.17	0.62 +/- 0.21	0.15 +/- 0.04	0.7 +/- 0.23	0.28 +/- 0.18	0.23 +/- 0.14	0.19 +/- 0.14	0.29 +/- 0.17	0.17 +/- 0.08

\diguanylate cyclase/phosphodiesterase (GGDEF & EAL domains) with PAS/PAC sensor(s)\""	0.11 +/- 0.14	0.14 +/- 0.01	0.12 +/- 0.07	0.25 +/- 0.08	0.24 +/- 0.14	0.26 +/- 0.02	0.38 +/- 0.06	0.13 +/- 0.08	0.3 +/- 0.18	0.33 +/- 0.14	0.16 +/- 0.08	0.5 +/- 0.1	0.28 +/- 0.08
\Chaperone protein DnaK\""	0.29 +/- 0.2	0.34 +/- 0.05	0.26 +/- 0.06	0.24 +/- 0.06	0.3 +/- 0.17	0.17 +/- 0.03	0.28 +/- 0.03	0.16 +/- 0.1	0.3 +/- 0.14	0.29 +/- 0.12	0.29 +/- 0.09	0.37 +/- 0.21	0.23 +/- 0.03
\UDP-glucose 4-epimerase (EC 5.1.3.2)\""	0.27 +/- 0.12	0.22 +/- 0.01	0.22 +/- 0.08	0.35 +/- 0.09	0.24 +/- 0.06	0.19 +/- 0.04	0.4 +/- 0.06	0.09 +/- 0.08	0.22 +/- 0.1	0.29 +/- 0.09	0.17 +/- 0.05	0.41 +/- 0.15	0.23 +/- 0.12
\Excinuclease ABC subunit A\""	0.11 +/- 0.11	0.35 +/- 0.02	0.4 +/- 0.14	0.33 +/- 0.12	0.21 +/- 0.07	0.22 +/- 0.04	0.28 +/- 0.07	0.2 +/- 0.12	0.18 +/- 0.11	0.37 +/- 0.27	0.34 +/- 0.03	0.2 +/- 0.03	0.2 +/- 0.05
\Aconitate hydratase (EC 4.2.1.3)\""	0.27 +/- 0.16	0.39 +/- 0.03	0.34 +/- 0.27	0.26 +/- 0.06	0.32 +/- 0.17	0.17 +/- 0.04	0.27 +/- 0.07	0.19 +/- 0.1	0.28 +/- 0.13	0.27 +/- 0.1	0.19 +/- 0.04	0.09 +/- 0.12	0.26 +/- 0.09
\Transcription-repair coupling factor\""	0.11 +/- 0.14	0.33 +/- 0.01	0.29 +/- 0.12	0.27 +/- 0.05	0.13 +/- 0.11	0.22 +/- 0.08	0.22 +/- 0.05	0.64 +/- 0.94	0.26 +/- 0.09	0.25 +/- 0.08	0.18 +/- 0.12	0.23 +/- 0.17	0.25 +/- 0.06
\Acriflavin resistance protein\""	0.32 +/- 0.17	0.16 +/- 0	0.41 +/- 0.01	0.26 +/- 0.1	0.16 +/- 0.09	0.32 +/- 0.12	0.23 +/- 0.06	0.23 +/- 0.12	0.2 +/- 0.11	0.31 +/- 0.08	0.5 +/- 0.05	0.33 +/- 0.15	0.2 +/- 0.1
\Alcohol dehydrogenase (EC 1.1.1.1)\""	0.24 +/- 0.24	0.26 +/- 0.01	0.16 +/- 0.08	0.25 +/- 0.05	0.25 +/- 0.09	0.27 +/- 0.04	0.27 +/- 0.07	0.22 +/- 0.16	0.24 +/- 0.09	0.26 +/- 0.1	0.11 +/- 0.02	0.16 +/- 0.11	0.31 +/- 0.05
\Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39)\""	0.53 +/- 0.11	0.16 +/- 0.01	0.28 +/- 0.14	0.22 +/- 0.11	0.36 +/- 0.17	0.15 +/- 0.05	0.25 +/- 0.06	0.19 +/- 0.12	0.2 +/- 0.08	0.21 +/- 0.11	0.22 +/- 0.04	0.2 +/- 0.03	0.3 +/- 0.07
\Glutamate synthase [NADPH] large chain (EC 1.4.1.13)\""	0.45 +/- 0.29	0.38 +/- 0.02	0.34 +/- 0.15	0.26 +/- 0.06	0.2 +/- 0.11	0.17 +/- 0.05	0.27 +/- 0.03	0.1 +/- 0.07	0.19 +/- 0.14	0.26 +/- 0.09	0.2 +/- 0.03	0.28 +/- 0.16	0.24 +/- 0.05
\Carbamoyl-phosphate synthase large chain (EC 6.3.5.5)\""	0.14 +/- 0.06	0.38 +/- 0	0.24 +/- 0.09	0.25 +/- 0.07	0.2 +/- 0.05	0.16 +/- 0.03	0.24 +/- 0.04	0.12 +/- 0.07	0.27 +/- 0.28	0.31 +/- 0.17	0.26 +/- 0.08	0.28 +/- 0.16	0.17 +/- 0.1
\Heat shock protein 60 family chaperone GroEL\""	0.2 +/- 0.06	0.34 +/- 0	0.48 +/- 0.5	0.28 +/- 0.08	0.2 +/- 0.13	0.14 +/- 0.01	0.28 +/- 0.04	0.08 +/- 0.07	0.26 +/- 0.09	0.23 +/- 0.11	0.48 +/- 0.04	0.15 +/- 0.22	0.2 +/- 0.08
\Thioredoxin reductase (EC 1.8.1.9)\""	0.2 +/- 0.19	0.24 +/- 0.01	0.22 +/- 0.11	0.3 +/- 0.08	0.18 +/- 0.11	0.2 +/- 0.02	0.29 +/- 0.09	0.16 +/- 0.08	0.26 +/- 0.07	0.24 +/- 0.1	0.11 +/- 0.03	0.18 +/- 0.02	0.24 +/- 0.03
\ClpB protein\""	0.24 +/- 0.12	0.34 +/- 0.02	0.31 +/- 0.1	0.28 +/- 0.08	0.22 +/- 0.08	0.2 +/- 0.03	0.26 +/- 0.05	0.13 +/- 0.08	0.21 +/- 0.09	0.27 +/- 0.11	0.21 +/- 0.04	0.13 +/- 0.19	0.17 +/- 0.09
\DNA gyrase subunit A (EC 5.99.1.3)\""	0.33 +/- 0.11	0.3 +/- 0.02	0.32 +/- 0.12	0.25 +/- 0.05	0.21 +/- 0.09	0.16 +/- 0.02	0.19 +/- 0.04	0.16 +/- 0.17	0.22 +/- 0.1	0.22 +/- 0.12	0.37 +/- 0.04	0.3 +/- 0.06	0.2 +/- 0.04
\Multimodular transpeptidase-transglycosylase (EC 2.4.1.129) (EC 3.4.-.-)\""	0.14 +/- 0.05	0.2 +/- 0	0.2 +/- 0.02	0.26 +/- 0.12	0.23 +/- 0.17	0.24 +/- 0.04	0.25 +/- 0.05	0.17 +/- 0.09	0.23 +/- 0.09	0.24 +/- 0.15	0.09 +/- 0.01	0.18 +/- 0.02	0.23 +/- 0.08
\Alkaline phosphatase (EC 3.1.3.1)\""	0.21 +/- 0.04	0.07 +/- 0.01	0.19 +/- 0.03	0.18 +/- 0.06	0.16 +/- 0.09	0.24 +/- 0.04	0.35 +/- 0.05	0.19 +/- 0.08	0.2 +/- 0.12	0.24 +/- 0.14	0.4 +/- 0.12	0.29 +/- 0.2	0.18 +/- 0.08
\Asparagine synthetase [glutamine-hydrolyzing] (EC 6.3.5.4)\""	0.12 +/- 0.14	0.12 +/- 0	0.13 +/- 0.02	0.15 +/- 0.08	0.31 +/- 0.21	0.13 +/- 0.03	0.17 +/- 0.06	0.16 +/- 0.11	0.25 +/- 0.12	0.29 +/- 0.2	0.07 +/- 0	0.29 +/- 0.07	0.31 +/- 0.17
\Translation elongation factor G\""	0.24 +/- 0.17	0.35 +/- 0.03	0.24 +/- 0.11	0.25 +/- 0.11	0.22 +/- 0.05	0.11 +/- 0.04	0.22 +/- 0.04	0.08 +/- 0.06	0.25 +/- 0.21	0.24 +/- 0.16	0.23 +/- 0.08	0.21 +/- 0.29	0.2 +/- 0.05
\5-methyltetrahydrofolate--homocysteine methyltransferase (EC 2.1.1.13)\""	0.14 +/- 0.12	0.3 +/- 0.01	0.27 +/- 0.06	0.24 +/- 0.1	0.19 +/- 0.07	0.13 +/- 0.01	0.29 +/- 0.05	0.11 +/- 0.08	0.19 +/- 0.08	0.25 +/- 0.11	0.16 +/- 0.09	0.27 +/- 0.02	0.2 +/- 0.12
\Threonine dehydrogenase and related Zn-dependent dehydrogenases\""	0.13 +/- 0.05	0.37 +/- 0.1	0.16 +/- 0.02	0.27 +/- 0.11	0.15 +/- 0.1	0.1 +/- 0.04	0.3 +/- 0.06	0.21 +/- 0.26	0.17 +/- 0.1	0.25 +/- 0.15	0.01 +/- 0.02	0.07 +/- 0.09	0.29 +/- 0.09
\Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1)\""	0.25 +/- 0.02	0.29 +/- 0.01	0.21 +/- 0.02	0.18 +/- 0.08	0.19 +/- 0.1	0.14 +/- 0.07	0.25 +/- 0.04	0.15 +/- 0.1	0.18 +/- 0.07	0.23 +/- 0.12	0.24 +/- 0.04	0.23 +/- 0.04	0.22 +/- 0.08
\DNA polymerase I (EC 2.7.7.7)\""	0.14 +/- 0.03	0.2 +/- 0.01	0.4 +/- 0.03	0.22 +/- 0.05	0.18 +/- 0.13	0.21 +/- 0.02	0.24 +/- 0.06	0.13 +/- 0.11	0.19 +/- 0.04	0.22 +/- 0.15	0.19 +/- 0.01	0.2 +/- 0.03	0.17 +/- 0.09
\Catalase (EC 1.11.1.6)\""	0.08 +/- 0.09	0.28 +/- 0.03	0.15 +/- 0.11	0.22 +/- 0.05	0.22 +/- 0.14	0.15 +/- 0.05	0.17 +/- 0.05	0.13 +/- 0.09	0.28 +/- 0.18	0.22 +/- 0.12	0.05 +/- 0	0.05 +/- 0.08	0.21 +/- 0.06
\Type I restriction-modification system, restriction subunit R (EC 3.1.21.3)\""	0.26 +/- 0.14	0.25 +/- 0.01	0.25 +/- 0.03	0.23 +/- 0.04	0.17 +/- 0.04	0.31 +/- 0.09	0.21 +/- 0.03	0.23 +/- 0.1	0.15 +/- 0.14	0.14 +/- 0.07	0.06 +/- 0.03	0.35 +/- 0.13	0.19 +/- 0.08

Table S6: SIMPER analysis results on groups identified on the PCoA analysis based on all sequences.

G1 = air marine/air polar/air arctic samples

G2 = soil/sediment/snow/air terrestrial samples

G3 = surface seawater samples

Compared groups	**SEED functional functions (comparison of its relative abundance between the 2 groups)**
G1 versus G2	Arylsulfatase (EC 3.1.6.1) (G1 > G2)
	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3) (G1 > G2)
	Heme O synthase, protoheme IX farnesyltransferase (EC 2.5.1.-) COX10-CtaB (G1 > G2)
	TonB-dependent receptor (G1 < G2)
	5-FCL-like protein (G1 < G2)
G1 versus G3	Arylsulfatase (EC 3.1.6.1) (G1 > G3)
	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3) (G1 > G3)
	TonB-dependent receptor (G1 < G3)
	Choline-sulfatase (EC 3.1.6.6) (G1 > G3)
	Heme O synthase, protoheme IX farnesyltransferase (EC 2.5.1.-) COX10-CtaB (G1 > G3)
G3 versus G2	TonB-dependent receptor (G3 > G2)
	Beta-galactosidase (EC 3.2.1.23) (G3 > G2)
	Cobalt-zinc-cadmium resistance protein CzcA (G3 > G2)
	Sarcosine oxidase alpha subunit (EC 1.5.3.1) (G3 < G2)
	Aldehyde dehydrogenase (EC 1.2.1.3) (G3 < G2)

Supplementary figures

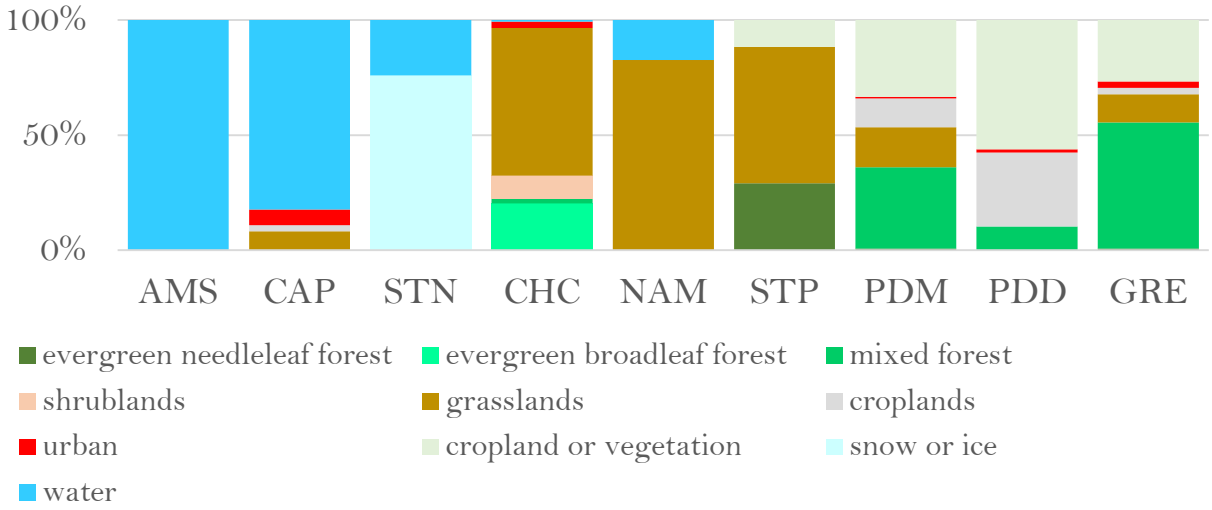


Figure S1: Surrounding landscapes of the sampling sites in a 50 km perimeter based on the land cover MODIS approach.

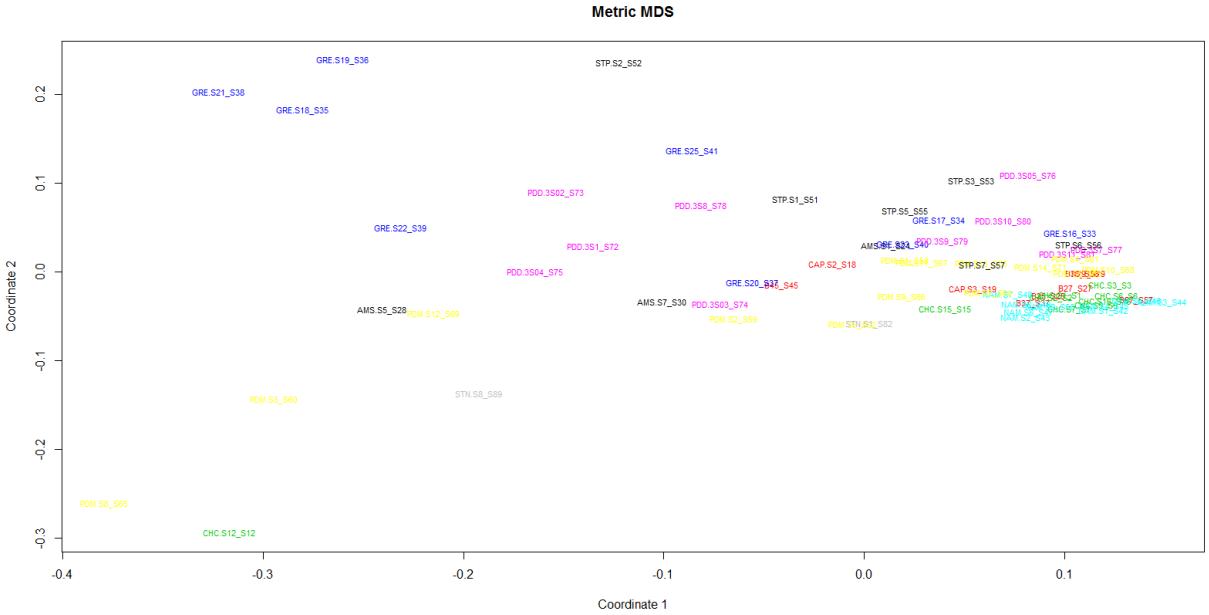


Figure S2: PCoA analysis of the Bray-Curtis dissimilarity matrix based on the functional potential structure of each air sample. The colors indicate the sites in which samples belong to (AMS: black, CAP:red, CHC: green, GRE: blue, NAM: blue-green, PDM: yellow, PDD: pink, STN: grey, STP: black).

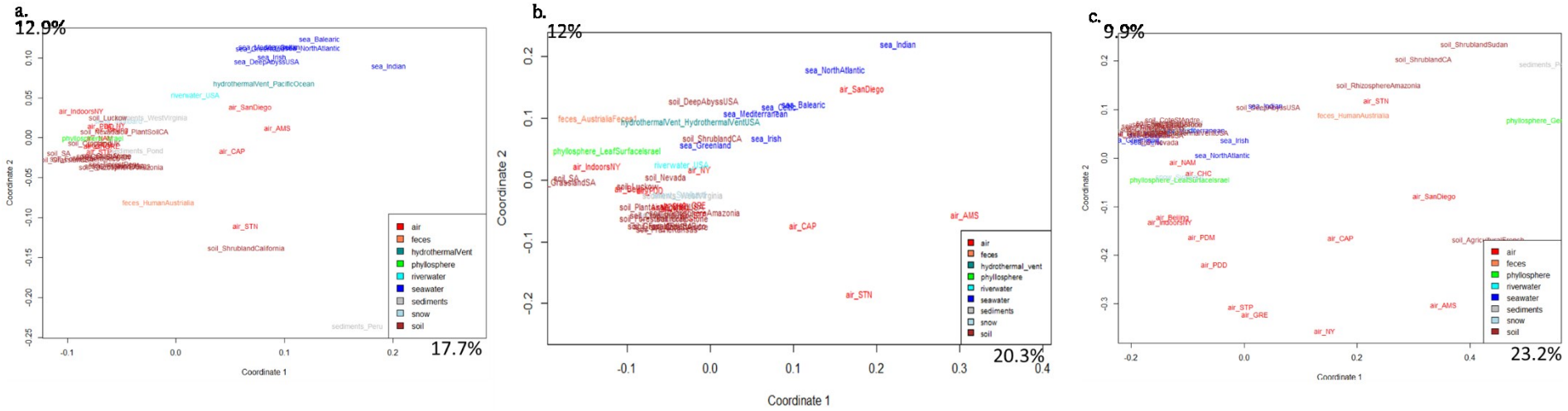


Figure S3: PCoA analysis of the Bray-Curtis dissimilarity matrix based on the functional potential structure of each site. All sequences (a), bacteria-associated sequences (b) and fungi-associated sequences (c) have been used for functional annotation. For the site including several metagenomes, the average profile was calculated. Colors indicate the ecosystems in which the sites belong to.

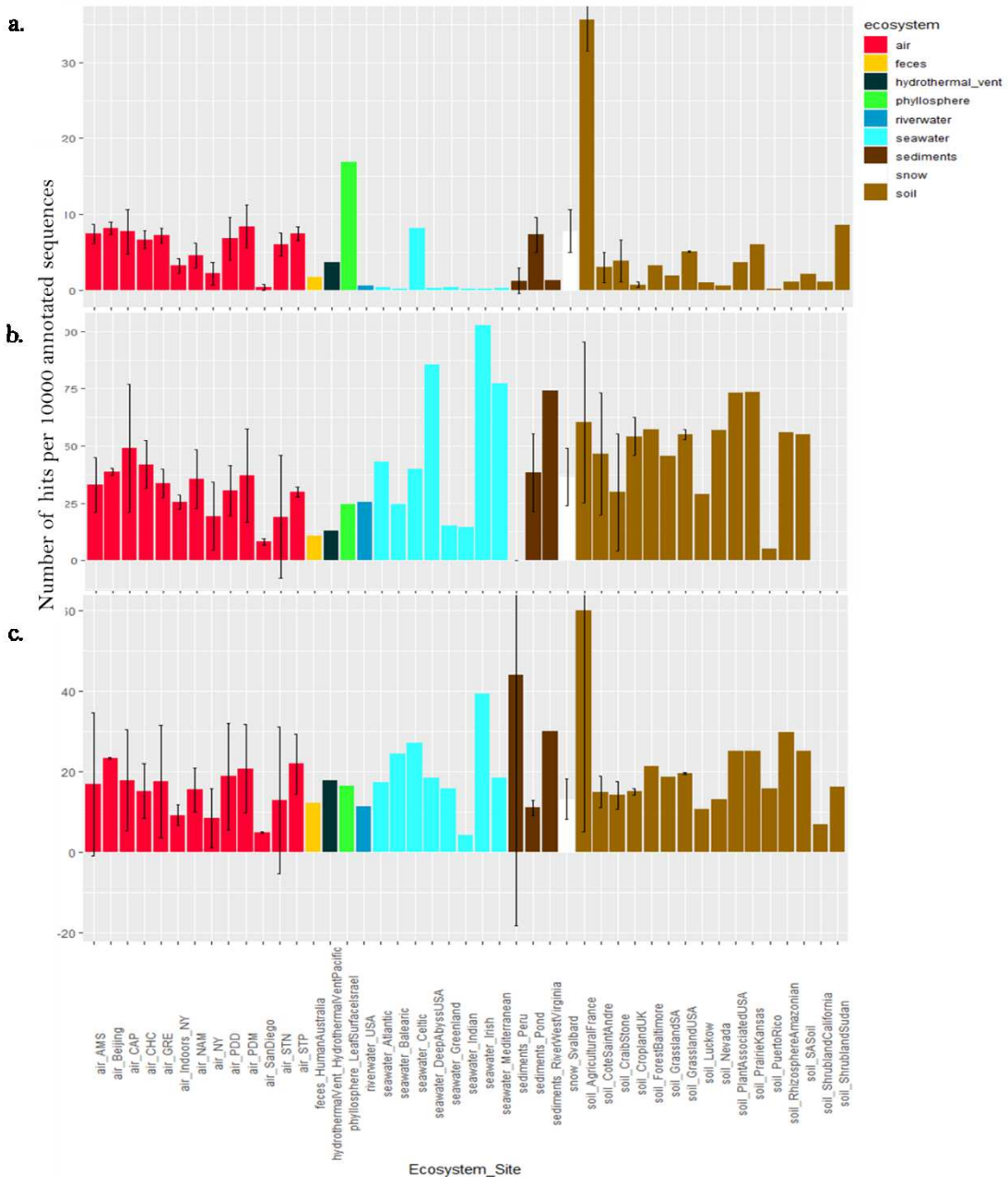


Figure S4: Average number of hits of hydrogen peroxide catabolic process related functions per 10000 annotated sequences from (a) all sequences, (b) fungi-associated sequences and (c) bacteria-associated sequences per site. Colors indicate the ecosystems in which the sites belong to. For the sites including several metagenomes, the standard deviation was added.

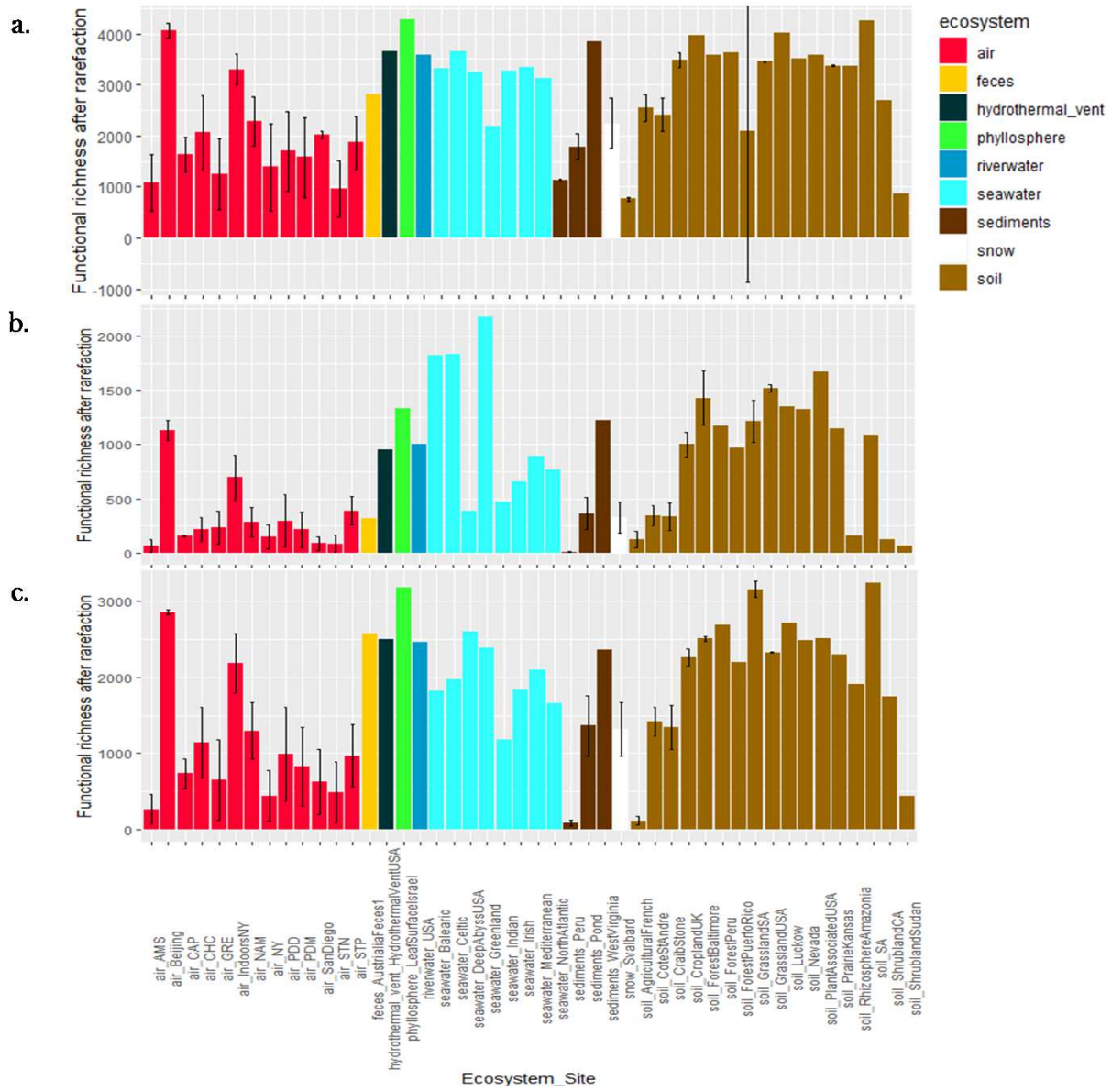


Figure S5: Functional richness after rarefaction from (a) all sequences, (b) fungi-associated sequences and (c) bacteria-associated sequences per site. Colors indicate the ecosystems in which the sites belong to. For the sites including several metagenomes, the standard deviation was added.

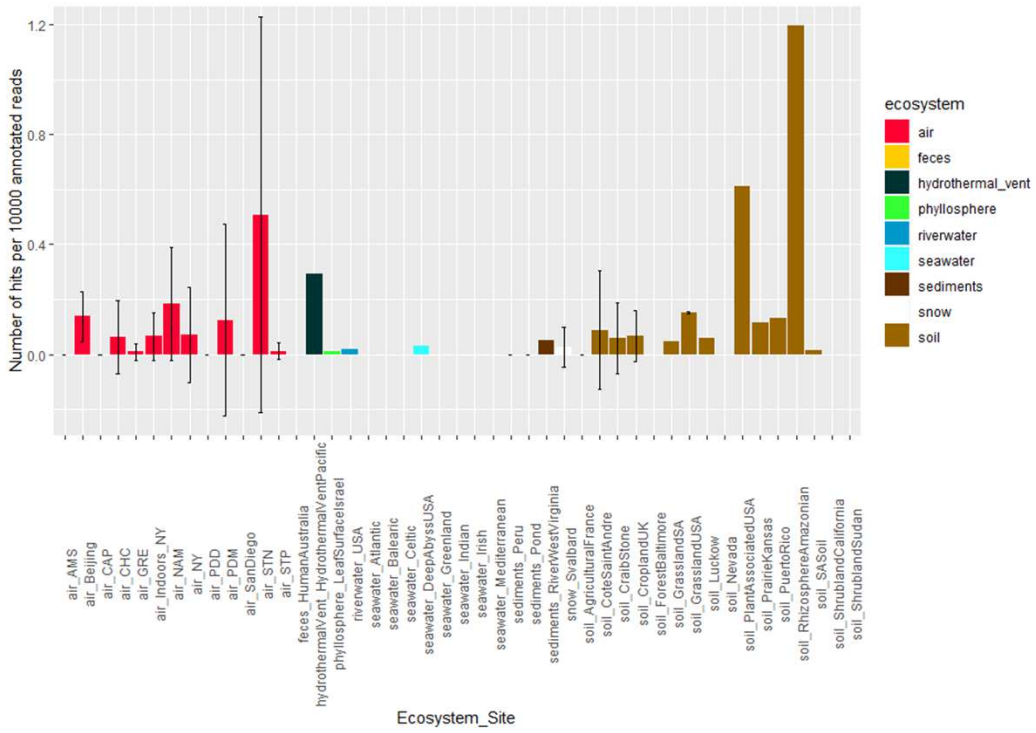


Figure S6: Average number of hits of methane monooxygenase process related functions per 10000 annotated sequences from all sequences per site. Colors indicate the ecosystems in which the sites belong to. For the sites including several metagenomes, the standard deviation was added.

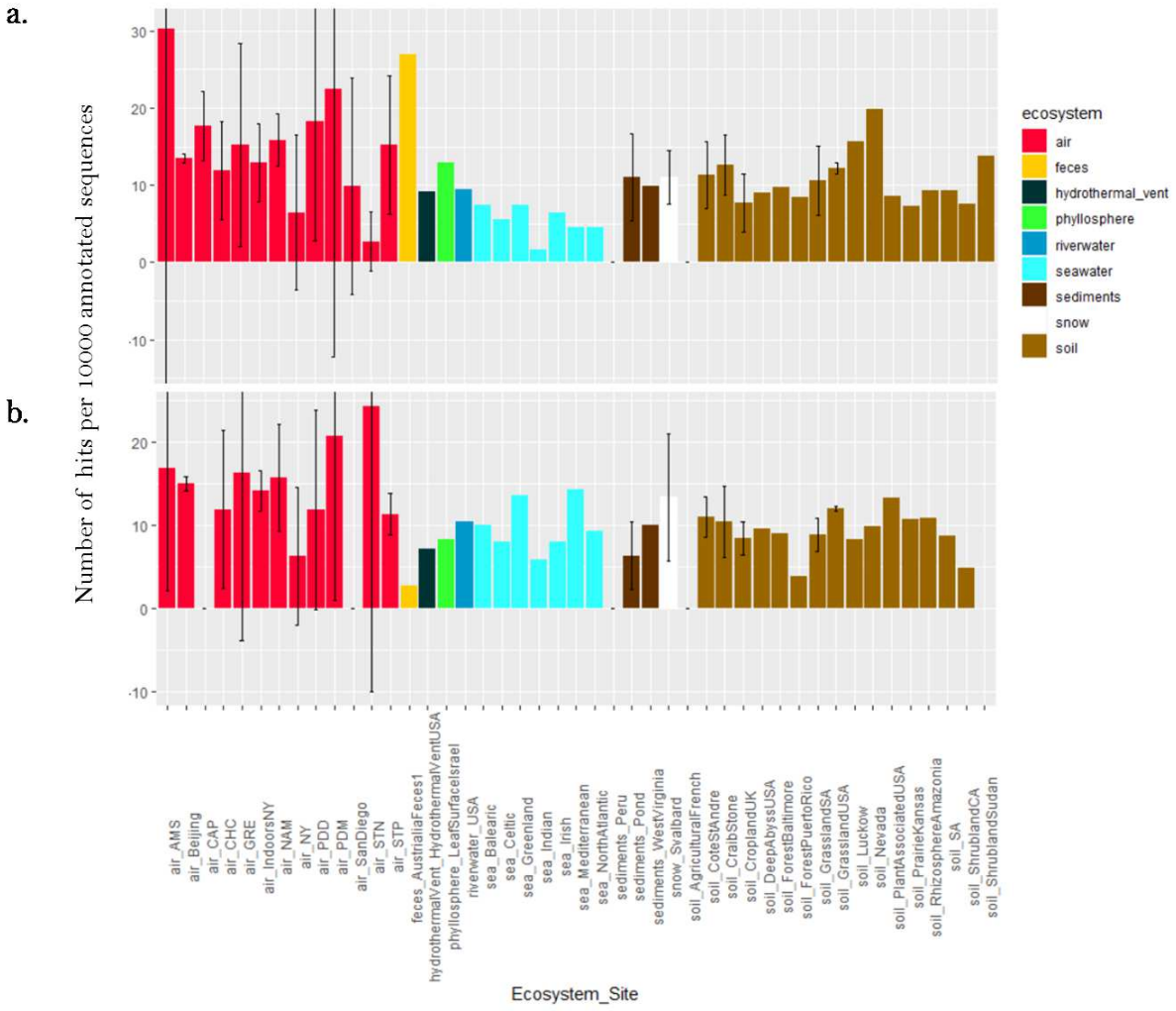


Figure S7: Average number of hits of lipoate synthase (a) and chromosome plasmid partitioning protein ParA (b) per 10000 annotated sequences from all sequences per site. Colors indicate the ecosystems in which the sites belong to. For the sites including several metagenomes, the standard deviation was added.

Chapter 6: Conclusion and perspectives

The troposphere represents an immense volume in which microorganisms are found in the gas phase (as free cells), solid phase (as attached to aerosols) and liquid phase (in cloud, rain, snow or fog water). These phases constitute potentially different biological niches associated with different airborne microbial communities that are, in general, poorly investigated in aeromicrobiology (discussion in **Chapter 1**). This PhD focused on free- and particulate-matter-associated microorganisms found in the dry phase of the troposphere (*i.e.* gas and solid phases) and the dry phase of the lower part of the troposphere (*i.e.* the planetary boundary layer). The planetary boundary layer interacts with humans, crops and diverse ecosystems like soils and oceans and, thus, understanding its microbial composition is of utmost importance. While planetary boundary layer microorganisms (from the dry phase) could vary significantly through space and time, only a few small spatial-scale studies have addressed local to regional scale effects. The most recent studies (using molecular biology tools) investigated the structure of microbial communities and sometimes the possible controversial controlling factors of airborne microbial communities including the local sources, local meteorology and long-range transport (*i.e.* transport over distances of hundreds of kilometers). The different experimental set-ups and sampling strategies that were used hampered a full understanding of atmospheric microbial composition distribution (discussion in **Chapter 1**).

We investigated both the taxonomic and functional distributions of troposphere microbial communities (of the dry phase) using a global-scale and molecular approach (*i.e.* qPCR, high throughput sequencing technologies). Our main objective was to answer the following questions: "*How is airborne microbial community composition driven?*" and "*Can we predict airborne microbial composition?*" We also determined the relative contribution of the different structuring factors (already cited above) on microbial communities of the dry troposphere. Finally, the physical selection processes and microbial adaptation that might occur during aerosolization and aerial transport were also evaluated.

We collaborated with research teams at atmospheric monitoring stations around the world and collected 150 samples and 45 controls at nine sites over a time period of 2 months to up to a year (for Puy-de-Dôme in France). For some of the locations, such as at the high altitude

station at Chacaltaya in Bolivia (5400 m above sea level) and the very remote site of Amsterdam Island in the south Indian Ocean, airborne microorganisms have never been explored in such a detailed manner. We designed a sampling strategy in order to access both particle (PM10) chemistry and microbiology while guaranteeing the collection of sufficient biomass with good efficiency and limited contamination (**Chapter 2**). A DNA extraction technique specific for quartz fiber filters was developed and implemented. It led to sufficient DNA yield to perform the subsequent molecular biology analyses, such as amplicon and metagenomic sequencing (**Chapter 2**).

Our work highlighted the significant contribution of local sources (*i.e.* surrounding landscapes) in the taxonomic and functional composition of microbial communities of the dry troposphere (**Chapter 3, 4** and **5**). The temporal evolution of the microbial composition was strongly associated to the surface conditions of the surrounding landscapes (changes in surface conditions encountered throughout the seasons, for example) (**Chapter 4**). Our results also showed that aerosolization could select for most resistant microorganisms in the air (**Chapter 5**). While mainly aerosolized by the surrounding landscapes, planetary boundary layer microbial communities are not the sum of the microbial communities of the different surrounding sources (soil, plants etc.), but selected microorganisms that are more resistant to atmospheric conditions (*i.e.* desiccation, UV radiation etc.), more aerosolized (due to membrane properties for example) and/or protected in atmospheric particulate matter. Microbial cell inputs from the different local sources vary through time and are modulated by local meteorological conditions, such as wind direction and speed as they would affect aerosolization and deposition of microbial cells (**Chapter 3**). Contribution of long-range transport and, thus, microbial cell inputs from distant sources appeared minor compared to inputs from local sources. Yet, our sampling strategy (one-week sampling) was adapted to identify large inputs from distant sources. We suggested that inputs resulting from long-range transport, whether intermittent or continuous, were not sufficiently large to be detected in our samples. **Figure 1** presents the role of the different factors in controlling microbial communities of the planetary boundary layer.

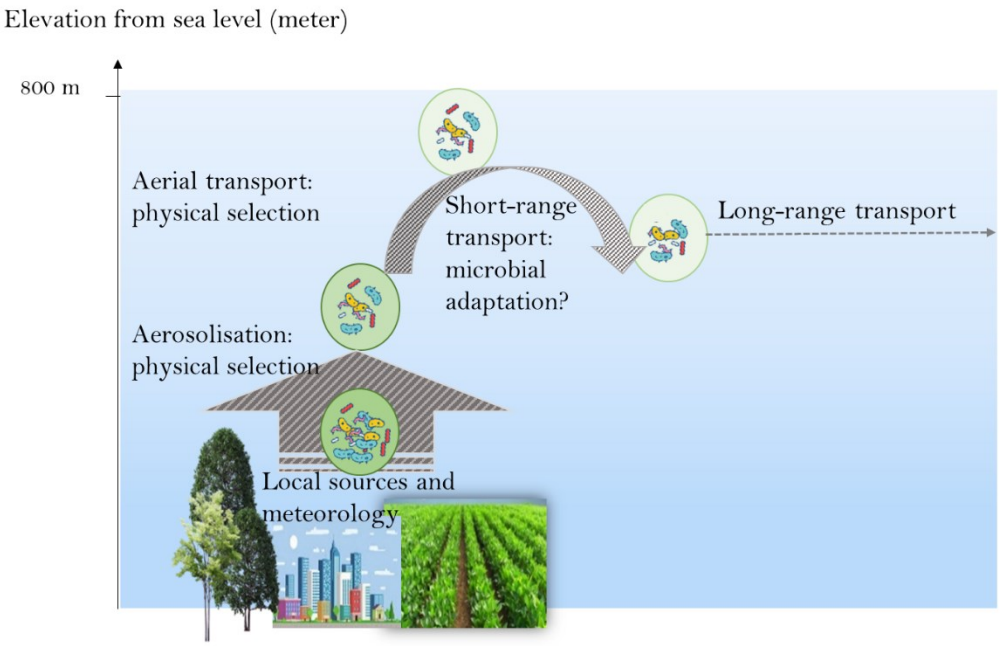


Figure 1: Overview of the role of different factors in controlling microbial communities of the planetary boundary layer.

As stated before, our studies led to significant results in the field of aeromicrobiology. They also highlight remaining gaps of knowledge and open new avenues for future projects. Our first geographical investigation on boundary layer microbial community composition (**Chapter 3**) showed that local sources and meteorology were the major contributors influencing the abundance and taxonomic composition of airborne microbial communities. While our investigation highlighted the major role of the surrounding sources in the observed taxonomic and functional composition of airborne microbial communities, the exact contribution of the local and the distant sources remains unknown and might not be technically possible to evaluate at the moment. Long-range transport might be important (both quantitatively and qualitatively with the dissemination of pathogens over long distances) during extreme meteorological events such as large dust storms and cyclones. A different behavior between PM10 chemistry and airborne microbial communities was also observed. During their transport in the atmosphere, the dynamic behavior of microorganisms might differ from that of the PM10 chemistry, as their sources might be different and/or their fate may diverge due to differences in their transformation and deposition while transported in the atmosphere. Specific atmospheric chemicals, such as polyols (**Chapter 3** and **4**), that correlate significantly

with airborne microorganisms are components of microorganisms. Atmospheric chemists might have overlooked the contribution of microorganisms to the presence and concentration of specific atmospheric chemicals. A recent example is the atmospheric sodium salt over the central Amazon basin that had been previously reported as salt coming from marine sources through air circulation, but was due to the presence of sodium-containing airborne fungal spores emitted locally[1]. **Chapter 4** dealt with an annual time-series of boundary layer microbial communities and showed that the seasonal changes in the composition of airborne microorganisms were linked to changes in the surface conditions of the surrounding landscapes. At our continental site at Puy-de-Dôme (France), the structure of airborne microbial communities was correlated to the seasonal changes in the relative abundance of crop and vegetation-associated microorganisms (higher relative abundance in spring/summer) and the relative abundance of soil and dead material-associated microorganisms (higher relative abundance in fall/winter). Agriculture in Puy-de-Dôme area represented 241 265 hectares of agricultural land and 7377 farms (data of 2016[2]). The potential impact of airborne microorganisms in disseminating diseases between crops in Puy-de-Dôme should be evaluated in order to prevent and/or control disease transmissions. Our global-scale functional study of airborne microbial communities using a MiSeq Illumina sequencing approach (**Chapter 5**) showed no apparent specific signature of planetary boundary layer microbial communities in terms of their functional potential. Air samples tended to group with their underlying ecosystems. Air metagenomes were characterized by a relatively high percentage of fungi-associated sequences compared to the source ecosystems (soil, surface seawater etc.). This higher percentage could be explained by a higher survival rate of fungi (compared to bacteria) after aerosolization and/or aerial transport due to their innate resistance to physical atmospheric conditions. Airborne fungi drove the higher proportions of stress-related functions observed in air metagenomes compared to the other ecosystems (*e.g.*, soil and oceans). Airborne bacterial communities did not necessarily show higher stress or resistance-related functions or functions linked to potential atmospheric chemical metabolisms (such as CH_4 and H_2O_2 metabolism).

The lack of evidence for physical selection and/or microbial adaptation of airborne bacterial communities might be due to the large local input of new microbial cells that continuously enter the planetary boundary layer. Similar global investigations should be done in the free

troposphere using balloon as this layer is not as influenced by the planetary surface as the planetary boundary layer and, thus, might harbor microbial cells that have been suspended longer. Aerosolization might have an important role in controlling the composition of airborne microbial communities by selecting which microorganisms enter the planetary boundary layer. A subsequent selective process might occur during aerial transport, but microbial cells embedded in particulate matter might survive atmospheric conditions even if they are not specifically resistant. Atmospheric chemicals might also affect the structure of airborne microbial communities by selecting adapted microorganisms, especially if microorganisms are metabolically active and grow in air. Thus, airborne microbial community composition remains not fully predictable at the moment, although it is strongly linked to local meteorology and the microbial composition of the underlying ecosystems. A better understanding of the atmospheric physical selection (how atmospheric chemicals and physics influence airborne microbial survival and adaptation depending on microbial taxon) is needed. Complex microbial communities (and not individual strains) might be investigated under controlled environmental conditions in atmospheric chambers similar to the atmospheric conditions. These investigations might help detect if microbial communities (especially bacteria) undergo a selective or adaptive process while airborne. The potential selective processes occurring during aerosolization should also be further explored in controlled settings. Questions regarding which microbial taxa tend to be more aerosolized and more resistant to aerosolization in complex microbial communities under different meteorological conditions should be addressed.

Despite the knowledge gained on airborne microbiota in this thesis, the field of aeromicrobiology is still emerging with major questions remaining such as "*What do microorganisms do once in the air?*". "*Are they metabolically active?" "If so, do they have an impact on atmospheric chemistry and more largely on biogeochemical cycles?" "Do they grow?"* are all crucial related questions that should be further addressed. The contribution of airborne microbial activity in transforming chemical species might be greater than currently expected, and might explain inconsistencies found in some atmospheric chemical cycles (for example the cycle of secondary organic aerosols[3,4]). Although the activity of microorganisms of airborne origin has been shown on culture medium, airborne microbial activity should be evaluated *in situ*, *i.e.* in the field or under more controlled settings such as in atmospheric

chambers. Field meta-omic investigations, especially metatranscriptomics and metaproteomics are currently very limited approaches in aeromicrobiology due to the low biomass represented by airborne microorganisms and the sensitivity of sequencing technologies. To our knowledge, only one metatranscriptomic investigation has been done on cloud water (which required RNA amplification before sequencing[5]) and one metaproteomic investigation (which is a methodology paper[6]) has been reported. Yet, these approaches might be useful for evaluating airborne microbial activity and should be further developed.

References

1. China, S. *et al.* Fungal spores as a source of sodium salt particles in the Amazon basin. *Nat Commun* **9**, 1–9 (2018).
2. Auvergne-Rhône-Alpes, C. d'agriculture. Agriculture du Puy-de-Dôme. (2019). Available at: https://aura.chambres-agriculture.fr/notre-agriculture/agriculture-du-puy-de-dome/. (Accessed: 11th September 2019)
3. Yang, W., Li, J., Wang, M., Sun, Y. & Wang, Z. A Case Study of Investigating Secondary Organic Aerosol Formation Pathways in Beijing using an Observation-based SOA Box Model. *Aerosol and Air Quality Research* **18**, (2018).
4. Khan, M. a. H. *et al.* A modeling study of secondary organic aerosol formation from sesquiterpenes using the STOCHEM global chemistry and transport model. *Journal of Geophysical Research: Atmospheres* **122**, 4426–4439 (2017).
5. Amato, P. *et al.* Metatranscriptomic exploration of microbial functioning in clouds. *Sci Rep* **9**, 1–12 (2019).
6. Liu, F. *et al.* Metaproteomic analysis of atmospheric aerosol samples. *Analytical and Bioanalytical Chemistry* **408**, (2016).

Side publications

Sequencing depth rather than DNA extraction determines soil bacterial richness discovery

Concepcion Sanchez-Cid[1,2], Romie Tignat-Perrier[1,3], Laure Franqueville[1], Laurence Delaurière[2], Trista Schagat[4], Timothy M. Vogel[1]

[1]Environmental Microbial Genomics, Laboratoire Ampère, Ecole Centrale de Lyon, Université de Lyon, Ecully, France

[2]Promega France, Charbonières-les-Bains, France

[3]Institut des Géosciences de l'Environnement, Université Grenoble Alpes/CNRS/Institut de Recherche pour le Développement, G-INP, Grenoble, France

[4]Promega Corporation, Madison, Wisconsin, United States

Abstract

Soil ecosystems harbour the highest biodiversity on our planet. Although Next Generation Sequencing techniques have increased our access to the soil microbiome, each step of soil metagenomics presents inherent biases that prevent the accurate definition of the soil microbiome and its ecosystem function. Biases related to DNA extraction have been particularly well documented with no one method being able to avoid them. In this study, we compared the phylogenetic and functional richness detected by five DNA extraction methods from two soil samples. The V3-V4 hypervariable region of the 16S rRNA gene was sequenced to determine the taxonomical DNA richness measured by each method at the genus level. The functional richness was evaluated by metagenomics sequencing. Sequencing depth had a greater influence on bacterial richness discovery at both taxonomical and functional levels than the DNA extraction methods. Furthermore, at an equal sequencing depth, the differences observed between methods in both soils were more likely a product of random subsampling than that of the DNA extraction itself. Therefore, an optimization of each step of soil

metagenomics workflow is needed in order to sequence samples at an equal depth and to be able to perform more accurate metagenomics comparisons.

Introduction

The soil ecosystem arguably harbours the highest diversity of microorganisms of any ecosystem[1]. Unravelling the composition and function of the soil microbiome is critical to better understanding the role of these microbial communities in soil function and ecosystem services. The use of Next Generation Sequencing (NGS) techniques has increased our access to the microbial communities present in soil, especially to the large proportion of uncultured microorganisms[2,3]. Nevertheless, each methodological step from soil sampling to sequence annotation presents inherent biases that limit the depth and reliability of soil microbiome analyses[4–7]. Of all these biases, the ones associated with DNA extraction have been particularly highlighted for their effects[8,9]. DNA can be adsorbed by soil compounds such as clay[10,11], which when combined with the presence of lysis-recalcitrant bacteria[12], reduces DNA extraction efficiency. Moreover, organic matter and humic acids, which are known to potentially inhibit enzymatic reactions[13], are often co-extracted with DNA. Many studies have compared DNA extraction methods and documented the biases imposed by lysis procedures[14,15]. This has provoked the proposal of different methods for DNA extraction and purification from soil over the past few decades[14,16,17] in an attempt to obtain an unbiased picture of soil microbiome biodiversity[9]. Nevertheless, no method has been shown to overcome all the biases described above and the debate on the choice of a DNA extraction method is still ongoing.

The criteria used to define the performance of a DNA extraction method vary between studies and range from DNA yield to phylogenetic diversity. However, higher DNA yields, purity, and integrity do not always imply an improvement in bacterial diversity discovery[4]. Analyses of the relative abundance of taxonomic groups have a limited potential for selecting DNA extraction methods, since the biases associated with soil metagenomics prevent us from determining the actual distribution of soil microbial populations within a community. Furthermore, DNA extraction methods may modify the relative abundance of detected communities without affecting bacterial richness discovery. The goal of this paper was not to resolve the debate concerning the use of relative abundance versus richness. We believe that bacterial richness measurements provide a more objective comparison of the performance of DNA extraction methods, since biodiversity calculations with a limited set of sequences are strongly biased by

evenness, which depends on the number of sequences. In addition, the actual relative abundances of the soil microbiome remain unknown.

In this study, we compared the phylogenetic and functional richness detected by two novel semi-automated methods for DNA extraction and purification to that measured by two commercial kits and the phenol/chloroform method as described by Griffiths *et al* [18]. DNA was extracted from two soil samples and the V3-V4 hypervariable region of the 16S rRNA gene was sequenced to determine the Operational Taxonomic Unit (OTU). DNA richness measured by each method was defined at the genus level for each sequence using the Ribosome Data Project (RDP) database and the RDP Bayesian classifier[19]. Finally, metagenomics sequences were annotated using MEGAN6[20] and the SEED hierarchical subsystems, and the third level of the SEED classification was selected for functional richness discovery analysis.

Results

DNA quantification and quantitative PCR (qPCR) assays

DNA samples were quantified and amplified before library preparation to assess the concentrations and number of copies of the 16S rRNA gene obtained by each method (Table 1). Both DNA concentrations and 16S rRNA gene copies obtained using the ZymoBIOMICS™ DNA Kit were lower than those measured by the other methods for both soils. The highest DNA concentrations were obtained from both soils using the Phenol/Chloroform method and the DNeasy® PowerSoil® Kit. However, when performing DNA amplification, similar numbers of 16S rRNA gene copies were obtained after DNA extraction using the Maxwell 1 and 2 methods, which are variants of the Maxwell® PureFood GMO and Authentication Kit (Promega). Therefore, PCR inhibition was greater in samples with higher DNA concentrations (*i.e*. Phenol/chloroform and Powersoil). In order to reduce amplification biases and normalize the amount of starting material for library preparation, all samples were diluted to 2.5 ng/µl before 16S rRNA gene PCR amplification and sequencing.

Table 1: DNA concentrations and number of copies of the 16S rRNA gene per µl of qPCR reaction obtained from 250 mg of soil from the Scottish Agricultural College (Craibstone, Scotland) and La Côte de Saint André (France) using different DNA extraction and purification strategies. DNA concentrations were assessed using the Qubit™ dsDNA HS Assay Kit. 2 µl of DNA were amplified by qPCR using the GoTaq® qPCR Master Mix and 341F/534R primers. Obtained average and standard deviation values are shown in the table. Maxwell 1 and 2: variants of the Maxwell® PureFood GMO and Authentication Kit (Promega). N=3.

Soil	**Method**	**DNA concentration (ng/µl)**	**16S rRNA gene copies/µl**
Scottish Agricultural College soil	Maxwell 1	9.05 ± 1.35	153,955 ± 52,412
	Maxwell 2	6.01 ± 1.50	64,279 ± 9,133
	Phenol/Chloroform method	23.40 ± 1.90	189,213 ± 15,036
	DNeasy® PowerSoil® Kit	17 ± 3.12	165,311 ± 65,886
	ZymoBIOMICS™ DNA Mini Kit	2.79 ± 0.64	20,056 ± 27,979
La Côte de Saint André soil	Maxwell 1	6.43 ± 0.97	160,608 ± 14,701
	Maxwell 2	5.06 ± 0.54	170,391 ± 16,323
	Phenol/Chloroform method	15.20 ± 2.76	211,949 ± 45,407
	DNeasy® PowerSoil® Kit	22.90 ± 1.15	88,904 ± 25,638
	ZymoBIOMICS™ DNA Mini Kit	0.12 ± 0.06	2,557 ± 1,376

Sequencing depth effect on bacterial richness discovery

In order to assess the relative contribution of sequencing depth and DNA extraction methods on taxonomical and functional bacterial discovery, the richness measured in each sample was determined and plotted as a function of sequencing depth (**Fig. 1**). Less than 10000 reads were obtained from ZymoBIOMICS™ DNA Mini Kit (Zymo Research) triplicate pools after

metagenomics sequencing and annotation. Therefore, they were excluded from functional richness analysis. Sequencing depth showed a higher influence on bacterial richness discovery at both taxonomical and functional levels than DNA extraction methods: measured richness increased proportionally with sequencing depth regardless of the method used for DNA extraction. Furthermore, DNA extraction triplicates that had been sequenced at different depths had access to a different proportion of bacterial DNA richness (**Fig. 1**). Given these findings, we normalized samples by size (sequence or read numbers) before comparing the richness measured by different DNA extraction methods in order to limit the effect of sequencing depth. After removing singletons and pooling method triplicates, subsamples of each method were randomly rarefied at 58645 sequences annotated as genera for the Scottish Agricultural College soil, 36311 sequences annotated as genera for La Côte de Saint André soil, 32693 sequences annotated as functions for the Scottish Agricultural College soil and 159748 sequences annotated as functions for La Côte de Saint André soil, as this represented values that all samples could meet, even the samples with the lowest number of reads.

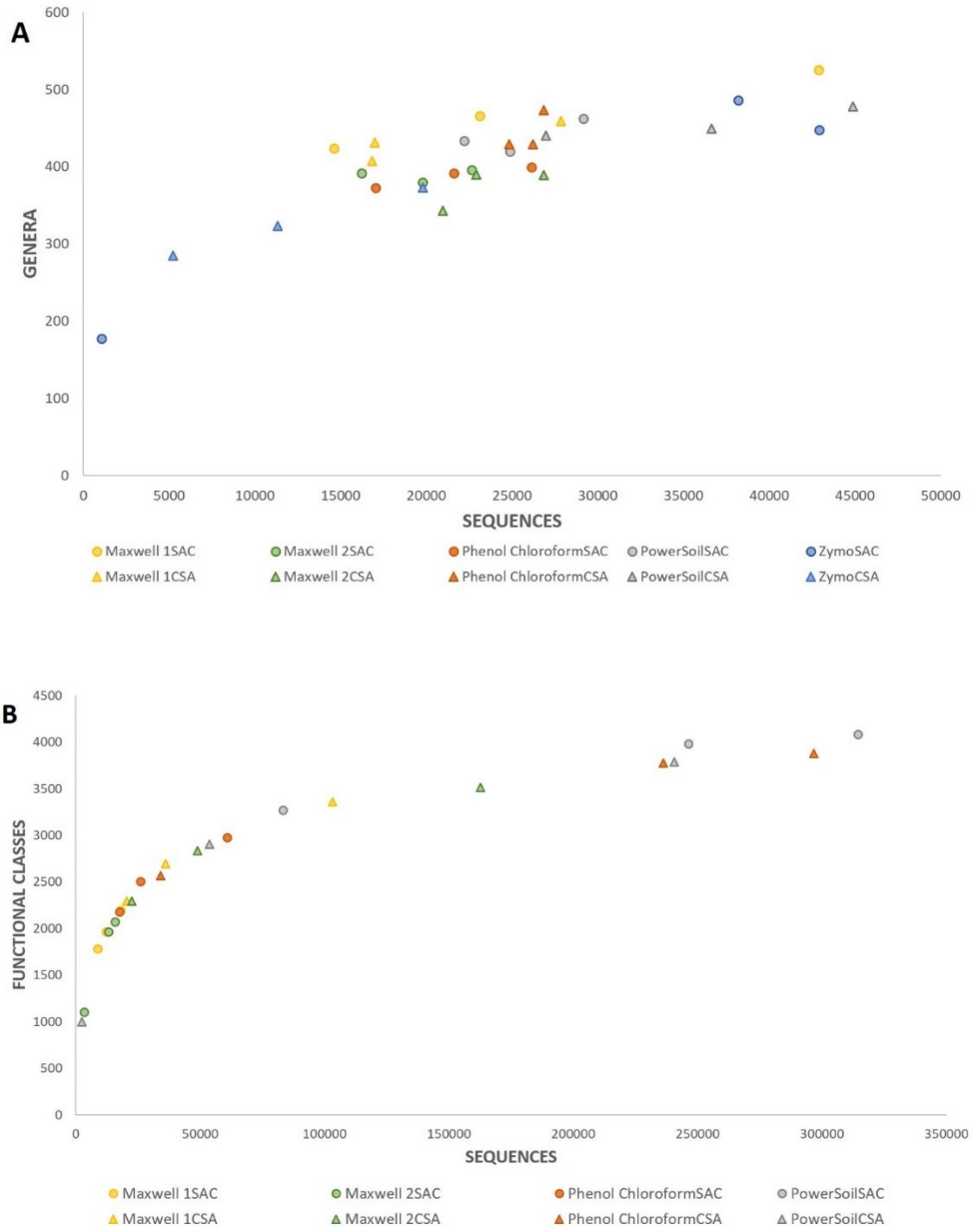


Figure 1: Effect of sequencing depth on taxonomic (A) and functional (B) richness discovery. Maxwell 1 and 2: variants of the Maxwell® PureFood GMO and Authentication Kit (Promega). POWERSOIL: DNeasy® PowerSoil® Kit (QIAGEN); ZYMO: ZymoBIOMICS™ DNA Mini Kit (Zymo Research). The taxonomical and functional richness measured in each sample was determined and plotted as a function of sequencing depth. Circles represent samples obtained from the Scottish Agricultural College soil, whereas samples obtained from La Côte de Saint André soil are represented by triangles. Colours were assigned to different DNA extraction method used (blue and orange for Promega's Maxwell 1 and 2 methods, respectively, grey for the Phenol/Chloroform method, yellow for the DNeasy® PowerSoil® Kit (QIAGEN) and green for the ZymoBIOMICS™ DNA Mini Kit (Zymo Research). For each method/soil pair, triplicates were performed and plotted in the graph. Less than 10000 reads were obtained from ZymoBIOMICS™ DNA Mini Kit (Zymo Research) triplicate pools after metagenomics sequencing and annotation. Therefore, they were excluded from functional richness analysis.

DNA extraction methods comparison at an equal sequencing depth

At an equal sequencing depth, genera and functional classes shared by all methods as well as those detected by only one method for each of the two soils were represented using Venn Diagrams (**Fig. 2**). The proportion of genera detected by a single method represented 8.29% of the Scottish Agricultural College soil pool and 12.84% of the pool from La Côte de Saint André soil. Regarding functional classes, the ones measured by a single method represented 20.22% of the Scottish Agricultural College pool and 12.11% of La Côte de Saint André pool. The average relative abundance of the 25 most abundant genera and functional classes and their relative abundances in each method subsample were inferred from the rarefied pool (**Tables S1a, b, c** and **d** in Supplementary Material). Regarding genera, most values fit into the average plus/minus one standard deviation for the Maxwell 1 and phenol/chloroform methods and the DNeasy® PowerSoil® Kit. The Maxwell 2 method globally showed higher relative abundances than the average plus one standard deviation in both soils, whereas the relative abundances of the most abundant genera were often lower than the average minus one standard deviation in the ZymoBIOMICS™ DNA Mini Kit rarefied pools. The most abundant functional classes were more likely to be between the average plus/minus one standard deviation than the genus distribution. Most of the function values outside one standard deviation from the mean belonged to the Maxwell 2 method for both soils.

Complementary analysis had shown that only a small proportion of annotated genera and functional classes are relatively abundant in the analysed ecosystems. Most of the annotated genera and functional classes were present at low abundance, with less than 30 associated sequences each (see **Fig. S1** and **S2** in Supplementary Material). Therefore, random subsampling of low abundant genera and functional classes could increase the rate of false positives by detecting artificially unique genera and functional classes. To determine whether the genera and functional classes classified as unique were a product of random rarefaction, their distribution between the different methods before rarefaction was determined. A genus or functional class was validated as unique when it was detected by only one method both before and after rarefaction. Otherwise, it was classified as a false assignation due to random rarefaction (**Table 2a, b**). The percentage of false assignation of unique genera or functional classes and its correlation to the number of sequences before rarefaction were assessed. Regarding genera, the percentage of false assignment ranged between 33% and 87% and it

showed no correlation to the sample size before rarefaction, with R^2 values of 0.31 and 0.003 in the Scottish Agricultural College soil and La Côte de Saint André soil, respectively (**Table 2a, b**). On the other hand, there was a negative correlation between the percentage of false assignment of functional classes and the sample size before rarefaction, with R^2 values of 0.99 and 0.96 in the Scottish Agricultural College soil and La Côte de Saint André soil, respectively (**Table 2a, b**).

Finally, the number of sequences associated with each validated unique genus and functional class before sample size normalization was inferred. Most genera detected by only one method had less than six annotated sequences before rarefaction (Tables S2a and b in Supplementary Material). The only exception was *Staphylococcus,* which was detected by the ZymoBIOMICS™ DNA Mini Kit and had 95 and 130 annotated sequences before rarefaction in the Scottish Agricultural College soil and La Côte de Saint André soil, respectively. Regarding functions, the same tendency was generally observed with less than 6 sequences annotated per unique function before size normalization. However, in the Scottish Agricultural College soil, more than 6 sequences were associated to half of the functional classes detected only by the DNeasy® PowerSoil® Kit, the largest pool of the cohort before rarefaction (Table S3 in Supplementary Material). These findings suggest that size normalization was not enough to account for the sequencing depth effect on bacterial richness discovery.

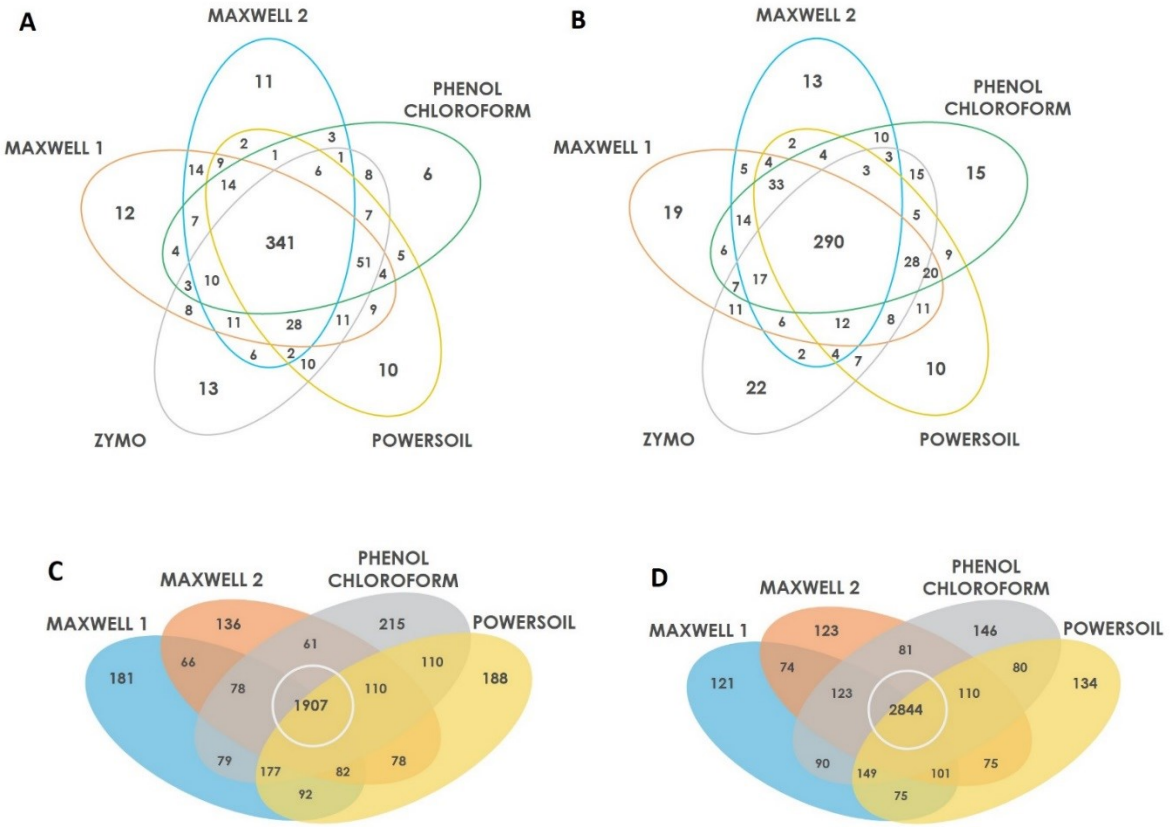


Figure 2: Venn Diagrams representing shared and unique (A) genera from DNA extracted from the Scottish Agricultural College soil; (B) genera from DNA extracted from La Côte de Saint André soil; (C) functional classes from DNA extracted from the Scottish Agricultural College soil; (D) functional classes from DNA extracted from La Côte de Saint André soil. Maxwell 1 and 2: variants of the Maxwell® PureFood GMO and Authentication Kit (Promega). POWERSOIL: DNeasy® PowerSoil® Kit (QIAGEN); ZYMO: ZymoBIOMICS™ DNA Mini Kit (Zymo Research). Less than 10000 reads were obtained from ZymoBIOMICS™ DNA Mini Kit (Zymo Research) triplicate pools after metagenomics sequencing and annotation. Therefore, they were excluded from functional richness analysis. Sequences from triplicates for each method/soil couple were pooled and singletons removed. Then, a subsample of each method pool was randomly rarefied using the vegan package in R at the lowest size of the cohort (58645 sequences annotated as genera for the Scottish Agricultural College soil; 36311 sequences annotated as genera for La Côte de Saint André soil; 32693 sequences annotated as functions for the Scottish Agricultural College soil; 159748 sequences annotated as functions for La Côte de Saint André soil). Venn Diagrams were obtained using the R package VennDiagram. N=3.

Table 2a: Percentage of false assignation of unique genera due to sample size normalization and correlation to sample size before rarefaction. Maxwell 1 and 2: variants of the Maxwell® PureFood GMO and Authentication Kit (Promega). PowerSoil: DNeasy® PowerSoil® Kit (QIAGEN); Zymo: ZymoBIOMICS™ DNA Mini Kit (Zymo Research). Detected unique genera: genera identified in a single method after sample rarefaction at the lowest size of the cohort (58645 sequences for the Scottish Agricultural College soil and 36311 sequences for La Côte de Saint André soil). Validated unique genera: genera identified in a single method both before and after rarefaction. False assignation of unique genera: genera detected as unique after rarefaction but detected by more than one method before rarefaction. N=3.

Soil	Method	Sequences before rarefaction	Detected unique genera	Validated unique genera	% False assignation of unique genera
SCOTTISH AGRICULTURAL COLLEGE	Maxwell 1	80628	12	8	33
	Maxwell 2	58645	11	4	64
	Phenol Chloroform	64805	6	3	50
	PowerSoil	76270	10	3	70
	Zymo	82187	13	8	38
	TOTAL	**362535**	**52**	**26**	**50**
	R^2 between sequences before rarefaction and % false assignation of unique genera = 0.31				
LA COTE DE SAINT ANDRE	Maxwell 1	61656	19	3	84
	Maxwell 2	70676	13	3	77
	Phenol Chloroform	77866	15	2	87
	PowerSoil	108447	10	3	70
	Zymo	36311	22	7	68
	TOTAL	**354956**	**79**	**18**	**77**
	R^2 between sequences before rarefaction and % false assignation of unique genera = 0.003				

Table 2b: Percentage of false assignation of unique functions due to sample size normalization and correlation to sample size before rarefaction. Maxwell 1 and 2: variants of the Maxwell® PureFood GMO and Authentication Kit (Promega). PowerSoil: DNeasy® PowerSoil® Kit (QIAGEN). Less than 10000 reads were obtained from ZymoBIOMICS™ DNA Mini Kit (Zymo Research) triplicate pools after metagenomics sequencing and annotation. Therefore, they were excluded from functional richness analysis. Detected unique functional classes: functional classes identified in a single method after sample rarefaction at the lowest size of the cohort (32693 sequences for the Scottish Agricultural College soil and 159748 sequences for La Côte de Saint André soil). Validated unique functional classes: functional classes identified in a single method both before and after rarefaction. False assignation of unique functional classes: functional classes detected as unique after rarefaction but detected by more than a method before rarefaction. N=3.

Soil	Method	Sequences before rarefaction	Detected unique functional classes	Validated unique functional classes	% False assignation of unique functional classes
SCOTTISH AGRICULTURAL COLLEGE	Maxwell 1	39204	181	11	94
	Maxwell 2	32693	136	12	91
	Phenol Chloroform	104693	215	21	90
	PowerSoil	644248	188	118	37
	TOTAL	**820838**	**720**	**162**	**78**
	R^2 between sequences before rarefaction and % false assignation of unique functions = 0.99				
LA COTE DE SAINT ANDRE	Maxwell 1	159748	121	39	68
	Maxwell 2	234007	123	54	56
	Phenol Chloroform	566982	146	95	35
	PowerSoil	296570	134	57	57
	TOTAL	**1257307**	**524**	**245**	**53**
	R^2 between sequences before rarefaction and % false assignation of unique functions = 0.96				

Discussion

In this study, we compared the bacterial richness detected by two new alternatives for semi-automated DNA extraction to methods commonly used in soil metagenomics. However, the contribution of different DNA extraction methods to bacterial richness discovery was relatively low compared to that of sequencing depth (**Fig. 1**); greater sequencing depth increased the discovery of OTUs and functional classes regardless of the method used for DNA extraction. In other words, at a cursory level, all methods detected a similar taxonomical and functional richness at equal sequencing depths. The main implication of the observed unequal sampling depths and their effect on richness discovery was the need for normalizing sample sizes in order to be able to perform comparisons between DNA extraction methods. One of the most popular approaches for sample normalization is the use of relative abundances[6], which have often been selected as a criteria to compare the performance of DNA extraction methods. However, as soil DNA complete diversity has not been extracted and sequenced yet[21], the actual relative abundances of soil microbiome cannot be known and the obtained profiles cannot be validated. Furthermore, the levels of classification chosen for these comparisons are often too high (*e.g*., phylum level) to be informative. More importantly, relative abundances are potentially influenced by sequencing depth, since the probability of detecting rare OTUs or functions increases with the number of sequences, and therefore, changes the relative proportions of high (and not unique) abundant versus low abundant OTUs or functions. Finally, differences in sample lysis and DNA purification may affect the relative abundances of the detected communities without modifying measured richness. In other words, the use of different DNA extraction methods could detect different proportions of the same communities rather than of different OTUs or functions. Altogether, the informative potential of relative abundance measurements appears to be considerably limited and the performance of DNA extraction methods should be compared using absolute values. Thus, after singletons removal and triplicates pooling, we normalized sample size by randomly rarefying every method pool at that for the lowest measured sequencing depth.

Only a few of the genera and functional classes present in the non-rarefied pools were abundant (see **Fig. S1** and **S2** in Supplementary Material) as both genera and function pools were mainly composed by low-abundant elements. This pattern of taxonomical distribution has already been observed in a study comparing soils from 237 different locations, where only

2% of the ensemble of bacterial phylotypes were found to be dominant[22]. Regarding functions, this distribution is consistent with a few functions being shared between different taxa and implicated in common bacterial ecology processes, while a large pool of low abundant genes confers functions specific to single taxonomical groups. Given this distribution of genera and functions, random sub-sampling of pools mainly composed by low-abundant sequences could result in the selection of false positives.

To determine whether the OTUs and functional classes identified as unique were a product of sample rarefaction, their presence in a single method was validated in the pools before rarefaction. Between 33% and 94% of the genera and functional classes identified as unique in the rarefied pool were present in more than one method pool before rarefaction and, therefore, were a product of random rarefaction rather than actual differences between DNA extraction methods (**Tables 2a and b**). Furthermore, there was a negative correlation between sample size before rarefaction and the percentage of false identification of unique functional classes, suggesting that size normalization of unequal pools is not enough to avoid the effect of sequencing depth and perform accurate comparisons between DNA extraction methods. The number of sequences associated to the OTUs and functional classes that were detected by a single method both before and after rarefaction was relatively low, which raises questions about their actual uniqueness in extracted DNA samples. Only *Staphylococcus* by the ZymoBIOMICS™ DNA Mini Kit from both soils had enough annotated sequences before rarefaction to provide a certain confidence on the uniqueness of the genera. Nevertheless, the possibility that *Staphylococcus* was present in DNA eluates extracted using other methods and was overseen during any of the steps following DNA extraction cannot be ruled out. Given that sequencing itself is a subsampling of extracted DNA, the differences detected between DNA extraction methods could be due to any other step of the soil metagenomics workflow, from sampling to sequence annotation, rather than DNA extraction. The higher numbers of sequences associated to unique functional classes are only observed in the largest pool of the cohort and, thus, support this hypothesis.

Although the effects of sequencing depth on bacterial discovery have been observed before [23,24], to the best of our knowledge this is the first study comparing the impacts of DNA extraction and sequencing depth on bacterial richness discovery. We believe this comparison is necessary to decide which investments are more urgently needed to improve the

metagenomics workflow. Our results show that sequencing depth has a more determinant effect on the bacterial richness measured from the same microbial community than the method chosen for DNA extraction. Furthermore, DNA extraction triplicates that had been sequenced at different depths had access to a different proportion of bacterial DNA richness. Similar results have been reported in a study analysing soil fungi, where higher levels of dissimilarity between replicates were found at low sequencing depths[25]. Therefore, differences found between DNA extraction methods are probably due to DNA subsampling for sequencing instead of to the choice of a DNA extraction method, and we are currently unable to determine which of these differences account for DNA extraction techniques.

However, since samples were sequenced at unequal depths, these comparisons needed a previous size normalization that inevitably caused a loss of information and the selection of false positives due to random subsampling. On the other hand, the effect of sequencing depth is still observed in functional analysis after rarefaction, which suggests that size normalization is not enough to compensate for the effect of sequencing depth on bacterial richness discovery. Whereas DNA extraction methods may affect sequencing depth, other factors, such as DNA amplification, library preparation, sequencing techniques and sequence annotation may also contribute to this inequality and lead to inaccurate comparisons between methods. Therefore, efforts should be made to optimize each of these steps in order to sequence representative samples of extracted DNA at a sufficient and equal depth. This would not only facilitate the accurate comparison between DNA extraction methods but would also help define standard methods for soil metagenomics that would improve metagenomic comparison and eventually lead to accurate profiles of soil microbiomes.

Methods

Soil sampling

Two soil samples were selected for this study (Table 3). Soils were sampled at an experimental farm (Scottish Agricultural College, Craibstone, Scotland, Grid reference NJ872104) and at a field planted with mature corn at La Côte Saint Andre, France. All samples were kept at 4°C before DNA extraction.

Table 3: Physical characterization of samples selected for this study. Details about the Scottish Agricultural College soil composition were provided by Kemp *et al.*[26]

	Scottish Agricultural College	**La Côte de Saint André**
Sand	73.85%	42.9%
Silt	20.04%	43.6%
Clay	6.11%	13.5%
pH	4.5	7.24
Organic matter	5.97%	2.92 %
Organic C	7.4%	1.7 %
Total N	0.38%	0.17 %

DNA extraction and purification

DNA was extracted from 250 mg of sample using the DNeasy® PowerSoil® Kit (QIAGEN), the ZymoBIOMICS™ DNA Mini Kit (Zymo Research) and the Phenol/Chloroform extraction method described by Griffiths *et al* [18], as well as a new semi-automated protocol in which the Maxwell® RSC Instrument (Promega) and the Maxwell® PureFood GMO and Authentication Kit (Promega) are used for DNA purification. Two variants of this protocol were tested, referred as Maxwell 1 and Maxwell 2 methods. All DNA extractions were performed in triplicate. In the Maxwell 1 method, 250 mg of sample were diluted in 1 ml of CTAB Buffer (Promega) and heated for 5 minutes at 95 °C. Samples were bead-beated twice at 5.5 m/s for 30 seconds in Lysis Matrix E tubes (MP Biomedicals) and centrifuged at 10000 rpm for 5 minutes. Then, 300 µl of supernatant were added to 300 µl of Lysis Buffer (Promega) and loaded into a Maxwell® RSC cartridge containing magnetic beads for DNA purification on the Maxwell® RSC Instrument, according to the Technical Manual TM473. A second purification using the ProNex® Size-Selective Purification System (Promega) was carried out to reduce humic acids carryover. In the Maxwell 2 method, two variants were introduced in the previously described protocol: 500 µl of CTAB Buffer were mixed with 500 µl of 0.5 M Sodium Phosphate Buffer (0.5 M Monobasic Sodium Phosphate; 0.5 M Dibasic Sodium Phosphate; pH 7.0) and added to 250 mg of sample, and cells were lysed without bead-beating.

DNA quantification and quantitative PCR (qPCR) assays

DNA concentrations were assessed using the Qubit™ Fluorometer and the Qubit™ dsDNA HS Assay Kit (ThermoFisher). Then, the size of the total bacterial community was estimated by quantifying the V3 region of the 16S rRNA gene by qPCR using the universal primers 341F (5'-CCT ACG GGA GGC AGC AG- 3') and 534R (5'-ATT ACC GCG GCT GCT GGC A-3'). qPCR assays were carried out using the Corbett Rotor-Gene 6000 (QIAGEN) in a 20 µl reaction volume containing GoTaq® qPCR Master Mix (Promega), 0.75 µM of each primer and 2 µl of DNA at ≤2.5 ng/µl. Two non-template controls were also included in all the assays. Standard curves for all the assays were obtained using 10-fold serial dilutions of a linearized plasmid pGEM-T Easy Vector (10^2 to 10^7 copies) containing the 16S rRNA gene of *Pseudomonas aeruginosa* PAO1. Cycling conditions for qPCR amplification were 95 °C for 2 minutes followed by 30 cycles of 95 °C for 15 seconds, 60 °C for 30 seconds and 72 °C for 30 seconds. Melting curves were generated after amplification by increasing the temperature from 60 °C to 95 °C.

16S rRNA gene V3-V4 amplicon sequencing and analysis

All samples were diluted to 2.5 ng/µl before 16S rRNA gene PCR amplification and sequencing, except for some samples extracted using the ZymoBIOMICS™ DNA Kit, which were already at lower concentrations. The V3-V4 hypervariable regions of bacterial 16S rRNA gene were amplified using the Titanium® Taq DNA Polymerase (Takara Clontech) and forward (5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG-3') and reverse (5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G-3') primers. Cycling conditions for PCR amplification were 95 °C for 3 minutes followed by 25 cycles of 95 °C for 30 seconds, 55 °C for 30 seconds and 72 °C for 30 seconds and a final extension step at 72°C for 5 minutes. DNA libraries were prepared based on Illumina's "16S Metagenomics Library Prep Guide" (15044223 Rev. B) using the Platinum® Taq DNA Polymerase (Invitrogen) and the Nextera® XT Index Kit V2 (Illumina). DNA sequencing with a 15% PhiX spike-in was performed using the MiSeq™ System and the MiSeq™ Reagent Kit v3 (Illumina). Reads were trimmed to meet a quality score of Q20. Then, pair-ended reads were assembled using PANDAseq[27] at a sequence length between 410 and 500 bp and an overlap length between 20 and 100 bp, using the rdp_mle algorithm. Then, each of the DNA sequences was annotated using the Ribosome Data Project (RDP) database and the RDP Bayesian classifier using an assignment confidence cut-off of 0.6[19]. In order to evaluate the effect of sequencing depth in taxonomic richness assessment, the genus richness detected in each sample was determined using the vegan package in R and plotted as a function of

sequencing depth. Then, after pooling triplicates from each method/soil couple and removing singletons, a subsample of each method pool was randomly rarefied at the lowest size of the cohort using the vegan package in R. Venn Diagrams were obtained using the VennDiagram package in R and the number of sequences annotated as unique genera (*i.e.*, genera detected by a single method) were inferred from the rarefied sample. Finally, the distribution of each unique genera before rarefaction and the number of sequences associated to those that were detected by a single method before rarefaction were determined.

Metagenomics sequencing and analysis

Metagenomics libraries were prepared from <1 ng of DNA using the Nextera® XT Library Prep Kit and Indexes (Illumina), as detailed in Illumina's "Nextera XT DNA Library Prep Kit" reference guide (15031942 v03). DNA sequencing with a 1% PhiX spike-in was performed using the MiSeq™ System and the MiSeq™ Reagent Kit v2 (Illumina). Reads were trimmed and filtered using USEARCH and sequences with a minimum quality score of Q20, a maximum of one miss-called base and a minimum length of 120 bp were blasted against the nr database using Diamond default parameters and a coverage of 60%. Sequences were functionally annotated using MEGAN6[20] and the SEED hierarchical subsystems. The third level of hierarchical functional subsystems classification was selected for richness discovery analysis. Samples extracted from both soils using the ZymoBIOMICS™ DNA Mini Kit were excluded from the analysis, since the ensemble of annotated reads from extraction triplicates did not add up to 10000. To evaluate the effect of sequencing depth in functional richness assessment, the functional class richness detected in each sample was determined using the vegan package in R and plotted as a function of sequencing depth. Then, after pooling triplicates from each method/soil couple and removing singletons, a subsample of each method pool was randomly rarefied at the lowest size of the cohort using the vegan package in R. Venn Diagrams were obtained using the VennDiagram R package, and the number of sequences annotated as unique functional classes (*i.e*., functional classes detected by a single method) were inferred from the rarefied sample. Finally, the distribution of each unique functional class before rarefaction and the number of sequences associated to those that were detected by a single method before rarefaction were determined.

Acknowledgements

We thank Graeme Nicol and Christina Hazard for soil sampling. This study funded in part by Promega Corporation.

Competing interests

This study was partially funded by Promega Corporation. **CS, LD** and **TS** are employed by Promega Corporation. **RT, LF** and **TMV** declare to have no competing interests.

Authors contributions

CS did the experimental work, contributed to experimental design, data analysis and manuscript writing. **RT** created the R scripts for data analysis and contributed to metagenomics sequencing, data analysis and manuscript writing. **LF, LD and TS** contributed to experimental design. **TMV** supervised this study and contributed to experimental design, data analysis and manuscript writing. All authors reviewed the manuscript.

Data Availability

The datasets generated and analysed during the current study are publically available in the Environmental Microbial Genomics Group repository: http://www.genomenviron.org/Research/Metagenom.html.

References

1. Roesch, L. F. W. *et al.* Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* **1,** 283–290 (2007).
2. Rappé, M. S. & Giovannoni, S. J. The Uncultured Microbial Majority. *Annu. Rev. Microbiol.* **57,** 369–394 (2003).
3. Hugenholtz, P., Goebel, B. M. & Pace, N. R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180,** 4765–4774 (1998).
4. Cruaud, P. *et al.* Influence of DNA extraction method, 16S rRNA targeted hypervariable regions, and sample origin on microbial diversity detected by 454 pyrosequencing in marine chemosynthetic ecosystems. *Appl. Environ. Microbiol.* **80,** 4626–4639 (2014).
5. Klindworth, A. *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41,** 1–11 (2013).

6. Hugerth, L. W. & Andersson, A. F. Analysing microbial community composition through amplicon sequencing: From sampling to hypothesis testing. *Front. Microbiol.* **8,** 1–22 (2017).
7. Tremblay, J. *et al.* Primer and platform effects on 16S rRNA tag sequencing. *Front. Microbiol.* **6,** 1–15 (2015).
8. Lombard, N., Prestat, E., van Elsas, J. D. & Simonet, P. Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. *FEMS Microbiol. Ecol.* **78,** 31–49 (2011).
9. Robe, P., Nalin, R., Capellano, C., Vogel, T. M. & Simonet, P. Extraction of DNA from soil. *Eur. J. Soil Biol.* **39,** 183–190 (2003).
10. Levy-Booth, D. J. *et al.* Cycling of extracellular DNA in the soil environment. *Soil Biol. Biochem.* **39,** 2977–2991 (2007).
11. Paget, E., Monrozier, L. J. & Simonet, P. Adsorption of DNA on Clay-Minerals - Protection against DNaseI and Influence on Gene-Transfer. *FEMS Microbiol. Lett.* (1992). doi:DOI 10.1111/j.1574-6968.1992.tb05435.x
12. Frostegård, Å. *et al.* Quantification of bias related to the extraction of DNA directly from soils. *Appl. Environ. Microbiol.* **65,** 5409–5420 (1999).
13. Tebbe, C. C. & Vahjen, W. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Appl. Environ. Microbiol.* **59,** 2657–2665 (1993).
14. Kauffmann, I. M., Schmitt, J. & Schmid, R. D. DNA isolation from soil samples for cloning in different hosts. *Appl. Microbiol. Biotechnol.* **64,** 665–670 (2004).
15. Miller, D. N., Bryant, J. E., Madsen, E. L. & Ghiorse, W. C. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl. Environ. Microbiol.* **65,** 4715–4724 (1999).
16. Roose-Amsaleg, C. L., Garnier-Sillam, E. & Harry, M. Extraction and Purification of Microbial DNA from Soil and Sediment Samples. *Appl. Soil Ecol.* **18,** 47–60 (2001).
17. Marketa Sagova-Mareckova, Ladislav Cermak, Jitka Novotna, Kamila Plhackova, Jana Forstova, J. K. Innovative Methods for Soil DNA Purification Tested in Soils with Widely Differing Characteristics. *Appl. Environ. Microbiol.* **74,** 2902–2907 (2008).
18. Griffiths, R. I., Whiteley, A. S., O'Donnell, A. G. & Bailey, M. J. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl. Environ. Microbiol.* **66,** 5488–5491 (2000).
19. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73,** 5261–5267 (2007).
20. Huson, D., Mitra, S., Ruscheweyh, H., Weber, N. & Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* **21,** 1552–1560 (2011).
21. Delmont, T. O., Simonet, P. & Vogel, T. M. Describing microbial communities and performing global comparisons in the omic era. *ISME J.* **6,** 1625–1628 (2012).
22. Delgado-Baquerizo, M. *et al.* A global atlas of the dominant bacteria found in soil. *Science (80-.).* **359,** 320–325 (2018).
23. Zaheer, R. *et al.* Impact of sequencing depth on the characterization of the microbiome and resistome. *Sci. Rep.* **8,** 1–11 (2018).
24. Lundin, D. *et al.* Which sequencing depth is sufficient to describe patterns in bacterial α- and β-diversity? *Environ. Microbiol. Rep.* **4,** 367–372 (2012).

25. Smith, D. P. & Peay, K. G. Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. *PLoS One* **9,** (2014).
26. Kemp, J. S., Paterson, E., Gammack, S. M., Cresser, M. S. & Killham, K. Leaching of genetically modified Pseudomonas fluorescens through organic soils: Influence of temperature, soil pH, and roots. *Biol. Fertil. Soils* **13,** 218–224 (1992).
27. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: Paired-end assembler for illumina sequences. *BMC Bioinformatics* **13,** 31 (2012).

Sequencing depth rather than DNA extraction determines soil bacterial richness discovery

Concepcion Sanchez-Cid[1,2], Romie Tignat-Perrier[1,3], Laure Franqueville[1], Laurence Delaurière[2], Trista Schagat[4], Timothy M. Vogel[1]*

1. Environmental Microbial Genomics, Laboratoire Ampère, Ecole Centrale de Lyon, Université de Lyon, Ecully, France
2. Promega France, Charbonières-les-Bains, France
3. Promega Corporation, Madison, Wisconsin, United States
4. Institut des Géosciences de l'Environnement, Université Grenoble Alpes/CNRS/Institut de Recherche pour le Développement, G-INP, Grenoble, France

*concepcion.sanchezcid-torres@ec-lyon.fr

Supplementary Material

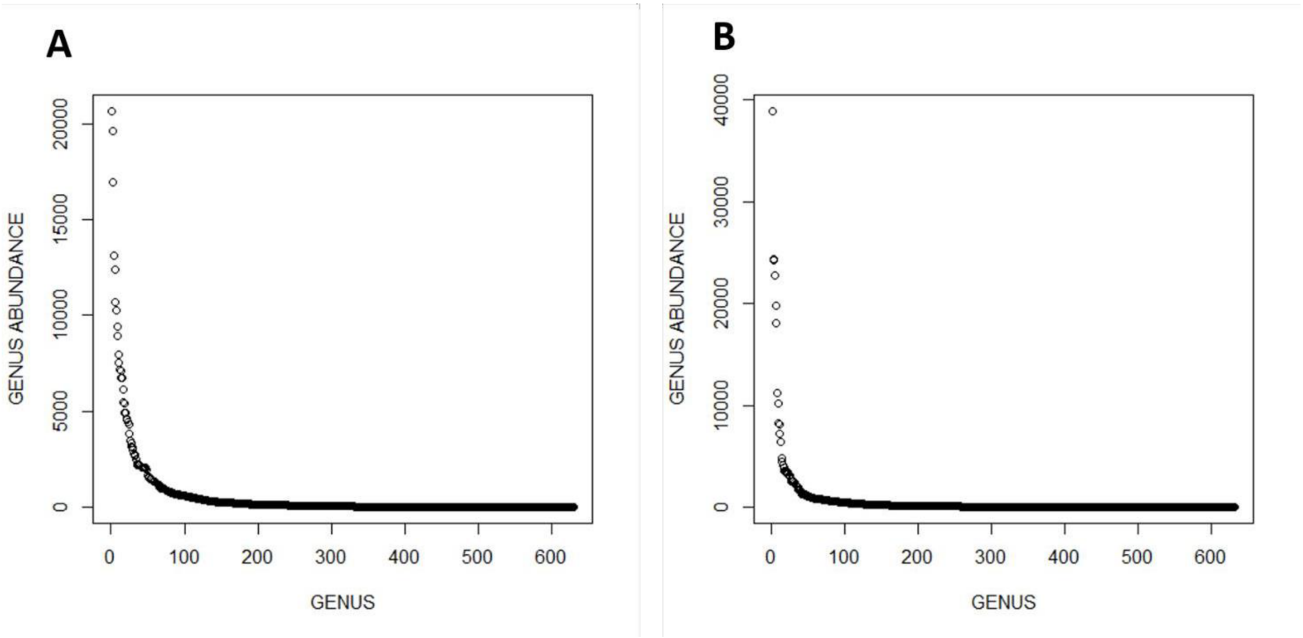


Figure S1. Total abundance of the ensemble of genera annotated using RDP classifier from DNA extracted from (A) the Scottish Agricultural College soil; (B) La Côte de Saint André soil. The abundance of each measured genus was determined and plotted in a decreasing order using R. N=3.

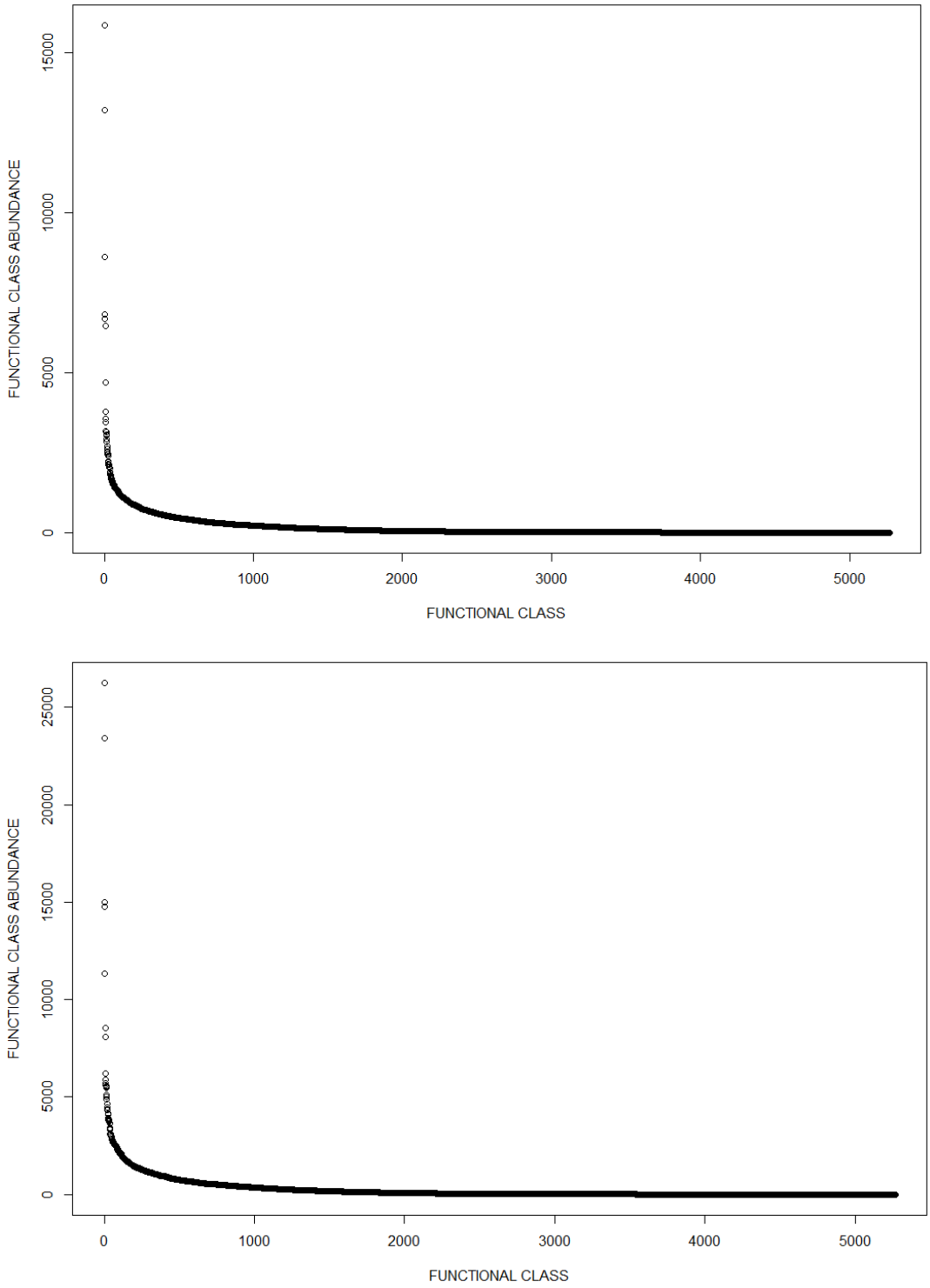


Figure S2. Total abundance of the ensemble of functional classes classified using SEED from DNA extracted from (A) the Scottish Agricultural College soil; (B) La Côte de Saint André soil. The abundance of each measured functional class was determined and plotted in a decreasing order using R. N=3.

SCOTTISH AGRICULTURAL COLLEGE SOIL								
Genus	**Number of sequences**	**Average RA**	**Standard Deviation**	**RA Maxwell 1**	**RA Maxwell 2**	**RA Phenol/ Chloroform**	**RA PowerSoil**	**RA Zymo**
Gp1	17018	5,80	1,46	5,31	7,64	5,47	6,76	3,84
Rhodanobacter	16415	5,60	1,75	4,05	6,46	7,82	6,05	3,61
Gp2	14188	4,84	1,56	3,73	7,21	4,85	5,22	3,18
Saccharibacteria genera incertae sedis	10833	3,69	1,56	4,85	5,37	2,58	4,04	1,64
Bacillus	9385	3,20	1,90	5,58	0,29	2,96	3,53	3,63
Subdivision3_genera incertae sedis	8834	3,01	1,25	3,86	4,58	2,22	2,94	1,46
Gp3	8514	2,90	0,71	2,63	4,13	2,45	2,91	2,40
Gemmatimonas	8012	2,73	2,11	1,38	2,21	5,56	4,16	0,36
WPS-2 genera incertae sedis	7130	2,43	0,39	2,10	2,00	2,81	2,39	2,85
Gaiella	6304	2,15	1,35	1,31	1,35	2,64	1,13	4,31
Conexibacter	5824	1,99	1,64	0,93	0,78	2,11	1,36	4,77
Rudaea	5821	1,99	0,50	1,48	2,53	1,42	2,30	2,20
Arthrobacter	5708	1,95	0,97	1,62	0,64	3,08	1,66	2,73
Rhizomicrobium	5563	1,90	0,37	1,46	2,45	2,03	1,70	1,85
Spartobacteria genera incertae sedis	5446	1,86	0,34	1,48	1,66	1,71	2,26	2,17
Candidatus Solibacter	5050	1,72	0,47	1,54	2,54	1,40	1,43	1,70
Phenylobacterium	4489	1,53	0,71	1,07	2,46	1,12	0,87	2,13
Roseiarcus	4313	1,47	0,93	0,83	1,56	0,92	0,99	3,05
Povalibacter	4030	1,37	0,27	1,00	1,64	1,60	1,20	1,44
Ktedonobacter	3968	1,35	0,89	1,61	0,38	2,55	1,66	0,57
Mucilaginibacter	3811	1,30	0,61	1,54	1,85	1,01	1,72	0,37
Parcubacteria genera incertae sedis	3733	1,27	0,48	1,51	1,86	1,14	1,28	0,57
Rhodoferax	3613	1,23	0,44	1,01	1,69	1,67	1,12	0,67
Nakamurella	3537	1,21	0,86	0,65	0,09	2,14	1,22	1,92
WPS-1_genera incertae sedis	3171	1,08	0,30	0,91	1,57	0,99	1,15	0,78

Table S1a. Average relative abundance (RA) and relative abundances per method of the 25 most abundant genera from the rarefied pool of sequences from the Scottish Agricultural College soil. Maxwell 1 and 2: variants of the Maxwell® PureFood GMO and Authentication Kit (Promega). PowerSoil: DNeasy® PowerSoil® Kit (QIAGEN); Zymo: ZymoBIOMICS™ DNA Mini Kit (Zymo Research). Relative abundances higher than the average plus one standard deviation are highlighted in orange, and those lower than the average minus one standard deviation are highlighted in blue. N=3.

LA COTE DE SAINT ANDRE SOIL								
Genus	**Number of sequences**	**Average RA**	**Standard Deviation**	**RA Maxwell 1**	**RA Maxwell 2**	**RA Phenol/ Chloroform**	**RA PowerSoil**	**RA Zymo**
Spartobacteria_genera_incertae_sedis	20362	11,22	1,79	9,91	13,50	9,37	10,62	12,68
Gp1	11785	6,49	1,94	7,52	8,29	5,15	7,73	3,76
Gaiella	11774	6,49	5,06	3,15	1,72	14,74	6,80	6,02
Gp6	10806	5,95	1,89	6,68	6,96	6,05	7,39	2,68
Subdivision3 genera incertae sedis	10177	5,61	2,26	6,08	8,62	2,30	5,17	5,86
Gp3	9035	4,98	1,02	5,46	6,22	4,20	5,32	3,68
Bacillus	5040	2,78	2,32	3,53	0,30	2,09	1,57	6,38
Gemmatimonas	4987	2,75	1,73	3,50	3,15	2,43	4,65	0,01
Gp2	4807	2,65	1,53	3,57	4,10	1,38	3,55	0,64
Candidatus Solibacter	4164	2,29	0,28	2,21	2,59	1,88	2,53	2,26
WPS-1 genera incertae sedis	3450	1,90	0,53	1,91	2,49	1,70	2,27	1,13
Gp16	3329	1,83	1,26	0,84	0,72	3,75	1,50	2,36
Saccharibacteria genera incertae sedis	2483	1,37	0,50	1,08	1,67	1,01	0,97	2,11
Gp5	2440	1,34	0,66	1,53	2,33	0,63	1,37	0,85
Bradyrhizobium	1981	1,09	0,62	0,90	0,45	1,27	0,78	2,06
Ktedonobacter	1919	1,06	0,67	1,42	0,23	1,90	1,18	0,55
Nitrospira	1879	1,03	0,65	1,41	1,74	0,59	1,29	0,14
Paenibacillus	1872	1,03	0,81	1,24	0,10	0,85	0,69	2,28
Chryseolinea	1820	1,00	0,51	1,23	1,40	0,44	0,48	1,47
Rhizomicrobium	1789	0,99	0,05	1,06	0,99	0,94	0,98	0,95
Candidatus Koribacter	1742	0,96	0,14	1,04	1,09	0,97	0,95	0,74
Parcubacteria genera incertae sedis	1676	0,92	0,40	0,85	1,58	0,62	0,59	0,98
Gp4	1675	0,92	0,52	1,01	1,72	0,66	0,88	0,34
Aridibacter	1645	0,91	0,42	0,87	1,56	0,40	0,76	0,94
Pirellula	1599	0,88	0,42	0,95	1,45	0,75	0,96	0,29

Table S1b. Average relative abundance (RA) and relative abundances per method of the 25 most abundant genera from the rarefied pool of sequences from La Côte de Saint André soil. Maxwell 1 and 2: variants of the Maxwell® PureFood GMO and Authentication Kit (Promega). PowerSoil: DNeasy® PowerSoil® Kit (QIAGEN); Zymo: ZymoBIOMICS™ DNA Mini Kit (Zymo Research). Relative abundances higher than the average plus one standard deviation are highlighted in orange, and those lower than the average minus one standard deviation are highlighted in blue. N=3.

SCOTTISH AGRICULTURAL COLLEGE SOIL							
Functional class	**Number of sequences**	**Average RA**	**Standard Deviation**	**RA Maxwell 1**	**RA Maxwell 2**	**RA Phenol/ Chloroform**	**RA PowerSoil**
5-FCL-like protein	20362	11,22	1,79	9,91	13,50	9,37	10,62
Long-chain-fatty-acid--CoA ligase	11785	6,49	1,94	7,52	8,29	5,15	7,73
3-oxoacyl-[acyl-carrier protein] reductase	11774	6,49	5,06	3,15	1,72	14,74	6,80
TonB-dependent receptor	10806	5,95	1,89	6,68	6,96	6,05	7,39
Cobalt-zinc-cadmium resistance protein CzcA	10177	5,61	2,26	6,08	8,62	2,30	5,17
Adenylate cyclase	9035	4,98	1,02	5,46	6,22	4,20	5,32
COG2363	5040	2,78	2,32	3,53	0,30	2,09	1,57
Butyryl-CoA dehydrogenase	4987	2,75	1,73	3,50	3,15	2,43	4,65
Enoyl-CoA hydratase	4807	2,65	1,53	3,57	4,10	1,38	3,55
Aldehyde dehydrogenase	4164	2,29	0,28	2,21	2,59	1,88	2,53
Acriflavin resistance protein	3450	1,90	0,53	1,91	2,49	1,70	2,27
Aspartate aminotransferase	3329	1,83	1,26	0,84	0,72	3,75	1,50
Beta-lactamase	2483	1,37	0,50	1,08	1,67	1,01	0,97
UDP-glucose 4-epimerase	2440	1,34	0,66	1,53	2,33	0,63	1,37
3-ketoacyl-CoA thiolase	1981	1,09	0,62	0,90	0,45	1,27	0,78
DNA-directed RNA polymerase beta' subunit	1919	1,06	0,67	1,42	0,23	1,90	1,18
Diguanylate cyclase/phosphodiesterase	1879	1,03	0,65	1,41	1,74	0,59	1,29
DNA-directed RNA polymerase beta subunit	1872	1,03	0,81	1,24	0,10	0,85	0,69
DNA polymerase III alpha subunit	1820	1,00	0,51	1,23	1,40	0,44	0,48
Thioredoxin reductase	1789	0,99	0,05	1,06	0,99	0,94	0,98
FIG039061: hypothetical protein related to heme utilization	1742	0,96	0,14	1,04	1,09	0,97	0,95
5-methyltetrahydrofolate--homocysteine methyltransferase	1676	0,92	0,40	0,85	1,58	0,62	0,59
Acetyl-coenzyme A synthetase	1675	0,92	0,52	1,01	1,72	0,66	0,88
D-3-phosphoglycerate dehydrogenase	1645	0,91	0,42	0,87	1,56	0,40	0,76
Carbamoyl-phosphate synthase large chain	1599	0,88	0,42	0,95	1,45	0,75	0,96

Table S1c. Average relative abundance (RA) and relative abundances per method of the 25 most abundant functional classes from the rarefied pool of sequences from the Scottish Agricultural College soil. Maxwell 1 and 2: variants of the Maxwell® PureFood GMO and Authentication Kit (Promega). PowerSoil: DNeasy® PowerSoil® Kit (QIAGEN). Relative abundances higher than the average plus one standard deviation are highlighted in orange, and those lower than the average minus one standard deviation are highlighted in blue. N=3.

LA COTE DE SAINT ANDRE SOIL							
Functional class	**Number of sequences**	**Average RA**	**Standard Deviation**	**RA Maxwell 1**	**RA Maxwell 2**	**RA Phenol/ Chloroform**	**RA PowerSoil**
5-FCL-like protein	12027	1,88	0,146	1,82	1,70	2,01	2,00
Long-chain-fatty-acid--CoA ligase	10829	1,69	0,065	1,77	1,62	1,72	1,67
3-oxoacyl-[acyl-carrier protein] reductase	6756	1,06	0,096	1,08	0,92	1,14	1,08
Adenylate cyclase	6685	1,05	0,070	1,03	0,96	1,07	1,13
Cobalt-zinc-cadmium resistance protein CzcA	5687	0,89	0,190	0,96	1,11	0,67	0,82
COG2363	4041	0,63	0,047	0,65	0,69	0,61	0,58
TonB-dependent receptor	3978	0,62	0,158	0,68	0,81	0,44	0,56
Arylsulfatase	2761	0,43	0,021	0,46	0,44	0,41	0,42
Butyryl-CoA dehydrogenase	2753	0,43	0,051	0,41	0,37	0,49	0,46
Beta-lactamase	2666	0,42	0,022	0,42	0,44	0,39	0,42
Acriflavin resistance protein	2666	0,42	0,044	0,44	0,47	0,36	0,41
UDP-glucose 4-epimerase	2550	0,40	0,015	0,38	0,41	0,39	0,41
Enoyl-CoA hydratase	2532	0,40	0,041	0,38	0,35	0,41	0,44
Aldehyde dehydrogenase	2450	0,38	0,039	0,40	0,32	0,41	0,40
Aspartate aminotransferase	2406	0,38	0,019	0,38	0,40	0,37	0,36
DNA-directed RNA polymerase beta' subunit	2360	0,37	0,016	0,36	0,37	0,36	0,39
DNA polymerase III alpha subunit	2290	0,36	0,018	0,34	0,35	0,38	0,36
DNA-directed RNA polymerase beta subunit	2122	0,33	0,014	0,32	0,35	0,34	0,33
Excinuclease ABC subunit A	2079	0,33	0,027	0,35	0,35	0,30	0,30
3-ketoacyl-CoA thiolase	2045	0,32	0,040	0,32	0,27	0,36	0,33
Acetyl-coenzyme A synthetase	2009	0,31	0,038	0,31	0,26	0,35	0,33
D-3-phosphoglycerate dehydrogenase	1988	0,31	0,022	0,29	0,34	0,30	0,31
Diguanylate cyclase/phosphodiesterase	1931	0,30	0,018	0,33	0,30	0,29	0,29
Thioredoxin reductase	1926	0,30	0,020	0,31	0,28	0,30	0,32
Hypothetical protein related to heme utilization	1908	0,30	0,010	0,29	0,31	0,29	0,31

Table S1d. Average relative abundance (RA) and relative abundances per method of the 25 most abundant functional classes from the rarefied pool of sequences from La Côte de Saint André soil. Maxwell 1 and 2: variants of the Maxwell® PureFood GMO and Authentication Kit (Promega). PowerSoil: DNeasy® PowerSoil® Kit (QIAGEN). Relative abundances higher than the average plus one standard deviation are highlighted in orange, and those lower than the average minus one standard deviation are highlighted in blue. N=3.

SCOTTISH AGRICULTURAL COLLEGE SOIL			
Method	**Genus**	**Annotated sequences after rarefaction**	**Annotated sequences before rarefaction**
Maxwell 1	Gp22	2	2
	Arenicella	2	2
	Phyllobacterium	2	2
	Falsirhodobacter	2	2
	Hydrogenophaga	1	4
	Geothermomicrobium	2	2
	Curtobacterium	2	3
	Nitrosopumilus	3	4
Maxwell 2	Parapedobacter	2	2
	Chelatococcus	2	2
	Amantichitinum	2	2
	Leucobacter	4	4
Phenol/Chloroform method	Acaricomes	2	2
	Sphaerobacter	3	3
	Salinispira	2	2
DNeasy® PowerSoil® Kit	Syntrophobacter	2	2
	Chitinibacter	2	2
	Thermomarinilinea	3	3
ZymoBIOMICS™ DNA Mini Kit	Fulvimonas	2	2
	Nitratireductor	2	2
	Paenochrobactrum	1	2
	Sorangium	2	2
	Desulfomonile	2	2
	Hungatella	2	2
	Staphylococcus	63	95
	Thermocrispum	2	2

Table S2a. List of genera detected as unique in both rarefied and non-rarefied pools of sequences from the Scottish Agricultural College soil. Maxwell 1 and 2: variants of the Maxwell® PureFood GMO and Authentication Kit (Promega). N=3.

LA COTE DE SAINT ANDRE SOIL			
Method	**Genus**	**Annotated sequences after rarefaction**	**Annotated sequences before rarefaction**
Maxwell 1	Limnobacter	2	2
	Terribacillus	1	2
	GpXIII	2	3
Maxwell 2	Epilithonimonas	1	3
	Pseudenhygromyxa	4	4
	Thioreductor	2	2
Phenol/Chloroform method	Sporomusa	2	2
	Actinocatenispora	3	3
DNeasy® PowerSoil® Kit	Tolumonas	1	2
	Nereida	1	2
	Smithella	1	2
ZymoBIOMICS™ DNA Mini Kit	Alistipes	2	2
	Desulfomonile	2	2
	Nitrosomonas	6	6
	Gracilibacillus	6	6
	Staphylococcus	130	130
	Anaerosinus	3	3
	Melioribacter	2	2

Table S2b. List of genera detected as unique in both rarefied and non-rarefied pools of sequences from La Côte de Saint André soil. Maxwell 1 and 2: variants of the Maxwell® PureFood GMO and Authentication Kit (Promega). N=3.

Aerobiologia
https://doi.org/10.1007/s10453-019-09606-x

ORIGINAL PAPER

Microbial composition in seasonal time series of free tropospheric air and precipitation reveals community separation

Nora Els · Catherine Larose · Kathrin Baumann-Stanzer · Romie Tignat-Perrier · Christoph Keuschnig · Timothy M. Vogel · Birgit Sattler

Received: 25 January 2019 / Accepted: 26 August 2019

Abstract Primary biological aerosols are transported over large distances, are traveling in various media such as dry air masses, clouds or fog, and eventually deposited with dry deposition, especially for larger particles, or precipitation like rain, hail or snow. To investigate relative abundance and diversity of airborne bacterial and fungal communities, samples have been collected with a liquid impinger (Coriolis μ) from the top of Mount Sonnblick (3106 m asl, Austrian Alps) from the respective sources under a temporal aspect over four seasons over the year to include all climatic conditions. Bacterial and fungal samples (16S rRNA and ITS) were sequenced using Illumina MiSeq paired-end sequencing, investigated for relative abundance by qPCR(16S rRNA and 18S rRNA) and ice nucleation activity. Results show that there is no stable free tropospheric air microbial community and air mass origin was different for the four sampling periods which exerted influence on the microbial composition of the atmosphere although a core microbiome could be identified consisting of 61 bacterial OTUs and eight fungal genera. Differentiation between seasons was stronger pronounced in air than in precipitation, with rain being most different and variable of precipitation types, indicating distinct forces driving microbial fate in the air. Microorganisms precipitated with snow, hail or rain or being transported by clouds differ in their species composition from free tropospheric air masses and do not mirror the air community structure. They were more diverse, distinct in composition, 16S:18S ratio and abundance from free-floating PBA. Hence, snow or cloud samples are not suitable proxies for free tropospheric air microbiome composition, since separation processes in aerosolization, transport and scavenging occur. The microbial composition of arriving precipitation or clouds represents only a part of the microbial air composition communities of the cumulative sources of origin. Relative abundance and composition of ice nucleation-active bacteria showed a higher share of relative % reads of known ice nucleation-active bacteria present in all wet phases compared to air. Results propose a separation of IN-

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10453-019-09606-x) contains supplementary material, which is available to authorized users.

N. Els (✉) · B. Sattler
Institute of Ecology, University of Innsbruck, Innsbruck, Austria
e-mail: nora.els@uibk.ac.at

N. Els · B. Sattler
Austrian Polar Research Institute, Vienna, Austria

C. Larose · R. Tignat-Perrier · C. Keuschnig · T. M. Vogel
Environmental Microbial Genomics, Laboratoire Ampère, École Centrale de Lyon, Université de Lyon, Écully, France

K. Baumann-Stanzer
Zentralanstalt für Meteorologie und Geodynamik, Vienna, Austria

Published online: 04 September 2019

active reads with higher shares occurring in precipitation. This study presents the first comparison of free tropospheric bacterial and fungal abundance and diversity in time series of air over several seasons in contrast to various precipitation forms in the free troposphere.

Keywords Free troposphere · 16SrRNA · 18SrRNA · ITS · Precipitation · Aerobiology

1 Introduction

Knowledge about and scientific interest in the composition and abundance of primary biological aerosols (PBA) (e.g., airborne bacteria and fungi) is rapidly increasing (180% increase in publications per year from 2000 to 2018, search term: "airborne bacteria"). Most aerobiology studies are restricted to ground-based sampling within the planetary boundary layer (PBL) or focus on extreme events such as heat events and dust storms (Fang et al. 2018; Jeon et al. 2011; Maki et al. 2017; Gat et al. 2017; Fierer et al. 2008; Bowers et al. 2013), civilization and health related questions, like disease spread (Park et al. 2014; Adams et al. 2015; Di Giulio et al. 2010; Kelley and Gilbert 2013; Hagan et al. 1995; Polymenakou 2012; Griffin 2007; Brown and Hovmøller 2002) or pollen events (Burge 2002; Damialis et al. 2015; Ziello et al. 2012; Clot 2003; Jäger 2000; Gioulekas et al. 2004; Sofiev et al. 2013; Hamaoui-Laguel et al. 2015). Microbial composition in air masses within the PBL is affected by ground emissions and mixing of air masses (Bowers et al. 2011; Carotenuto et al. 2017; Burrows et al. 2009a, b). Within the PBL (upper boundary around 100–1000 m above ground level, AGL), vertical mixing of air masses over mountainous terrain is driven by ground turbulence, thermal convection and dynamic forcing with strong diurnal and seasonal variations in elevation AGL and mixing force (Rotach et al. 2015). Above the PBL, laminar flowing air masses are stratified in the free troposphere (Wekker et al. 2015).

Studies focusing on free tropospheric air microbial communities often present data from single campaigns or few flights (Schmale and Ross 2015; Techy et al. 2010; Jimenez-Sanchez et al. 2018; Smith et al. 2018; Maki et al. 2017; Xia et al. 2013; Zweifel et al. 2012) and have shown that microbes present in the free troposphere might originate from long-range travel (Maki et al. 2017; Smith and David 2012; Smith et al. 2011, 2012; Smith 2013; Brown and Hovmøller 2002; Burrows et al. 2009a, b). Air masses above and below PBL differ significantly in microbial composition. Free tropospheric habitats tend to have more Firmicutes (Smith et al. 2018; Smith 2013), Proteobacteria, Burkholderiales (DeLeon-Rodriguez et al. 2013) and extremophile yeasts, Saccharomycetes and Microbotryomycetes (Els et al. 2019). While the existence of either a stable (DeLeon-Rodriguez et al. 2013) or a highly variable (Zweifel et al. 2012) extremophile free tropospheric PBA community is disputed, generally few studies addressing the question were mainly performed over marine and oceanic regions (DeLeon-Rodriguez et al. 2013). Data on free tropospheric continental background PBA composition and variation are even more sparse due to numerous infrastructural challenges, like accessability, need of aerial vehicles, suitable technologies for sufficient air sampling volume and challenging meteorological conditions. Thus, insight into seasonal variability and free-floating PBA composition compared to cloud and precipitation composition in free tropospheric conditions is limited.

Microorganisms are not only ubiquitous in dry air masses, but also in clouds (Amato et al. 2005, 2017; Joly et al. 2014; Evans et al. 2019; Fuzzi et al. 1997), snow (Amato et al. 2007a, b; Honeyman et al. 2018; Elster et al. 2007; Harding et al. 2011; Larose et al. 2010), hail (Temkiv et al. 2012; Vali 1971; Michaud et al. 2014) and rain (Christner et al. 2008a, b; Huffman et al. 2013). Their densities and diversity are higher in wet air masses and precipitation than in clear air due to source of moisture (Xia et al. 2013; Evans et al. 2019). Aerosolized microorganisms can act as cloud condensation nuclei (Bauer et al. 2003) and be incorporated in clouds. Clouds likely offer a more hospitable environment due to water availability and nutrients (Delort et al. 2017; Deguillaume et al. 2008; Bianco et al. 2016). Bacteria might be able to multiply (Sattler et al. 2001) and degrade substances (Lallement et al. 2018) in cloud doplets.

Microorganisms could even trigger their aerosolization and precipitation voluntarily by ice nucleation (IN) capacity as part of their life cycle, which was observed to express in enhanced concentrations of PBA and biological IN-activity in snow and

rain (Morris et al. 2014; Christner et al. 2008a, b; Huffman et al. 2013; Pouzet et al. 2017; Stopelli et al. 2015).

A range of processes, such as aerosolisation (Alsved et al. 2018), vapor to liquid phase change, particle coalescence, in-cloud circulation, temperature regime, cloud lifetime, cloud trajectories, nucleation properties and local air composition at the site of precipitation, effect on the microbial composition of precipitation (Möhler et al. 2007).

Recent publications used snow (Cáliz et al. 2018; Weil et al. 2017) or cloud sampling (Maki et al. 2017; Jiaxian et al. 2019) as a proxy for free tropospheric air microbial composition. However, the insights gained from precipitation and cloud studies above PBL might not directly depict the patterns in free tropospheric microbial communities.

Before being collected on the ground, precipitation travels through the PBL, where an inversion layer of accumulated particles is concentrated at the upper PBL (Brunet 2017). Scavenging impacts the microbial composition of droplets that pick up local PBA, but also release particles during transport and sedimentation (Jang et al. 2018; Huffman et al. 2013). Below-cloud scavenging of aerosol particles by snow is a known important mechanism of wet deposition (Paramonov et al. 2011), with differing effectivity for chemical components based on the forms of precipitation (Zhang et al. 2013; Zikova and Zdimal 2016). Thus, the scavenging effect might also be selective for PBA of certain properties (*e.g.*, size, hygroscopicity, density, sphericity and interface interactions) between precipitation types and thus foster a selection for specific PBA (Möhler et al. 2007).

This study presents the first comparison of free tropospheric bacterial and fungal abundance and diversity in time series of air over several seasons in contrast to various precipitation forms in the free troposphere. We further estimated the relative abundance and composition of ice nucleation-active bacteria in order to investigate whether the separation between air and precipitation forms in terms of ice nuclei can be observed on the molecular level, as it is observed on physical IN-effectivity.

2 Materials and methods

2.1 Study site

Samples were collected at the roof top terrace of Mount Sonnblick Observatory (3106 m asl.), Austria (47°3′14″N, 12°57′27″E, Fig. 1a, b). This location is operated as a meteorological observatory and is member of Global Atmosphere Watch of the World Meteorological Organisation (WMO). The site is exhaust neutral, supplied by electricity from the valley and equipped with an elevated air outlet to enable undisturbed air measurements. The vegetation in the valley consists of high alpine forest and shrubs that are dominated by pine (*Pinus cembra, Pinus mugo, Pinus silvestris*) and spruce (*Picea abies, Picea montana*). The closest settlement is the village Rauris village, which is 15 km away. The area within 100 km perimeter around the Sonnblick Observatory is sparsely populated.

In May, a complete snow cover was prevailing in the valley and on the mountain with high fresh snowfall (50–100 cm) before the sampling period, in August no snow cover was present at the mountain site or the valley, and in November and February a deep snow cover was present at the mountain site as well as at valley floor.

2.2 Air and precipitation sampling

A total of 17 snow, ten cloud samples, two rain and one hail sample were collected (Table 1). In total, 48 air samples comprising 288 m^3 air were taken on four sampling campaigns in the months of May (2 to 5 May 2017, 13 samples, 78 m^3), August (3 to 6 August 2017, 11 samples, 66 m^3), November (27 November to 1 December 2017, 12 samples, 72 m^3) and February (20 to 23 February 2018, 13 samples, 78 m^3) approximately every 4 h. We used a liquid impinger (Coriolis µ, bertin Technologies, Montigny-le-Bretonneux, France) at a flow rate of 300L min^{-1} for 10 min (i.e., 3 m^3 each) into separate sterile vials filled with 15 mL 0.2 µm filtered, distilled water. Before each sampling for sequencing, one run was taken as decontamination purging. Two temporally close samples (i.e., within the same hour) of three cubic meters each were pooled to account for low cell density in high elevations. Sampling could not always be conducted at the planned 4-hour interval, due to strong

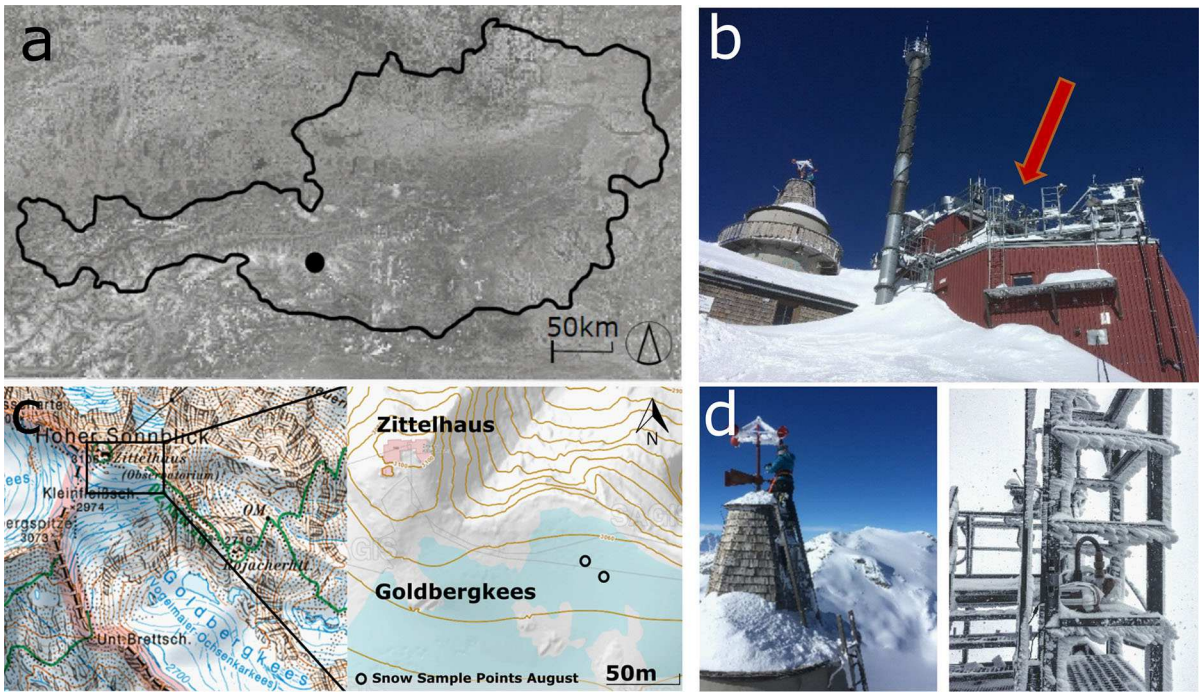


Fig. 1 **a** Location of mount Sonnblick in Austria, **b** location of the sampler on the roof terrace of the observatory, **c** location of August snow (slush) sample points, source:www.salzburg.gv.at/sagisonline, **d** rime ice (supercooled atmospherical liquid water droplets frozen on surfaces) sampling from roof and railing

winds or precipitation events or too cold temperatures which hampered the functionality of the sampling device (compare Table 1).

Snow was sampled into sterile Nasco Whirl-Pak (Roth, Karlsruhe, Germany) after fresh snow fall with a stainless steel shovel (Roth, Karlsruhe, Germany). Before each sampling, we rinsed the shovel three times with 0.2 µm filtered milli-Q and turned it ten times in the snow, like the type of snow we would then sample, to precontaminate, usually this was conducted around 2 m apart from the actual sampling site. Thus, the idea of precontamination was to dispose non-site microbes and cover the shovel with the microbes we wanted to sample before the actual sampling. The shovel was kept in a sterile Nasco Whirl-Pak (Roth, Karlsruhe, Germany) for transport, to avoid contamination, and was washed three times with HCl (15%) and milli-Q before taken to the field. We did not use EtOH, as this can leave remnants that can then interfere with the aquatic chemistry measurements, we also conducted with these samples. Only the uppermost fresh snow layer was sampled.

In August, slush (melting snow and ice) was sampled from the upper section of the Goldbergkees, in approx. 500 m direct distance from the observatory (see map Fig. 1c). Cloud water (i.e., supercooled atmospherical liquid water droplets) was collected as rime ice frozen to the railing and meteorological devices on the roof terrace and tower and roof of the observatory (see picture for sampling illustration and locations in Fig. 1d) with a precontaminated stainless steel shovel (Roth, Karlsruhe, Germany) into sterile Nasco Whirl-Pak (Roth, Karlsruhe, Germany). Rime in direct contact with surfaces was not sampled. Rain was sampled as wet deposition in a WADOS (wet and dry only precipitation sampler, Kroneis GmbH, Vienna, Austria) as a 24-h cumulative sample at 10:00 h the 5.8.2017 and 6.8.2017, respectively. Hail was collected directly into sterile Nasco Whirl-Pak (Roth, Karlsruhe, Germany) during a strong thunderstorm event on the 5.8.2017 (20:00 h).

Table 1 Sampling dates and times of analyzed air samples and precipitation, times indicated in UTC + 1 (summer and winter time), samples were taken within one hour around indicated time

May						
02.05.2017			T1	T2		Snow
			13:00 h	19:00 h		
03.05.2017	T3	T4	T5	T6	T7	Cloud
	07:00 h	11:00 h	15:00 h	19:00 h	23:00 h	
04.05.2017	T8	T9	T10	T11	T12	Cloud tower I, II
	07:00 h	11:00 h	15:00 h	19:00 h	23:00 h	Cloud terrace
05.05.2017	T13					Snow
	07:00 h					
August						
03.08.2017				T1		
				19:00		
04.08.2017	T2	T3	T4			
	07:00 h	11:00 h	15:00 h			
05.08.2017	T5	T6	T7	T8		Slush snow I, II
	07:00 h	13:00 h	15:00 h	19:00 h		
06.08.2017	T9	T10	T11			Rain, hail
	07:00 h	11:00 h	15:00 h			
November						
27.11.2017				T1	T2	Rain
				19:00 h	23:00 h	
28.11.2017	T3	T4	T5	T6		Snow
	07:00 h	11:00	15:00 h	19:00 h		
29.11.2017		T7	T8			Cloud, snow
		11:00 h	15:00 h			
30.11.2017	T9	T10		T11		Cloud I, II, Snow
	07:00 h	11:00 h		19:00		
01.12.2017	T12					Snow
	07:00 h					
February						
20.02.2018			T1	T2		Cloud, snow I, II
			16:00 h	19:00 h		
21.02.2018	T4	T5	T6	T7		Snow I, II, III
	07:00 h	11:00 h	15:00 h	19:00 h		Cloud,
22.02.2018	T8	T9	T10	T11		Snow I, II, III, IV
	07:00 h	11:00 h	15:00 h	19:00 h		Cloud
23.02.2018	T12	T13				
	07:00 h	11:00 h				

2.3 DNA extraction and sequencing

Snow and cloud water were molten directly after sampling at room temperature (20 °C), and 1L of molten sample was filtered though a 0.2 µm polycarbonate filter (47 mm, Isopore, Merck, Darmstadt, Germany) as soon as liquid, and then stored frozen at − 20 °C until DNA extraction. The air samples, collected in 15 mL solution of sterile distilled water, were stored at − 20 °C, until filtered

on a 0.2 µm polycarbonate filter (47 mm, Isopore, Merck, Darmstadt, Germany). DNA was extracted from the filters using DNeasy Power Water extraction kit (QIAGEN, Venlo, Netherlands) following the protocol provided with the kit, as this kit was stated to yield good results for aerobiological studies (Dommergue et al. 2019). Amplification, library prep (MiSeq Illumina sequencing, 2 × 250 bp, Nextera XT Library Preparation Kit, Illumina, San Diego, USA) and sequencing were done with the Environmental Microbial Genomics group at the Laboratoire Ampère (ECL Lyon, University of Lyon, France). Community diversity was targeted: the V3-V4 region of the bacterial 16S rRNA SSU gene was amplified using 341F/785R (S-D-Bact-0341-b-S-17/S-D-Bact-0785-a-A-21, Klindworth et al. 2013) primers, and the fungal internal transcribed spacer (ITS) regions were amplified with primer pair 5.8S_Fung/ITS4 targeting the ITS2 region (Taylor et al. 2016).

Raw sequences were submitted to NCBI BioProject database under Project ID PRJNA516816 (http://www.ncbi.nlm.nih.gov/bioproject/516816).

2.4 Bioinformatics and statistics

For 16S sequences, the base quality of the reads 1 and reads 2 was controlled using the functions fastx_quality_stats and fastq_quality_boxplot_graph of the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). PANDAseq (Masella et al. 2012) was used to assemble the read 1 and the read 2 using the RDP algorithm, a minimum and maximum length of the resulting sequence of 410 bp and 500 bp, respectively, a minimum and maximum overlap length of 20 bp and 100 bp, respectively. The resulting sequences were stripped out from the primers and annotated at the genus or family level by RDP Classifier (Wang et al. 2007) using the RDP 16S rRNA database and an assignment confidence cutoff of 0.6. This signifies that 60% of K-mers will match at the genus level. Singletons were removed.

For ITS sequences, forward and reverse reads were merged using vsearch (Rognes et al. 2016). Sequences were quality filtered and assembled in QIIME pipeline (Caporaso et al. 2010). Chimeras were removed using UCHIME (Edgar et al. 2011) with closed reference and de-novo approach. OTUs (operational taxonomic units) were assembled at 97% similarity using vsearch clustering algorithm at default settings (Rognes et al. 2016) and blasted against UNITE 7.2 database (Kõljalg et al. 2013). Singletons were removed. Fungal OTUs were merged on genus level (i.e., 603 genera) for all analyses apart from alpha diversity, to account for species length polymorphisms in fungal ITS regions below genus level (Gomes et al. 2002).

Negative control bacterial genera resp. fungal OTUs from the kit and the sampling liquid were subtracted for all samples (see online supplementary material Table S10 and Fig. S1 for sequence statistics and rarefaction curves). Statistical analyses were done in *R* (R Core Team 2015) using the phyloseq (McMurdie and Holmes 2013), vegan (Oksanen et al. 2018) and ggplot (Wickham 2009) packages.

The alpha diversity Chao1 richness index S_1 was calculated as follows:

$$S_1 = S_{\text{obs}} + \frac{F_1^2}{2F_2}$$

where S_{obs} is the number of species in the sample, F_1 is the number of singletons, and F_2 is the number of doubletons.

The alpha diversity Shannon evenness-Index H' was calculated as follows:

$$H' = \sum_i p_i * \ln p_i \,\text{mit}\, p_i = \frac{n_i}{N}$$

with p_i being the relative abundance of a certain species i of the total number of individuals N and n_i being the absolute number of individuals belonging to one species.

The data set was run both rarefied and untreated, which did not show differences in statistical analyses, but removed rare genera. To include otherwise removed genera and to present the data characteristics appropriately, relative abundance analysis was carried out on non-filtered data (Weiss et al. 2017). The fungal dataset was third-root transformed. Bray–Curtis distances for bacteria and fungi were calculated on datasets normalized to relative abundance (Weiss et al. 2017) and ordinated with non-metric multidimensional scaling (NMDS). The samples May_T12, May_5, May_T6 and May_T12 had to be excluded from the ITS analyses, as they had too few reads to calculate NMDS.

Pairwise PERMANOVA was conducted using the package "pairwiseAdonis" (Martinez Arbizu 2017); Bonferroni p value correction was applied as default for multiple corrections. Statistical parameters of ANOSIM (999 permutations), ADONIS (999

permutations), Kruskal–Wallis, Dunns PostHoc and PERMANOVA analysis are reported in the online supplementary data (Tables S5–S9).

SIMPER (Similarity Percentage), describing the species contribution of Bray–Curtis distance between two groups, was calculated in vegan (supplementary Tables S12–S18).

Obtained air and precipitation samples were blasted (blast version 2.7.1) against a nucleotide database of organisms containing the ice nucleation protein (search for "ice nucleation protein"), created from ncbi (www.ncbi.nlm.nih.gov), with a threshold of 97% similarity. The obtained reads were filtered for matches longer than 400 base pairs and merged on genus level.

2.5 Microbial enumeration

2.5.1 16S rRNA gene qPCR

The bacterial cell concentration was approximated by the number of 16S rRNA gene copies per cubic meter of air. The V3 region of the 16S rRNA gene was quantified using the Quantifast 2X SYBR Green dye (QIAGEN Venlo, Netherlands) and the following primer sequences: Eub 338f 5′-ACTCCTACGGGAGGCAGCAG-3′ as the forward primer and Eub 518r 5′-ATTACCGCGGCTGCTGG-3′ as the reverse primer (Øvreås and Torsvik 1998) on a Rotor-Gene 3000 machine (QIAGEN, Venlo, Netherlands). The reaction mixture of 20 µl contained 10 µl of SYBR master mix, 2 µl of DNA and RNAse-free water to complete the final 20 µl volume. The qPCR two-step program consisted of an initial step at 95 °C for 2 min for enzyme activation and then 35 cycles of 5 s at 95 °C and 20 s at 60 °C hybridization and elongation. A final step was added to obtain a denaturation from 55 °C to 95 °C with increments of 1 °C s^{-1}. The amplicon length was around 200 bp. PCR products obtained from DNA from a pure culture of Escherichia coli were cloned in a plasmid (pCRtm2.1-TOPO®-vector, Invitrogen, Carlsbad, USA) and used as a standard quantification with the Broad-Range Qubit Fluorometric Quantification (Thermo Fisher Scientific, Waltham, USA). Non-template controls were subtracted.

2.5.2 18S rRNA gene qPCR

The fungal cell concentration was approximated by the number of 18S rRNA gene copies per cubic meter of air. The region located at the end of the SSU 18S rRNA gene, near the ITS1 region, was quantified using the Quantifast 2X SYBR Green dye (QIAGEN, Venlo, Netherlands) and the following primer sequences: FR1 5′-AICCATTCAATCGGTAIT-3′ as the forward primer and FF390 5′-CGATAACGAACGAGACCT-3′ as the reverse primer (Chemidlin Prévost-Bouré et al. 2011) on a Rotor-Gene 3000 machine (QIAGEN, Venlo, Netherlands). The reaction mixture of 20 µl contained 10 µl of SYBR master mix, 2 µl of DNA and RNAse-free water to complete the final 20 µl volume. The qPCR two-step program consisted of an initial step at 95 °C for 5 min for enzyme activation and then 35 cycles of 15 s at 95 °C and 30 s at 60 °C hybridization and elongation. A final step was added to obtain a denaturation from 55 °C to 95 °C with increments of 1 °C s^{-1}. The amplicon length was around 390 bp. PCR products obtained from DNA from a soil sample were cloned in a plasmid (pCRtm2.1-TOPO®vector, Invitrogen, Carlsbad, USA) and used as a standard quantification with the Broad-Range Qubit Fluorometric Quantification (Thermo Fisher Scientific, Waltham, USA). Non-template controls were subtracted.

2.6 Meteorology

Meteorological and aerosol data were obtained from the Sonnblick Observatory repository. An aerosol backscatter profile from VAISALA CL51 ceilometer located at Kolm-Saigurn (the base of Mount Sonnblick at 1500 m asl.), and PBL heights were obtained from ZAMG (Zentralanstalt für Meteorologie und Geodynamik, Vienna). The mixing layer heights were calculated according to Lotteraner and Piringer (2016) (see online supplementary information Figs. S3–S6). The lagrangian model FLEXPART (Stohl et al. 2002) was used in backward mode to detect the most probable source regions for microorganisms which may have been transported with long-range air flows. The model output fields depict the regions with high source–receptor sensitivity due to the prevailing three-dimensional flow fields predicted by the ECMWF model for 120-h backcalculation (see online supplementary material Fig. S6 for model

Table 2 Mean meteorological values ± standard deviation for the sample periods derived from hourly values (*source*: ZAMG)

	Temperature (°C)	Humidity (%)	Precipitation (mm)	total Precipitation (mm)	Wind speed (m/s)	Wind Direction (dd)	maximum Mixing height (m AGL)	mean PBL (m AGL)	Pressure (hPa)
May	− 6.90 ± 1.48	92.67 ± 4.79	0.10 ± 0.42	12.30	7.2 ± 3.22	197.49 ± 88 (SSW)	1607	716 ± 453	692.3 ± 1.64
August	8.50 ± 2.90	91.47 ± 11.07	0.73 ± 1.81	70.30	6.17 ± 2.33	198.67 ± 84.94 (SSW)	2438	748 ± 667	704.6 ± 1.54
November	− 12.80 ± 2.81	89.45 ± 3.01	0.16 ± 0.41	17.10	7.5 ± 3.04	188.18 ± 106.91 (S)	819	349 ± 188	682.2 ± 5.22
February	− 14.40 ± 3.52	90.69 ± 1.45	0.00 ± 0.03	0.90	9.13 ± 1.88	111.25 ± 64.16 (ESE)	766	261 ± 138	681.6 ± 1.58

output). Mixing layer height calculation and FLEXPART backward calculation are described in the supplementary information of Els et al. (2019).

3 Results

3.1 Meteorology

Atmospheric temperatures were lowest during February and highest in August, with highest mean relative humidity in May (compare Table 2, compare Supplement Figure S3 for respective meteorological conditions at each sampling time).

Based on backward dispersion modeling results, air masses originated from different regions during each of the seasonal sampling campaigns (Supplement Figure S6). Samples collected in May were free tropospheric air mainly of maritime origin transported from the Atlantic over France to the Alpine region. Some transport from the Mediterranean region may have influenced the first sample collected on May 2, 2017. In August, the main source regions were identified as the Mediterranean Sea and North Africa and at the end of the sampling period, additional contributions from air masses traveling over Western Europe and the eastern Atlantic were observed. In November, the free tropospheric air was tracked back to the British Island and Iceland, even up to the northern Polar region. The sampling on November 11, 2017, was influenced by air that had been transported over Northern Spain, the Riviera and the Po Valley. In February, the air masses reaching the observatory during the sampling days had been influenced by flow from most of Europe and the Mediterranean area.

In the first days of May 2017, the core of a low-pressure system was centered over the Eastern Alpine region, resulting in cloudy conditions and strong snow fall in the morning of May 2, 2017 as well as in the night of May 4 to 5, 2017. In August, intermittent precipitation was observed during the sampling periods T1, T5 and T13 (Table 1). On August 3 to 6, 2017, a series of frontal passages dominated the meteorological conditions at Sonnblick Observatory. Heavy precipitation, thunderstorms and hail occurred during sampling, and mean wind speeds were lowest as compared to the other measurement periods.

For the period between November 27 and 30, 2017, weak pressure gradients dominated the weather

conditions in the Alpine area, resulting in low temperatures. In February, no precipitation was observed. While low-pressure gradients prevailed during the first day of this period, a cyclone over the Mediterranean caused an increase in wind speeds up to 14.8 ms^{-1} at the Sonnblick Observatory on February 21, 2018.

Samples collected in May, November and February were entirely in the free troposphere (Supplement Figure S4). A deep convective boundary layer reached above the mountain top on August 3 and 4, 2017, and the mixing heights determined from the ceilometer measurements indicated that the boundary layer air remained below the observatory on August 5 and 6, 2017.

Stable stratification of the valley boundary layer, characterized by low mixing heights, prevented significant transport of valley air to the mountain top in February and November. Mixing heights were low during these periods, most of the time below 400 m AGL (compare Supplement Figures S4 and S5).

3.2 Seasonality of bacterial and fungal communities

3.2.1 Abundance

The median of 16S rRNA gene abundance per m^3 air was highest in November (1.14×10^6 m^{-3} air) and lowest in February (2.67×10^5 m^{-3}air), with the biggest range (Coefficient of Variation CV: 2.47) and highest (1.67×10^7 m^{-3}air) and lowest values (1.18×10^5 m^{-3}air) in May. The smallest variation coefficient was found in August (0.58). Dunn pairwise comparison revealed significant difference for all groups except between May and February ($p = 0.054$) and May and August ($p = 0.44$) (Fig. 2a).

The highest median and absolute maximum, but lowest variation of 18S rRNA gene abundance per m^3 air was found in November (median: 5.1×10^4 m^{-3}air, max: 9.98×10^4 m^{-3}air, CV: 0.65), the lowest median and absolute value, but highest variation in May (median: 1.0×10^2 m^{-3}air, min: 6.13×10^0 m^{-3}air, CV: 2.45). Dunns pairwise comparison revealed significant difference between all groups but February and August ($p = 0.069$) (Fig. 2a).

The lowest median of the 16S to 18S rRNA gene ratio was found in November (3.86×10^1 16S:18S m^{-3}air). May had the highest median ratio (4.02×10^3 16S:18S m^{-3}air) and maximum value (2.28×10^5 16S:18S m^{-3}air), but also highest variation (2.31). The lowest ratio and variation coefficient was found in August (9.74×10^0 16S:18 m^{-3}air, CV: 0.72). Dunns pairwise comparison revealed significant differences between all but August and February ($p = 0.14$) (Fig. 2a). The seasonal abundance of 16S rRNA genes, 18S rRNA genes and their ratio was significantly different (Kruskal–Wallis $p < 0.001$).

3.2.2 Alpha diversity

Bacterial and fungal read counts per sample ranged from 13478 to 41882 and 568 to 34638, respectively, after quality filtering. The whole dataset comprised 1590 bacterial genera and 2573 fungal OTUs, which accounted for 963 bacterial genera and 1912 fungal OTUs (resp. 603 fungal genera) after removal of blanks.

The overall median and mean of bacterial Chao1 (describing richness, Fig. 2b) were 48, with the lowest median in August (40.75) and the highest in November (70.06), and the lowest value in August (19.33) and the highest in February (160.00). November differed significantly from May ($p = 0.023$) and August ($p = 0.0368$), whereas in February significant differences to August were detected ($p = 0.0368$, Kruskal–Wallis $p = 0.03$). The highest variation was found in May (0.61) and the lowest in November (0.34).

The overall median of Shannon (describing evenness, Fig. 2b) alpha diversity index for bacteria was 2.37 and the mean 2.31, ranging from 1.99 (February) to 2.61 (May). The lowest value was found in November (0.25) and the highest in May (3.26). Variation was highest in November (0.39) and lowest in August (0.11). Shannon diversity differed significantly in February from all other months (Feb-May $p = 0.0006$, Feb-Aug $p = 0.0008$, Feb-Nov $p = 0.0247$, Kruskal–Wallis $p < 0.001$).

The highest fungal richness (median (48.5) and maximum value (81)) was observed in August, while the lowest value was recorded for May (median (3) and absolute value (1), variability (0.36)). Variability was highest in February (0.81). All values were significantly different from each other (Kruskal–Wallis $p < 0.001$) apart from November and February (Dunns $p = 0.3053$).

The highest fungal evenness (Shannon) median (2.89) and absolute value (3.65) was recorded in

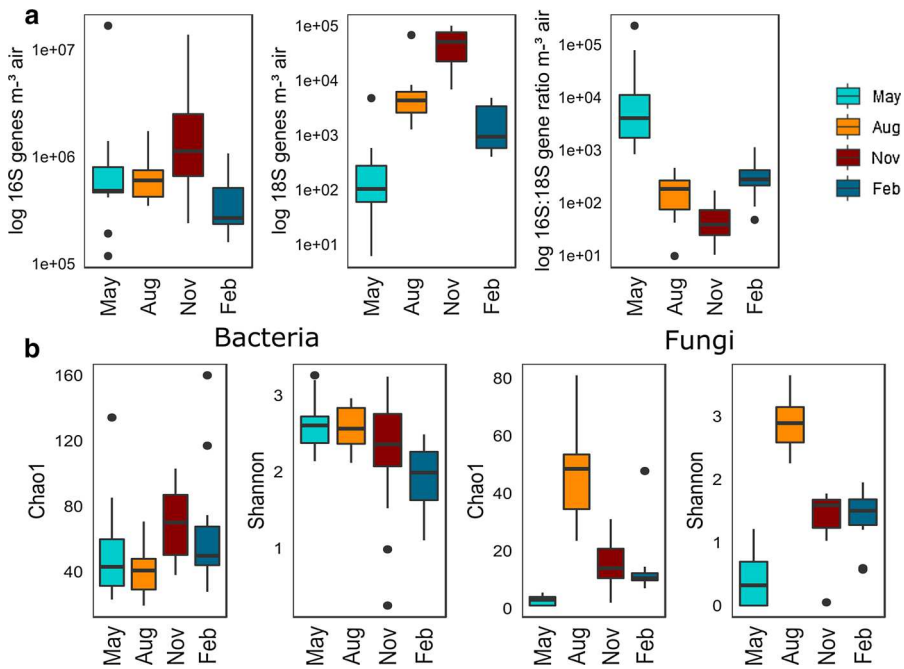


Fig. 2 **a** Abundance of 16S, 18S and 16S to 18S ratio m^{-3} air per season, **b** alpha diversity measures Chao1 and Shannon for bacteria and fungi

August, and the lowest median occurred in May (0.32) and in November (0.05).

November and February did not differ significantly ($p = 0.4647$), but all other groups did (Kruskal–Wallis $p < 0.001$). Variability was highest in May (1.10) and lowest in August (0.15).

3.2.3 Beta diversity

Seasonal air bacterial communities were significantly different (ADONIS $p = 0.001$, $R^2 = 0.166$, ANOSIM $p = 0.001$, $R = 0.114$). They differed significantly between February and May (PERMANOVA adj.p = 0.006), February and August (PERMANOVA adj.p = 0.006), May and November (PERMANOVA adj.p = 0.006) and August and November (PERMANOVA adj.p = 0.006). February samples were the most variable and distinct as compared to the other communities. The February and November (winter) and May and August (summer) bacterial communities did not differ significantly. On phylum level, winter samples were most abundant in Firmicutes (Nov: 78.4%, Feb: 48.1%), whereas summer samples were abundant in Proteobacteria (May: 54.3%, Aug 49.4%), Bacteroidetes, Acidobacteria and Planctomycetes (Fig. 3c).

The highest value of unique air bacterial genera was reached in May, the lowest in August.

61 genera, i.e., 11% of the total bacterial genera, were common to all seasons (Fig. 3b). The most abundant of these common bacteria belonged to the genera *Geobacillus* (Bacilli), *Aminobacter* (Alphaproteobacteria) and *Schlesneria* (Planctomycetia) (see online supplement Tables S1 and S2).

Of all bacterial air genera, 54% were unique to one of the seasons. The most abundant genera only occurring in May and August were *GP2* (Acidobacteria), *Domibacillus* (Bacilli) and *Paludibacter* (Bacteroidia). Most abundant unique genera in winter samples were *Bifidobacterium* (Actinobacterium), *Vulcaniibacterium* (Gammaproteobacteria) and *Polaribacter* (Flavobacteria). SIMPER analysis (see supplement Tables S12 and S13) revealed *Geobacillus* abundance in February and November as the driver of differences between summer and winter samples. Further, high abundances in *Microbacterium*, *Tumebacillus* in winter and presence of *Aminobacter*, *Bosea*, *Burkholderia* and *Schlesneria* in summer differentiated samples.

Air fungal seasonal communities also significantly differed (ADONIS $p = 0.001$, $R^2 = 0.3075$, ANOSIM $p = 0.001$, $R = 0.5225$). All seasons differed from each other (PERMANOVA p.adj = 0.006, respectively); however, NMDS of Bray–Curtis distance (Fig. 3a) revealed that May samples were most dispersed, while August samples were most similar

and February and November samples had the highest overlap.

August samples were dominated Agaricomycetes, Lecarnomycetes and Leotiomycetes, while May fungal communities were characterized by Eurotiomycetes, Saccharomycetes, Tremellomycetes and Wallemiomycetes. Winter samples had a high amount of unidentified organisms, November was characterized by high abundance in Sordarimycetes (23.2%) and Tremellomycetes (12.1%), while February featured many Agaricomycetes (34.6%).

Only 3% of the identified fungal air genera were common in all seasons. The most abundant were unidentified Ascomycota, unidentified Agaricomycetes, *Formitopsis* (Agaricomycetes), *Aspergillus* (Eurotiomycetes) and *Trichaptum* (Agaricomycetes). The highest share of unique genera was found in August (53.9% of all identified fungal genera), with *Teloschistaceae* (Lecarnomycetes), *Gymnopus* (Agaricomycetes) and *Resupinatus* (Agaricomycetes) and range of Agaricomycetes identified as the most abundant. February contained the second most fungal genera detected in air (31.7% of all fungal genera detected), with *Exidia* (Agaricomycetes), *Talaromycetes* (Eurotiomycetes) and *Taphrina* (Taphrinomycetes) as most abundant unique genera (supplementary. Table S2). SIMPER analysis (supplement Table S16) revealed *Phanerochaete*, *Formitopsis*, *Aspergillus* and *Wallemia* as driving fungal genera distinguishing May samples from all other seasons. August had characteristically high abundances of *Sacrogyne*, *Acarospora* and *Formitopsis*, while *Mycena*, *Saccharomyces*, *Extremus* were characteristic in February, and *Exophilia*, *Filobasidium* for November.

3.3 Abundance and diversity of bacterial and fungal communities in air and precipitation

3.3.1 Comparison of abundance in air and precipitation

The median of 16S rRNA gene abundance m^{-3}air (5.23×10^5) is within one order of magnitude of the abundance per mL cloud (1.17×10^5 mL^{-1}) or snow (2.66×10^5 mL^{-1}). Rain water (2.29×10^7 mL^{-1}) and hail (1.88×10^7 mL^{-1}) showed two orders of magnitude higher values (Fig. 4).

Minimum values of 16S rRNA genes mL^{-1} in precipitation were reached in snow (1.87×10^2 mL^{-1}); maximal concentration of 16S rRNA genes was present in rain (4.28×10^7 mL^{-1}). Variation was lowest in cloud (0.94) and higher in snow (1.67), but highest in air (2.18).

Air and precipitation 16S rRNA gene abundances were significantly different (Kruskal–Wallis $p = 0.001$), with cloud and air (p.adj. = 0.0012) and cloud and rain (p.adj. = 0.0004) differing strongly. Further hail vs. rain and snow vs. air, cloud, hail and rain differed in unadjusted testing.

The median of 18S rRNA gene abundance m^{-3}air (3.01×10^3 m^{-3}air) was lower than in cloud (8.55×10^4 mL^{-1}) and snow (2.02×10^4 mL^{-1}), much higher in hail (1.46×10^7 mL^{-1}) and lower in rain (7.72×10^2 mL^{-1}).

16S to 18S ratio (Kruskal–Wallis $p < 0.001$) differed in air and cloud (p.adj = 0.0001), snow and air (p.adj. = 0.0002), and rain and snow (p.adj. = 0.0161). The highest median 16S:18S ratio occurred in rain (3.18×10^4), and a higher 16S:18S ratio in air (2.29×10^2) than in cloud (1.15×10^0), snow (2.81×10^1) and hail.

18S rRNA gene abundance in precipitation (Kruskal–Wallis $p = 0.01$) differed only in snow and air (p.adj. = 0.0375), but further in air and cloud, cloud and rain, hail and rain and rain and snow in unadjusted testing (supplement Tables S5, S8, S9).

3.3.2 Alpha diversity in air and precipitation

Chao1 for bacteria differed in air and cloud (p.adj. = 0.0027), air and snow (p.adj. = 0.0000), rain and snow (p.adj. = 0.0038) and cloud and rain (p.adj. = 0.0254) as well as overall (Kruskal–Wallis $p < 0.001$).

Bacterial Shannon evenness (Kruskal–Wallis $p < 0.001$) differed in air and cloud (p.adj. = 0.0535), air and snow (p.adj. = 0.0000), hail and snow (p.adj. = 0.0464).

Fungal Chao1 (Kruskal–Wallis $p < 0.001$) differed in cloud and air (p.adj. = 0.0000) and air and snow (p.adj. = 0.0000), further in cloud and rain, hail and rain and rain and snow in unadjusted testing. Fungal Shannon (Kruskal–Wallis $p < 0.001$) differed as well in air and cloud (p.adj. = 0.0090) and snow and air (p.adj. = 0.0001) and further in rain and cloud and rain and snow in unadjusted testing.

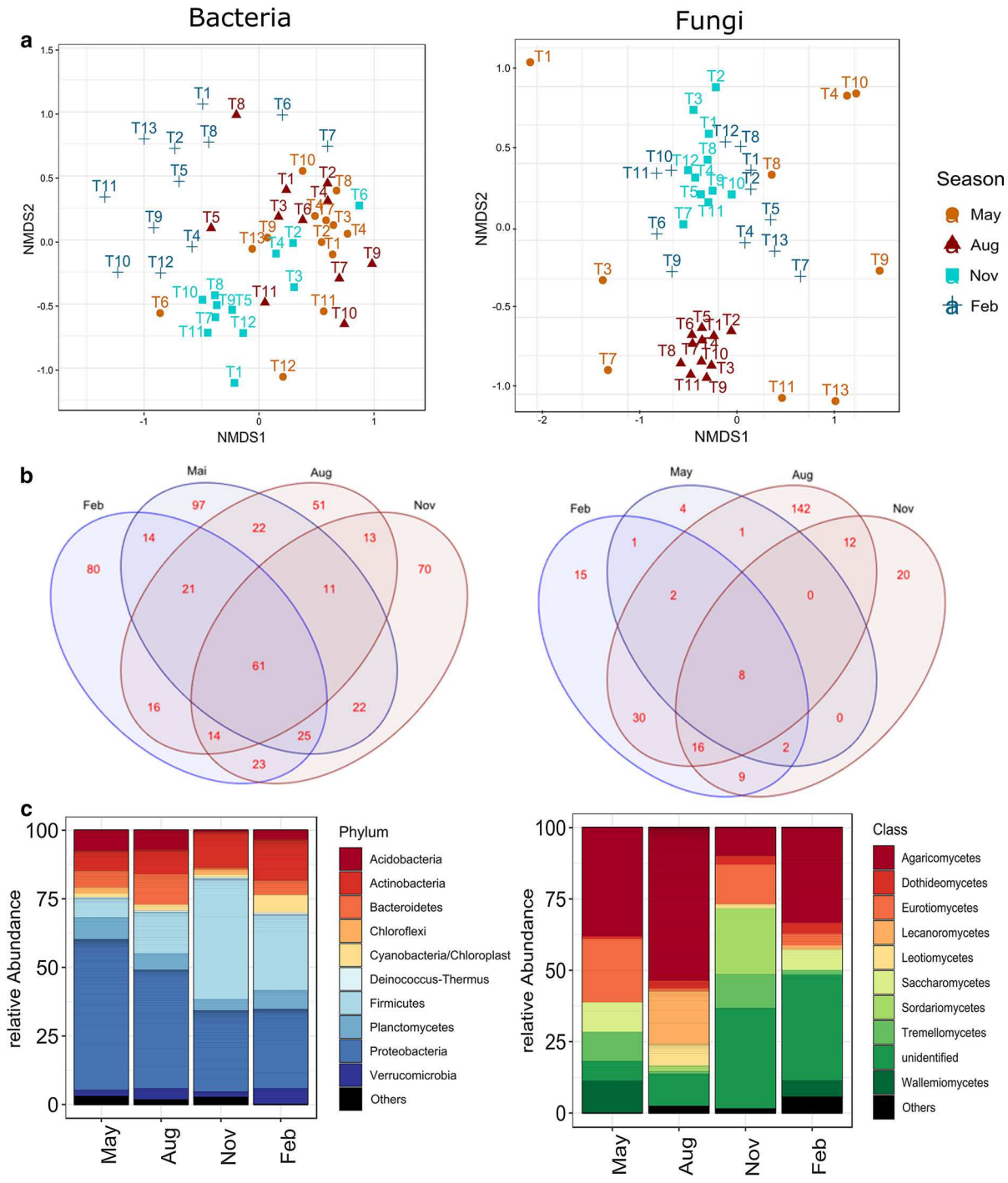


Fig. 3 **a** NMDS on Bray–Curtis distance on genus for bacteria and fungi for air composition by season, **b** Venn diagram of unique and common bacterial and fungal genera by season, **c** relative abundance (%) of top 10 most abundant bacterial phyla and fungal classes

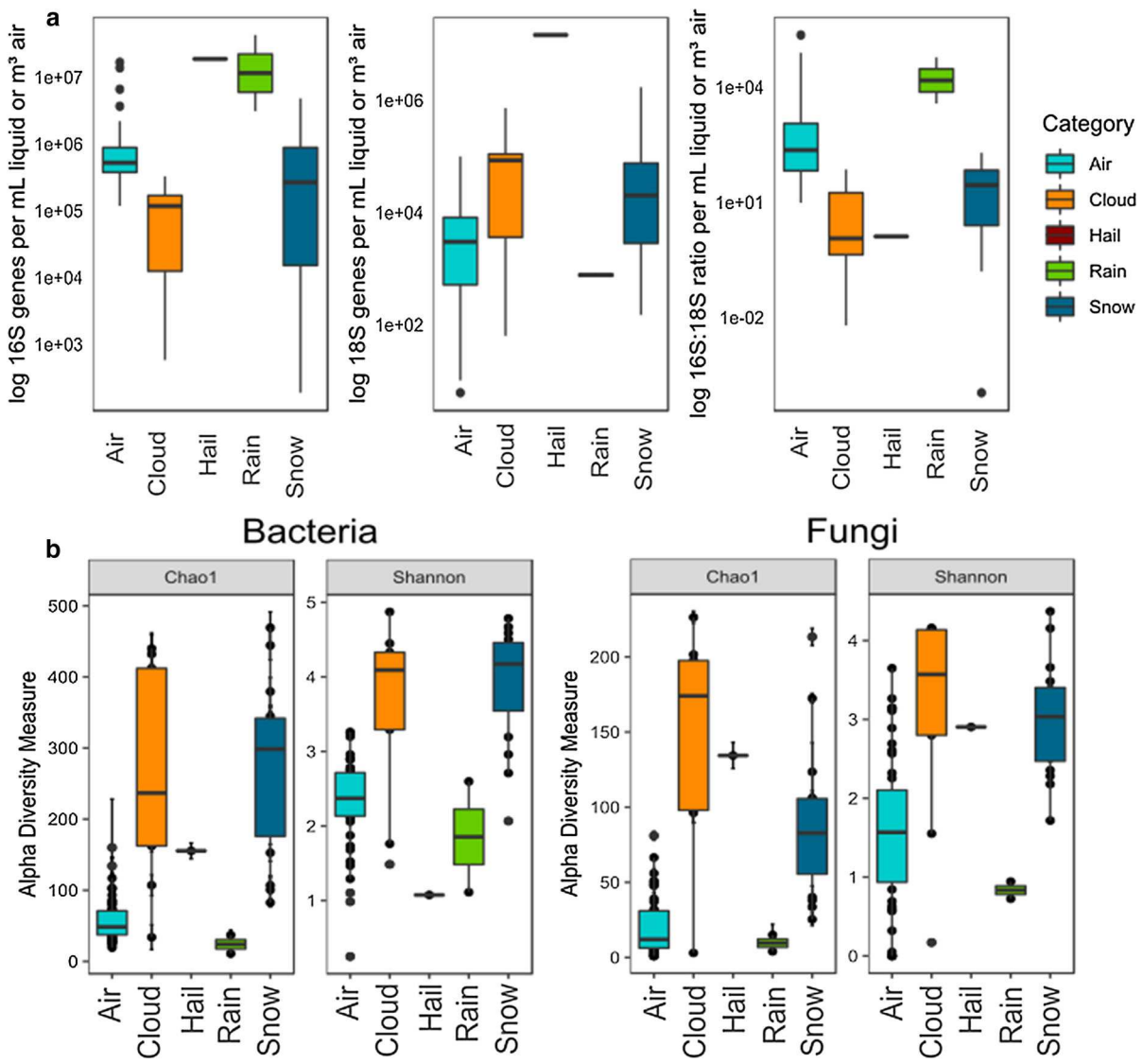


Fig. 4 **a** 16S rRNA genes, 18S rRNA genes and 16S:18S ratio m^{-3}air and mL^{-1} precipitation, **b** alpha diversity indices in air and precipitation for bacteria and fungi

3.3.3 Beta Diversity in air and precipitation

NMDS (Fig. 5a) revealed a strong clustering of bacterial communities of precipitation, with rain and one May cloud sample being most distinct. The February bacterial air samples are the most distinct from precipitation.

August fungal samples cluster closest together in air samples and closest to precipitation. Fungal precipitation samples cluster more densely than any season for air, apart from August rain samples and one May cloud sample, which are more distinct. Fungal February and November air samples strongly overlap, while May samples are most scattered.

Bacterial air samples are significantly different from precipitation samples (ADONIS $R^2 = 0.1118$, $p = 0.001$, ANOSIM $R = 0.3004$, $p = 0.001$). Air and snow resp. cloud bacterial communities were significantly different (air–snow, air–cloud: PERMANOVA p.adj = 0.01), whereas air and rain did not differ significantly (PERMANOVA p.adj = 0.84), but hail did in unadjusted testing (PERMANOVA unadj.p = 0.022, p.adj = 0.22). Snow was significantly different from rain (PERMANOVA unadj.p = 0.025).

Also fungal air communities differed significantly from precipitation (ADONIS $R^2 = 0.1076$, $p = 0.008$, ANOSIM $R = 0.0697$, $p = 0.001$). Air and snow resp. cloud fungal communities differed significantly (air–snow, air–cloud: PERMANOVA p.adj = 0.01). Fungi in air and rain differed in single testing (PERMANOVA unadj.$p = 0.046$, adj.$p = 0.46$). Snow and rain fungi were significantly different (PERMANOVA adj.$p = 0.04$).

When compared pairwise for each season, bacteria in air were most similar to snow, rain and hail in August (PERMANOVA air–snow p adj. = 0.096, air–rain p adj. = 0.156, air–hail p adj. = 0.552) and snow and cloud in May (PERMANOVA air–snow p adj. = 0.054, air–cloud p adj. = 0.048). Cloud, snow, rain and hail were never significantly different, when compared against each other within one season. In the total dataset, rain and snow showed the biggest difference (PERMAOVA p non-adj. = 0.025).

Snow and cloud in cold seasons (November and February) appear similar in bacterial composition, with Acidobacteria (3.2–6.2%), Actinobacteria (18.6–40.7%), Bacteriodetes (5.5–22.0%) and Cyanobacteria (13.4–20.8%) being abundant, whereas air is dominated by Firmicutes (48.1%- 78.4%). Proteobacteria make up a comparable share in air, cloud and snow in winter (11.2–27.2%). In May, air and cloud are more similar with high abundances in Proteobacteria (54.3% in air, 51.1% in cloud), while May snow features high abundances of Actinobacteria (53.8%) and Firmicutes (19.0%), while cloud and snow in May feature Cyanobacteria (7.8% in snow, 8.6% in cloud) that are very low in May air (0.9%, Fig. 6).

Air in August features also high numbers of Proteobacteria (49.4%) that are even higher in hail (88.8%). August snow is dominated by Bacteriodetes (47.4%), Proteobacteria (27.1%) and Actinobacteria (20.1%), with Firmicutes lacking, and the lowest abundances of Planctomycetes (0.6%), Chloroflexi (0.01%) and Cyanobacteria (2.7%) compared to snow from the other seasons.

Fungi show a similar pattern, with rain and snow being most different in the overall dataset (PERMANOVA p.adj. = 0.04) and precipitation never significantly differing within season comparison. Highest similarity occurred between air and snow in May (PERMANOVA p adj. = 1.0), but also all precipitations in August and did not differ significantly to air (PERMANOVA snow p adj. = 0.102, rain p adj. = 0.066, hail p adj. = 0.468). Further, air and cloud in February were only significantly different in unadjusted testing (PERMANOVA p non-adj. = 0.028, p adj. = 0.084).

In fungi, the composition of cloud and snow is similar in winter samples, but different to air. In November, air is dominated by unidentified (35.5%), Sordariomycetes (23.2%) and Eurotiomycetes (14.1%). In February, air is dominated by Agaricomycetes (34.6%) and unidentified (39.9%). Clouds and snow in winter were dominated by Leotiomycetes (9.8% cloud, 16.8% snow) and Agaricomycetes in November (27.4% snow, 26.2% cloud), Dothideomycetes (9.9–35.8%), Eurotiomycetes (15.8–23.9%), Lecarnomycetes (9.2–26.5%), further unidentified (9.4–26.8%, Fig. 6).

In May, air featured in Wallemiomycetes (11.1%) and Saccharomycetes (10.4) that rarely occurred in any other air or precipitation sample and May cloud samples featured Tremellomycetes (36.4%) and Dothideomycetes (9.8%).

Snow in both summer samples was high in unidentified organisms. Rain and air featured high abundance of Agariomycetes (59.9% in air, 55.5% in rain) in August. Rain was further rich in Sordariomycetes (37.2%), while hail was abundant in Dothideomycetes (36.5%) and Leotiomycetes (31.6%).

11.8% of bacterial genera and 9.7% of fungal genera were only found in the 288 m^3 of sampled air and not in any of the precipitation samples (Fig. 5b). The share of common bacterial genera (12.4%) and fungal genera (11.6%) was similar. The ratios of unique and common bacterial and fungal genera between precipitation types display a comparable picture, with the biggest %-differences in shared genera between cloud, snow and air in bacteria (21.7%) compared to 13.3% in fungal genera, unique cloud fungi (15%) compared to unique cloud bacteria (9.1%), and bacteria shared between cloud and snow (18.2%) compared to fungi in cloud and snow (21.5%).

Most abundant in all sample types were *Geobacillus* (Firmicutes), *Aminobacter* (Proteobacteria) and *Herpetrichiellaceae* (Eurotiomycetes), *Fomitopsis* (Agaricomycetes) and

Hypholoma (Agaricomycetes) (Supplementary Tables S3 and S4).

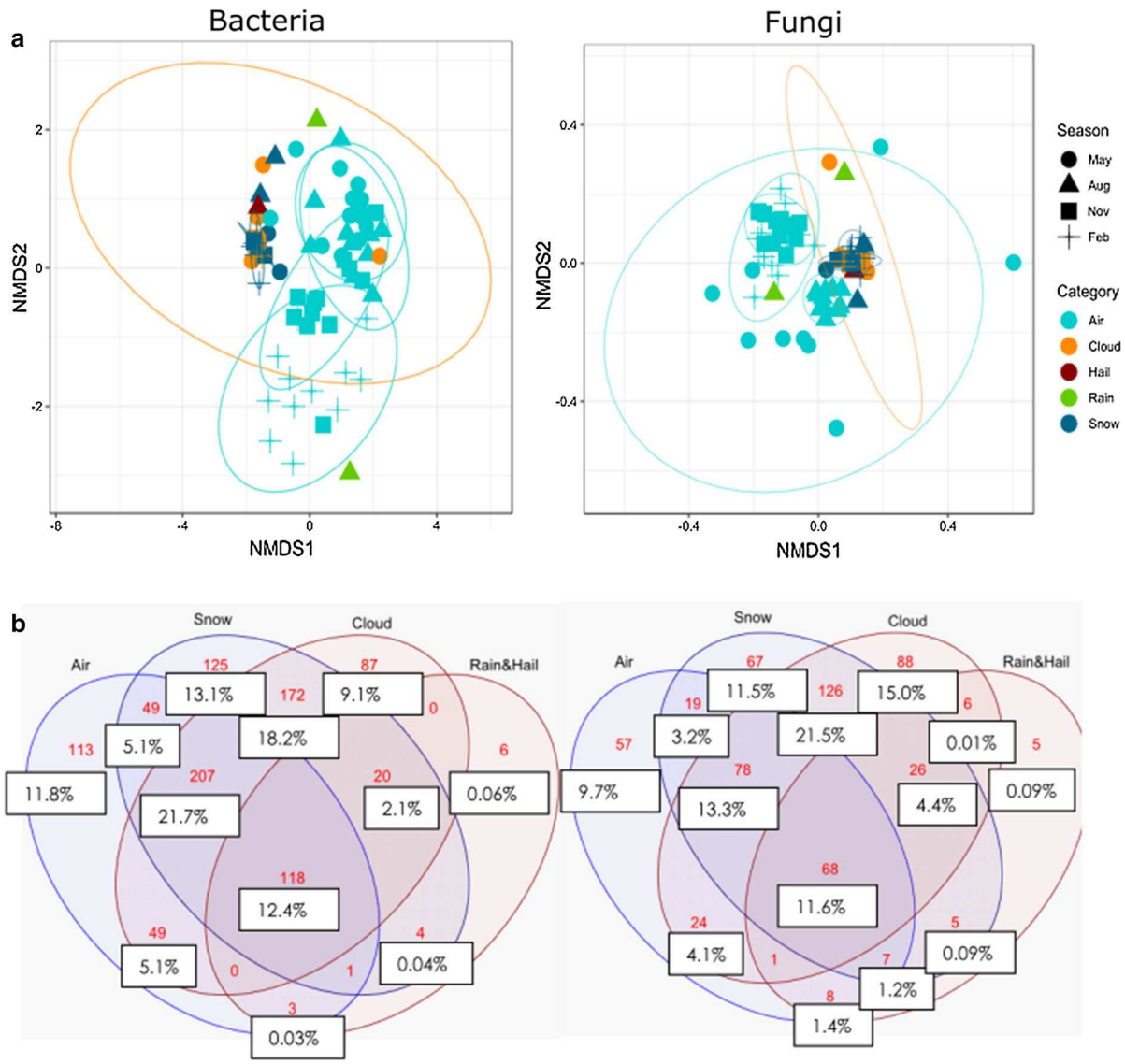


Fig. 5 **a** NMDS of Bray–Curtis distance of air and precipitation, **b** Venn diagram of common and unique bacterial and fungal genera

SIMPER analysis (supp Tables S1, S14, S15, S17, S18) revealed that the genera mostly driving differences between air and snow are *Geobacillus* and *Aminobacter* that are more abundant in air, and *Chlorophyta*, *Nakamurella* and *Ferruginibacter* that are more abundant in snow, on phylum level Firmicutes were dominant in air compared to Actinobacteria, Bacteriodetes and Cyanobacteria in snow.

The difference between cloud and air was driven by *Geobacilli*, *Bosea* and *Aminobacter* that were more abundant in air and *Janthinobacterium* and *Chlorophyta* that were more abundant in clouds, on phylum level Firmicutes were much more abundant in air, whereas Actinobacteria and Cyanobacteria were more abundant in clouds.

In rain, *Gp1* and *Dehalospirillum*, *Nakamurella* and *Aquisphaera* were more abundant than in air, and Actinobacteria and Acidobacteria were more abundant at the phylum level in rain than in the air.

In hail, *Janthiobacterium*, *Cytophagales* and *Subtercola* drove differences with air, and Proteobacteria was the dominant phylum as compared to Firmicutes, Actinobacteria and Planctomycetes for air samples.

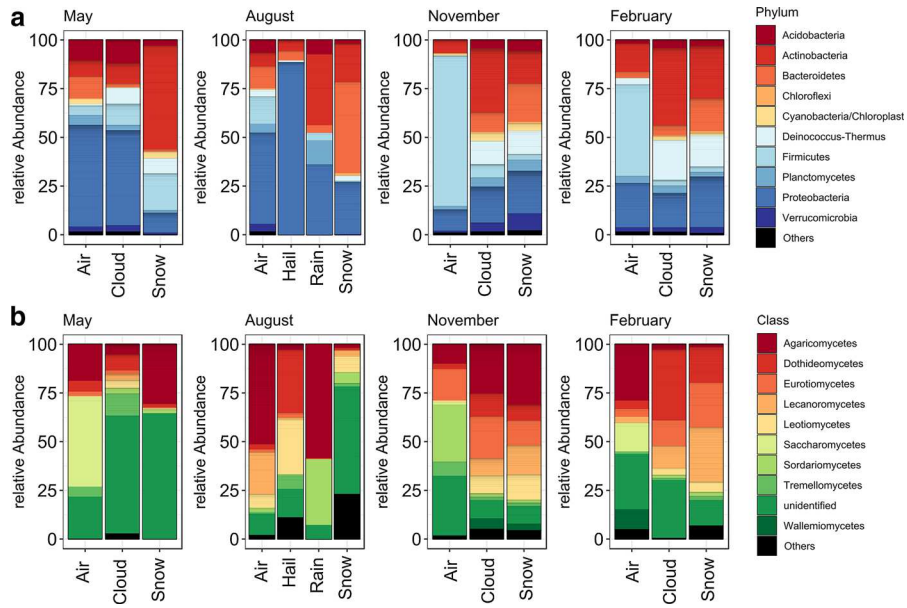


Fig. 6 Top 10 **a** bacterial phyla and **b** fungal classes for air and precipitation by season

Most abundant unique fungal genera in air were *Stereaceae* (Agaricomycetes), *Curvularia* (Dothideomycetes) and *Corticium* (Agaricomycetes), while most abundant unique genera in snow comprised *Phaeosclera* (Dothideomycetes) *Lactiflus* (Agaricomycetes) and *Phaerophyscia* (Lecarnomycetes), and *Phaeomycocentrospora* (Dothideomycetes), *Byssoloma* (Lecarnomycetes) and *Rhodosporidiobolus* (Microbotryomycetes) in clouds.

For fungi, SIMPER analysis was conducted on class level, as too many unidentified taxa occurred on lower levels. Agaricomycetes were abundant in air compared to snow and cloud, whereas Eurotiomycetes and Dothideomycetes were more abundant in snow and cloud. Lecarnomycetes in snow and Tremellomycetes in cloud also drove differences with air.

Agaricomycetes and Sordariomycetes were more abundant in rain than in air, while Eurotiomycetes, Tremellomycetes and Lecarnomycetes drove the difference of air to rain. In hail, the biggest difference to air was dominated by the presence of Agaricomycetes, Eurotiomycetes and Sordariomycetes in air together with Dothideomycetes, Leotiomycetes and Arthoniomycetes in hail.

3.4 Ice nucleation-active bacteria

In total, there were 53 genera containing the ice nucleation protein (INP) detected in the dataset. In air samples, the relative abundance of INP reads of the total reads per sample was lowest in November (1.56%) and highest in August (5.28%), but with high variation. In clouds, relative abundance of IN varied strongly between seasons displaying the highest value in the dataset (15.82%), whereas in May accounted for only 2.25% and in February 4.16%. In snow the mean IN abundance varied between 5.22% in May and 11.51% in August. Hail ranged at 7.04% and rain between 0 and 11% (Fig. 7a). The composition relative abundance of the most abundant INP-containing bacteria showed a similar pattern than the composition of most abundant phyla (discussed in Chapter 3.3, Fig. 6), with snow and cloud being similar in February and November, and more similarity of air and cloud in May, and high variability in August. (Fig. 7b)

NMDS (Fig. 8a) reveals that cloud and air have a higher overlap, than snow and air, with cloud at the interface between air and snow. The two rain samples are different in IN-composition. Hail clusters with snow and cloud. Most of the 53 identified genera containing the ice nucleation protein, present in the dataset, were associated with snow. The occurrence of ice nucleation-active reads is more similar in the frozen water samples (hail, snow) to the clouds, than air. However, the Venn diagram (Fig. 8b) reveals, that the most genera are shared between all of the samples

or air, snow and cloud, but by NMDS, indicating, that they are more abundant in snow and cloud.

The genera associated with higher abundance in air were *Corynebacterium, Pantoea, Enterobacteria, Micrococcus, Caulobacter, Paenibacillus, Acidovorax. Xanthomonas* and *Erwinia* were unique to air.

4 Discussion

4.1 Seasonal differences in air community composition

Seasonal variation and temporal vicinity of sample composition have been observed for near-surface airborne bacterial and fungal communities (Bowers et al. 2013; Fahlgren et al. 2010; Franzetti et al. 2011; Pickersgill et al. 2017), and even above PBL for bacteria (Bowers et al. 2012). These observations are consistent with our data, which showed that there was no stable free troposphere air microbial community during the year of sampling at Mount Sonnblick, although a core microbiome could be identified consisting of 61 (11% of all air genera) bacterial and 8 fungal genera (3% of all fungal air genera). Many of these have previously been identified as indicator species for above PBL air compared to below PBL air (Els et al. 2019). Air mass origin was different for the four sampling periods. Differential source regions and transport have been shown to influence microbial composition of the atmosphere (DeLeon-Rodriguez et al. 2013; Innocente et al. 2017). Zweifel et al. (2012) also suggested annual variation, succession of species in airborne communities or random variation as explanatory mechanisms. One mechanism that might impact microbial composition in air communities is thermal convection and air mass mixing in the boundary layer. The lowest amount of unique bacterial genera was found in August, but the highest number of unique genera was found in May. In August the thermal convection and mixing was most active, with the highest mean boundary layer elevation and variation of the sampling dates (748 ± 667 m AGL). The height of PBL-mixing even reached up to the observatory for a short moment (between T3 and T4, compare supplement data Fig. S4), thus it is likely that

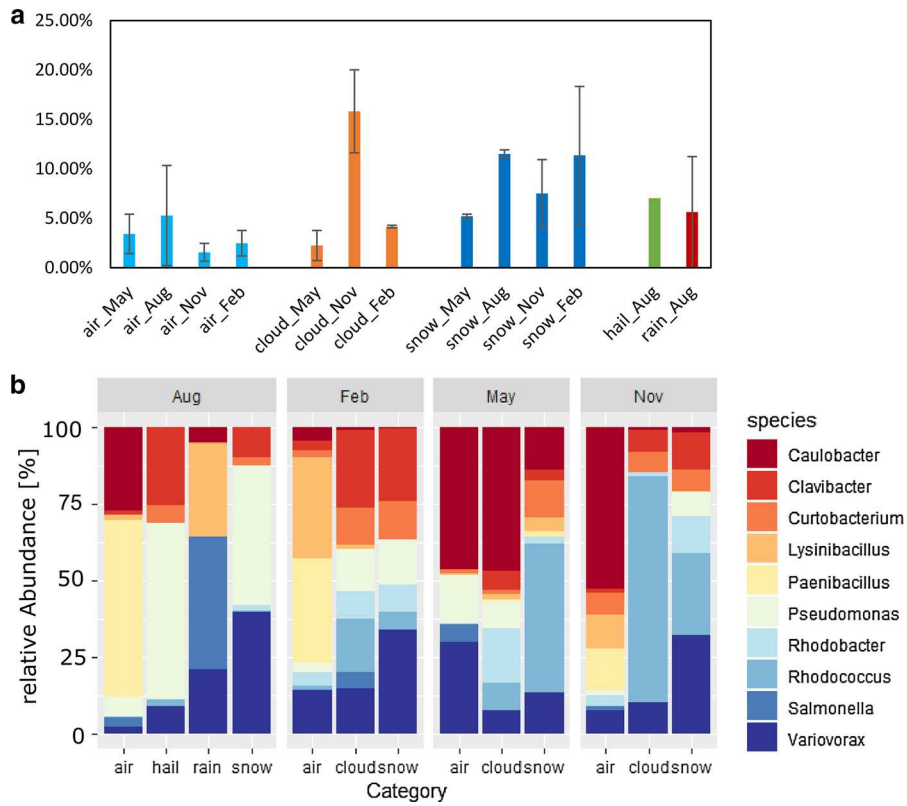


Fig. 7 **a** Relative abundance (%) of IN-active bacteria of the total number of 16S reads, **b** relative abundance of 10 most abundant IN-active genera

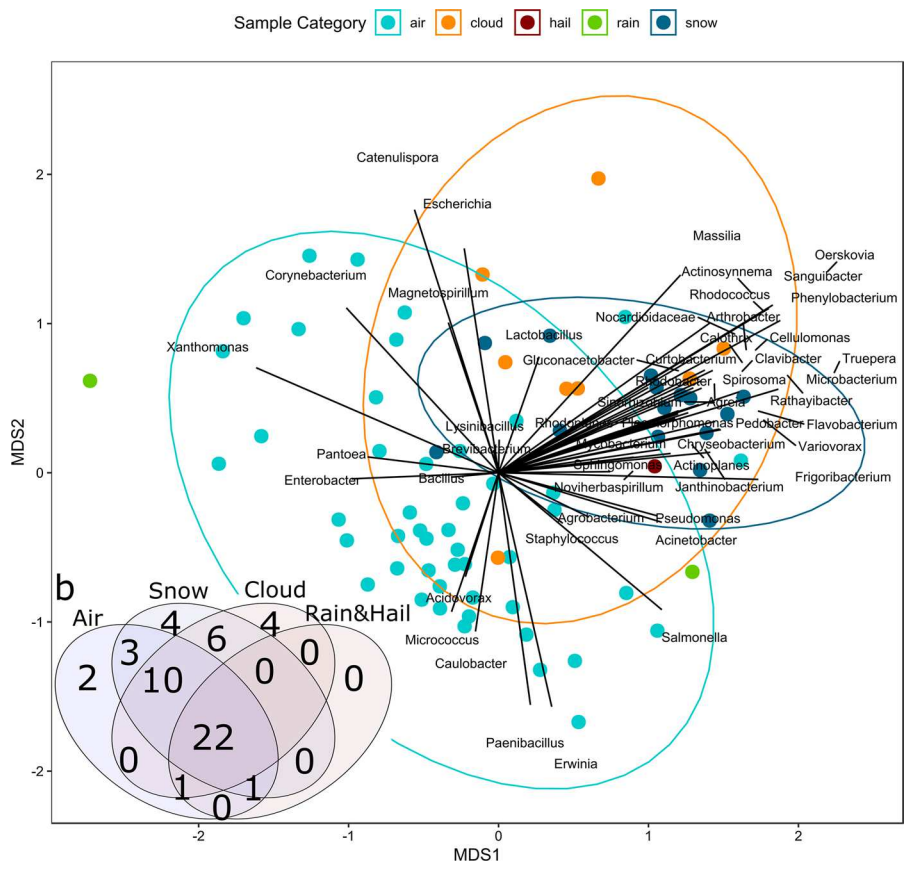


Fig. 8 **a** NMDS on Bray–Curtis distance of ice-nucleating communities contained in air, cloud, snow, rain and hail, **b** Venn diagram of common and unique genera, containing the ice nucleation protein

more mixing took place. August was also the time of strong precipitation events, highest humidity and highest temperature, thus highest thermodynamic capacity of air (Hänel 1976). In May, the PBL did not reach the sampling height, but PBL dynamics were more pronounced than in winter, due to higher temperatures and thus higher thermal convection processes.

4.2 Seasonal changes in microbial abundance

16S and 18S rRNA gene abundances lie within reported ranges of free tropospheric air and have been discussed in Els et al. (2019). 16S rRNA gene abundance showed highest variability in May, and highest 16S:18S ratio, indicating a filtering of fungi. High bacterial loads in spring and fall were previously reported at high elevations sites (Bowers et al. 2012). The overall 18S rRNA gene abundance was highest in November together with the lowest 16S:18S ratio. Fungal occurrences are strongly coupled to seasons and meteorological factors, depending on their life-style and sporulation strategy, and usually show high variations in air and cloud water (Pickersgill et al. 2017; Amato et al. 2007a, b). In May the valley was completely covered by fresh snow (50 cm before sampling), and in November, the first snowfall on vegetation in the valley occurred on the day before sampling at the end of the growing season. Hence, when soils were dry, and particles are not bound strongly, together with less dense canopy and bare fields, aerosolization might be effectuated. One possibility for the high fungal abundance despite ongoing snowfall might be the impeded scavenging by snow crystals due to fungal hydrophobins. These proteins occur uniquely in mycelial fungi such as Ascomycetes or Basidiomycetes (Wösten 2001), which have been detected in the respective samples. Hydrophobins are crucial in a specific life stage of those fungi to grow aerial hyphae (Wösten et al. 1999). In August the fungal diversity was highest, as fungal sporulation is most active in summer, also thermal mixing is

stronger, thus more spores were likely transported higher up in the atmosphere and injected to the free troposphere.

4.3 Different seasonal dispersal patterns for bacterial and fungal communities

Bacterial and fungal communities exhibited different patterns in changes in community structure throughout the year. Bacterial community structure differed significantly between summer and winter. Summer samples were dominated by ubiquitous, terrestrial associated genera like *Paludibacter* (Bacteriodetes), a gram-negative non-spore former involved in NH_4^+ oxidation (Zhang et al. 2018), *Domibacillus* (Firmicutes), a gram-positive, spore-forming, aerobic, non-motile, rod-shaped organism isolated from deep-sea sediment (Sun and Sun 2016), ocean sediment (Sharma et al. 2014), mangrove sediment soil and macroalgae (Verma et al. 2017) and *GP1* (Acidobacteria), found in alpine grassland on the Tibetian Plateau (Yuan et al. 2014). Winter samples were more characterized by extremophile but omnipresent bacteria like *Bifidobacterium* (Actinobacteria) that can withstand a high range of pH (from 8 to 2) (Charteris et al. 1998), *Vulcaniibacterium,* a thermotolerant Gammaproteobacteria isolated from a geothermally heated soil (Yu et al. 2013), and *Polaribacter* (Bacteriodetes), which has been found in several arctic habitats (Cuthbertson et al. 2017; Harding et al. 2011; Collins et al. 2010). Differences in the winter to the summer microbiome (in SIMPER analysis) were driven by changes in abundance in Firmicutes such as *Geobacillus* and *Tumebacillus*. *Geobacillus* is a known thermo-resistant, spore-forming, ubiquitous Firmicute (Zeigler 2014). *Tumebacillus* (aerobic, spore-forming, gram-positive, rod-shaped, chemolithoautotroph) has been found in high Arctic Canadian permafrost (Steven et al. 2008) and was identified as one of the most abundant lineages in Australian dry lake aerosols (Munday et al. 2016) and heavy-haze days in Beijing, China (Yan et al. 2018). Firmicutes are recurrently characterized as highly abundant in free tropospheric air masses and long-range transport air masses (Smith et al. 2011, 2012; David et al. 2018; Els et al. 2019) and were enhanced in free tropospheric fall samples (Bowers et al. 2012). They are known to withstand hostile, even extraterrestrial conditions, like heat, cold, radiation or chemicals, by forming resistant endospores and are found in all investigated environments (Setlow 2006; Driks and Eichenberger 2016; Nicholson et al. 2000).

Fungi displayed a high seasonal variability with a range of different genera and phyla dominating the fungal composition at the different seasons. Sporulation strategy, dissemination adaption and lifestyle-dependent patterns are critical aspects for airborne fungal occurrence (Pickersgill et al. 2017). 53.9% of all fungal genera were unique to August air masses. These samples were dominated by a large group of lichenized fungi (Miadlikowska et al. 2014) and the largest non-lichenous fungal group Leotiomycetes (Wang et al. 2006), whereas xerophilic and psychrophilic Eurotiomycetes (Geiser et al. 2006), highly resistant, yeast forming Saccharomycetes (Mühlhausen and Kollmar 2014), yeast forming Tremellomycetes (Liu et al. 2015) and xerophilic Wallemiomycetes (Zalar et al. 2005) were dominant in May. Tremellomycetes were also abundant in November, together with plant pathogenic, endophytic, saprophytic, epiphytic and lichenolous Sordariomycetes (Maharachchikumbura et al. 2016). In February, a time when potentially low sporulation happens, air abundant Agaricomycetes dominated. High wind speed in February yielded the second highest number of genera occurring in one season (31.7%) and the communities were dominated by Agaricomycetes, which is ground and plant associated and known to form propagules being dispersed by the air and mainly connected to wet deposition (Hibbett et al. 2014; Prasher 2015). Additionally, compared to the other phyla, they are characterized by a relative small aerodynamic diameter which enables the spores to avoid faster sedimentation by dry deposition (Woo et al. 2018) and usually contribute a big share of airborne fungi (Pickersgill et al. 2017; Fröhlich-Nowoisky et al. 2012). However, it has to be taken in to account that the choosen molecular primer and amplification method can result in an over- or underrepresentation of certain fungal classes (e.g., here highly abundant Agaricomycetes) or phyla (Blaalid et al. 2013).

4.4 Comparison of community structure and abundance in air and precipitation

Microorganisms precipitated with snow, hail or rain or being transported by clouds differ in their species

composition from free tropospheric air masses and do not mirror the air community structure. They are distinct in composition, 16S:18S ratio, abundance and diversity from free-floating PBA. Clouds and snow generally had the highest diversity in both bacterial and fungal datasets, with atmosphere and rain being the least diverse.

Cloud bacterial communities were dominated by Proteobacteria, Actinobacteria and Bacteriodetes, which has also been observed in several cloud water studies from China, France, southern Africa and the USA (Jiaxian et al. 2019; Amato et al. 2017; Kourtev et al. 2011; Zhu et al. 2018; Evans et al. 2019). Fifteen of overall identified fungal genera were unique to clouds, whereas only 9% of bacterial genera were unique to clouds. The most abundant unique fungal cloud genera were *Phaeomycocentrospora* (Dothideomycetes), a large cosmopolitan genus of plant-pathogenic leaf and fruit fungi (Crous et al. 2013). Further abundant unique cloud genera were *Byssoloma* (Lecarnomycetes), a lichen with pantropical spread and scattered temperate localities that was only recently first described in Sweden (Hermansson and Thor 2004), and *Rhodosporidiobolus* (Microbotrymycetes), which was found in arctic air and is associated with elevated phospholipase and hemolytic activity (Kirtsideli et al. 2017). Microbotrymycetes contain a range of psychrophilic yeasts. Thus, for those abundant unique cloud genera, their likely source and part of their dissemination strategy are long-range transport.

In air as compared to snow, cloud, hail and rain especially differences in Bray–Curtis distance were driven on genus level by *Geobacillus* (Firmicutes) and *Aminobacter* (Alphaproteobacteria), an organism found in various terrestrial environments (McDonald et al. 2005) that supports legume nodulation (Maynaud et al. 2012) and is capable of herbicide degradation (Schultz-Jensen et al. 2014). A further driving organism was *Bosea* (Alphaproteobacteria) an amoeba-resisting bacteria which is associated with hospital water supply and intensive care ventilator-acquired pneumonia (Khamis et al. 2003; La Scola et al. 2003) and rhizosphere (Ahern et al. 2007). On phylum level, air was especially enhanced in Firmicutes. Hail was the only precipitation form not yielding significant differences to air at the respective season in unadjusted testing for bacteria and fungi in our dataset.

The bacterial abundance of fresh deposited snow, graupel or rain accounted for approximately 1×10^4 mL^{-1} or higher in various samples (Amato et al. 2007a, b; Bauer et al. 2002; Carpenter et al. 2000; Sattler et al. 2001; Segawa et al. 2005) and 10^3–10^4 m^{-3} for bacterial density in air at remote sites (Bauer et al. 2002; Burrows et al. 2009a, b). This implies a higher density of microbes in precipitation at a range of $1{:}10^6$. PBA are more abundant in clouds than in the free atmosphere, and despite harsh conditions in clouds they might be more life-supporting due to the presence of water (Delort et al. 2017; Bauer et al. 2002). Clouds can highly vary in PBA abundance, with clouds formed during extreme events like desert dust uplift, containing higher than expected abundances (of 10^5 m^{-3} bacteria like particles) (Maki et al. 2017). Further, the median of 16S to 18S ratio was lowest in clouds, suggesting a higher capacity of cloud doplets to hold fungi in clouds in comparison with air, snow or rain. The result of lowest ratio of bacteria to fungi in clouds is in accordance with Bauer et al. (2002), whose counts result in a bacteria-to-fungi ratio of 4.4 for clouds, 5 for snow and 16.4 for air. Bacteria could provide a surface area for the condensation of water vapor on the surface of aerosolized particulate matter (Jiaxian et al. 2019). In rain, 16S:18S ratio was highest, suggesting a separation mechanism for fungal spores before rain forms.

Rain was high in 16S rRNA gene abundance but low in 18S abundance and in bacterial and fungal diversity. Air and rain were not significantly different in total bacterial air dataset, but in unadjusted testing in August dataset, fungi were significantly different in the total dataset.

Rain was the only precipitation form that yielded a significant difference when compared with another precipitation form in the total dataset, but it was not significant within the August dataset. However, the snow here was likely more influenced by melting, aging and aerial deposition. Rain was different to bacterial diversity in snow (unadjusted) and more significantly to fungi in snow.

Hail had the lowest 16S: 18S ratio, which might be due to the formation from cloud droplets that have also low 16S:18S ratio, where spores might possibly get stuck to ice surface during sudden freezing.

4.5 Seasonal effects related to temperature

A significant separation of air masses and precipitation samples was observed, and differences varied between seasons. In the winter seasons (November and February), the composition was similar in cloud and snow, but different to air, whereas in May bacterial composition of cloud and air was more similar, but pronouncedly different from snow. In May, the community of fungi in clouds featured high numbers of cold-adapted yeast and lichenous Tremellomycetes (Liu et al. 2015; Thomas-Hall et al. 2010) and plant-pathogenic Dothideomycetes (Schoch et al. 2009; Hane et al. 2007) that were less abundant in air or snow. These results indicate that cold environmental conditions lead to comparable bacterial and fungal composition of snow and cloud, but different composition of air, while in the warmer conditions of May, air and cloud water were more similar in bacterial composition. Here we found the highest number of shared genera of cloud and air in one season (8.6%) and highest total number of unique cloud genera (20.7%) and outstandingly high amount of unique fungi in May cloud (63.6%), with snow more different and a distinct fungal cloud composition. This indicates that warmer water vapor (but still negative ambient temperature in May around − 6.9 °C) might have higher momentum and exchange or even promote a distinct composition by its living conditions. Also the composition of ice nucleation-active bacteria was similar in air and cloud in May with higher abundances of *Pseudomonas*, than in the other seasons, that can efficiently nucleate ice at a temperature of − 7 °C or lower (Maki et al. 1974).

Bacteria in air were more similar to precipitation in August (especially hail and air) than in the other months. This could be due to higher temperatures and humidity, facilitating more thermodynamic momentum leading to exchange and incorporation of humid air into cloud formation, similar life conditions in humid air and fast, on-site precipitation rather that distant transport.

August snow was lacking Firmicutes and had a low abundance of Chloroflexi, Planctomycetes and Cyanobacteria compared to snow from the other seasons. Here melting and accumulation processes might have played a role, as the snowpack was percolated with melt and rain water, and sediment form the molten snow pack and dry deposition was accumulated on the surface.

August fungal air samples were the closest to precipitation fungal samples. This might be due to the high abundance of fungi in August, but further also the higher humidity during this time, as fungal species actively discharge spores via liquid jets into the air, preferentially under humid conditions (Pringle et al. 2005; Elbert et al. 2007) and fungal variability being associated with humidity and rain frequency (Bowers et al. 2013). Fungal composition was shown to be significantly different in wet and dry deposition, suggesting taxon and size specific involvement in cloud condensation and precipitation (Woo et al. 2018).

4.6 Selection processes and potential drivers

The differences observed in community structure and abundance among the different sample types might be related to physical processes such as scavenging. In general, scavenging depends on the aerodynamic diameter, particle density but also chemical composition. Microbial cells can show hydrophilic or hydrophobic membranes which can alter their deposition substantially (Maria and Russell 2005). Cell surface hydrophobicity is often found in gram-negative bacteria (Krasowska and Sigler 2014). Rain or snow precipitation removes below-cloud aerosol particles and is a key process in aerosol chemical transport (Zhang et al. 2013). The scavenging coefficients or removal rates of particles are complex functions (Slinn 1983), but theoretical and experimental studies have shown that snow is a more efficient scavenger of particles than rain because of its larger surface area (Franz and Eisenreich 1998). The relative scavenging efficiency between snow and rain depends on particle sizes and precipitation intensity (Wang et al. 2014). For ultrafine and fine particles, the snow scavenging coefficient is predicted to be $\sim$ 10 times larger than the rain scavenging coefficient at low precipitation rates (Wang et al. 2014). For rain, scavenging efficiency decreases with increasing particle size (Ardon-Dryer et al. 2015).

Clouds, fog and rain seem to represent media, where biological activity is significant due to the protection from desiccation and the potential use of chemical compounds in the aqueous solution (Deguillaume et al. 2008). Some bacteria can act as cloud

condensation nuclei and produce biosurfactants that increase droplet size and cloud lifetime, counteract desiccation and allow water scavenging and thus facilitate widespread dispersal (Ahern et al. 2007).

PBA originate from surface aerosolization and thus represent lager potential sources than precipitation. Stricter selective processes are occurring for PBA to reach clouds, survive and be precipitated.

Rain does not scavenge too many organisms, as the residence time in the atmosphere is lower in rain. Raindrops have a smaller surface and are potentially less active in atmosphere interactions. Snow, in contrast, acts more like a sponge.

Hail is a unique form of precipitation that allows direct characterization of particles present during atmospheric ice nucleation, as the nucleation event is captured in the hail stone embryo (Michaud et al. 2014). The embryo is then caught by updraught and cycling in a storm cloud while accreting supercooled water during hailstone formation and thus represents a vertical intersection of cloud microbial composition (Knight and Knight 2001; Michaud et al. 2014). Šantl-Temkiv et al. (2013) observe an enrichment in plant-associated bacterial groups (Gammaproteobacteria, Sphingobacteria, Methylobacteria). Hail is one of the most extreme environments to live (Šantl-Temkiv et al. 2013).

4.7 INP-containing bacteria are selected for precipitation

Another mechanism that might explain variation in community structure is the capacity of certain organisms to carry out ice nucleation (IN). IN-particles active at ≥ -10 °C deplete twice as rapidly in precipitating clouds than other particles at this size (Stopelli et al. 2015).

There was a higher share of relative % reads of known ice nucleation-active bacteria present in all wet phases compared to air. Within the wet phases, snow had the hightest mean % reads, whereas clouds showed a big variation thereof. It seems that a separation of IN-active reads occurred with higher shares occurring in precipitation.

Regarding seasonality, within air masses, the highest share of IN-active reads was reached in August, when also strongest precipitation occurred and samples were taken during rain events. IN-occurrence has been reported to increase during and after rain events (Huffman et al. 2013).

August precipitation was characterized by short but strong showers (compare supplement data Fig. S3) which have been observed to be significantly enhanced in IN in the first period of high intensity rain and hail, compared to continuous-type rain (Vali 1971). Ice nucleation-active bacteria were found to be relatively more abundant in rain than in air at the same site (Stephanie and Waturangi 2011). Christner et al. (2008a, b) report cumulative ice nuclei over several seasons to reach higher maximum values in rain (Louisiana rain 8–230 IN L^{-1}, Alberta rain 20–600 IN L^{-1}) compared to snow (mid-latitude snow 3–150 IN L^{-1}). Cloud IN varied between 1 and 200 mL^{-1} at -10 °C at Puy de Dome (Joly et al. 2014), thus ranging one order of magnitude higher than in rain or snow.

Christner et al. (2008a, b) estimated 0.4% of DNA-containing cells in mid-latitude snowfall being IN-active at -7 °C to -4 °C, whereas in cloud water, estimates range between 0% and 1.5% IN-active bacteria at -10 °C of total bacteria (Joly et al. 2014). In air, numbers were reported as 0.001% (Garcia et al. 2012; Xia et al. 2013) and showed significantly higher numbers in foggy air than in clear air (Bowers et al. 2009; Xia et al. 2013). Petters and Wright (2015) describe a variation in IN-concentration between rain, cloud water, hail and snow of up to five orders of magnitude.

Abundant rainfall was reported to drive near-surface air IN-concentration with a positive correlation of rain intensity and IN-concentration, while strong snow fall had an attenuating effect on IN-concentration (Conen et al. 2017).

In comparison, the values we estimate here based on relative abundance of reads of INP containing bacteria yield substantially higher numbers of IN-bacteria in air, cloud and precipitation. This could impact terrestrial biodiversity, with microbes landing in new environments faster (Morris et al. 2014), as our results imply, when precipitating.

The observed separation for community composition between air, snow, cloud and wet precipitation was also observed for IN-active bacteria, with clouds seeming like a mediator between air and snow in IN-composition. Only a few genera were unique to one sample type, the discrimination by NMDS was due to changes in relative abundance between sample types,

which indicated an accumulation of most INP-containing bacteria in the wet phases, and only single species associated with air. *Erwinia* and *Xanthomonas* were unique in air in this dataset, though they are usually abundant in the environment, but also absent in other water-based IN-surveys (Du et al. 2017). Zweifel et al. (2012) found *Erwinia* and *Pantoea* in low frequencies, but recurringly in free tropospheric air samples.

The microbes present in clouds, snow and rain survived long-range transport, ice nucleation and cold cloud processes. Thus, it is likely that wet or snow deposited microbes will outperform their dry deposited counterparts in numbers and adaptions. As clouds often form far from the place where they precipitate and biological nucleators are frequently involved, their microbial composition is different from the present surface air microbiome.

4.8 Precipitation is no proxy for free troposphere air communities

Cáliz et al. (2018), who recently used snow as a proxy for air mass composition in the free troposphere, identified "airborne" Chytridiomycota as abundant in spring snow. The overall relative abundance of Chytridiomycota in all air samples was 0.01%, compared to 2.49% in snow and 6.15% in rain but only 0.13% in hail and 0.22% in clouds in our dataset, indicating scavenging snow and rain precipitation. Bacteriodetes, a phylum identified by Cáliz et al. (2018) as abundant in snow, was identified as the one of three phyla driving most differences between air and snow (5.67% air, 18.40% snow). While there is no doubt that microorganisms are aerosolized, incorporated in snow and then deposited (Harding et al. 2011), based on our dataset, selection at the phylum level occurs. Cyanobacteria were high in snow and cloud compared to air. They often contain a hygroscopic mucilaginous covering, that effects in density reduction and dynamic streamlining (Reynolds 2007), but might also absorb atmospheric moisture, why they become heavy and drop down (Sharma and Singh 2010). Air bacterial composition was found to be more similar to soil, decaying material, lichen and plant surface assemblages than to snow in a survey in Greenland, where snow featured higher abundances of Beta- and Gammaproteobacteria and Bacteriodetes, while air was richer in Alphaproteobacteria, Cyanobacteria, Acidobacteria and Planctomycetes (Šantl-Temkiv et al. 2018).

Fog is known to have enriched concentrations of bacteria and yeast CFUs up to two orders of magnitude compared to clear air conditions, but no changes for mold CFUs (Fuzzi et al. 1997). Moreover, air samples derive from restricted altitudes, whereas the cloud samples can be addressed as a composite of this respective height including a thought black box from above. The composition of cloud chemistry and metabolically adapted cloud PBA community varies depending on the cloud origin (Deguillaume et al. 2014; Amato et al. 2017).

Evans et al. (2019) found that Shannon alpha diversity was greater in fog than clear conditions. Fog was more shaped by local sources than clear air, but also contained more remote marine signatures, indicating good conditions for long-range travel in clouds (Evans et al. 2019).

SIMPER analysis identified *Janthinobacterium* (Betaproteobacteria) and *Chlorophyta* (Chloroplast) as genera driving the difference between air and cloud and especially abundant in hail. *Janthinobacterium* in cloud water was found negatively correlated with SO_2, O_3, and particulate matter (PM_{10} and $PM_{2.5}$) and positively correlated with K^+ and Mg^{2+} (Jiaxian et al. 2019), which main sources are soil dust in big cloud droplets (Zhu et al. 2018). *Janthinobacterium* has antifungal properties (Rubio et al. 2018) and was found in high mountain lakes and Antarctica (Ahern et al. 2007) and alpine cryoconite, and is known for its psychrotolerant properties (Kim et al. 2012). It does fit the continental background location of Mount Sonnblick, where low concentrations of aerosol pollution markers (like SO_2 or high PM) would be expected (Deguillaume et al. 2014).

Chlorophyta (i.e., microalgae) have previously been found in cloud samples (Urbano et al. 2011). The reason for that might also be the higher retention capacity in droplets than in air for bigger size PBA. The aerosolization of chlorophyta is especially correlated with relative humidity, but also with temperature, sunshine and wind speed (Sharma and Singh 2010).

5 Conclusion

Significant differences were found in air microbial composition during different seasons, with pronounced summer and winter distinction in bacteria and strong discrepancies between all sampling events for fungi.

The season, thus temperature and humidity conditions, impact on free tropospheric microbial air composition and abundance, and should be kept in mind when interpreting single flight results. The differentiation between seasons was stronger pronounced in air than in precipitation, with rain being most different and variable of precipitation types, indicating distinct forces driving microbial fate in the air.

Microorganisms precipitated with snow, hail or rain or being transported by clouds differ in their species composition from free tropospheric air masses and do not mirror the air community structure. They are more diverse, potentially more viable due to the availability of water and as shown in this paper, distinct in composition, 16S:18S ratio and abundance from free-floating PBA.

Hence, snow or cloud samples are no suitable proxies for free tropospheric air microbiome composition, as used by several studies, due to the findings that separation processes in aerosolization, transport and scavenging occur. The microbial composition of arriving precipitation or clouds represents only a part of the microbial air communities of the cumulative sources of origin.

Air can travel thousands of kilometers, but it is even more important to consider what is migrating in clouds and wet air masses, respectively, as the precipitation of PBA is more likely than sedimentation from clear air masses. Thus, precipitation might even have a bigger impact in ecosystem input due to higher microbe density when considering the ratio of cell numbers in precipitation compared to air being $1{:}10^6$.

To understand global connectivity of bioaerosols, microbial migration and inoculation sources, a microbial air sampling network would be necessary, where precipitation cannot be used as a proxy but rather has to go alongside with air analyses.

With ongoing change in global water circulation, water availability and precipitation patterns this might have implications for microbial dispersal to ecosystems. Further, more continuous aerobiological monitoring in the free troposphere will be needed to understand interacting factors and feedbacks.

Acknowledgements Open access funding provided by University of Innsbruck and Medical University of Innsbruck. We thank for great support of three students: Christina Schweighofer, Karoline Wieser, Tamara Schober; the whole team of the Laboratoire Ampère of the École Centrale de Lyon, Elke Ludewig, Christian Meier, Alexander Hieden, Christoph Lotteraner and the whole team of the Sonnblick Observatory of the ZAMG and most important Klaus Unterberger.

Funding Open access funding provided by University of Innsbruck and Medical University of Innsbruck. This work received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie Grant No. 675546.

Data availability Raw sequences were submitted and are currently under evaluation at NCBI BioProject database under Project ID PRJNA516816 (http://www.ncbi.nlm.nih.gov/bioproject/516816).

Compliance with ethical standards

Conflicts of interests The authors declare that they have no conflict of interest.

References

Adams, R. I., Bateman, A. C., Bik, H. M., & Meadow, J. F. (2015). Microbiota of the indoor environment: A meta-analysis. *Microbiome, 3*(October), 49. https://doi.org/10.1186/s40168-015-0108-3.

Ahern, H. E., Walsh, K. A., Hill, T. C. J., & Moffett, B. F. (2007). Fluorescent pseudomonads isolated from Hebridean cloud and rain water produce biosurfactants but do not cause ice nucleation. *Biogeosciences, 4*(1), 115–124. https://doi.org/10.5194/bg-4-115-2007.

Alsved, M., Holm, S., Christiansen, S., Smidt, M., Ling, M., Boesen, T., et al. (2018). Effect of aerosolization and drying on the viability of pseudomonas syringae cells. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2018.03086.

Amato, P., Hennebelle, R., Magand, O., Sancelme, M., Delort, A.-M., Barbante, C., et al. (2007a). Bacterial characterization of the snow cover at Spitzberg, Svalbard. *FEMS*

Microbiology Ecology, 59(2), 255–264. https://doi.org/10.1111/j.1574-6941.2006.00198.x.

Amato, P., Joly, M., Besaury, L., Oudart, A., Taib, N., Moné, A. I., et al. (2017). Active microorganisms thrive among extremely diverse communities in cloud water. *PLoS ONE, 12*(8), e0182869. https://doi.org/10.1371/journal.pone.0182869.

Amato, P., Ménager, M., Sancelme, M., Laj, P., Mailhot, G., & Delort, A.-M. (2005). Microbial population in cloud water at the Puy de Dôme: Implications for the chemistry of clouds. *Atmospheric Environment, 39*(22), 4143–4153. https://doi.org/10.1016/j.atmosenv.2005.04.002.

Amato, P., Parazols, M., Sancelme, M., Mailhot, G., Laj, P., & Delort, A.-M. (2007b). An important oceanic source of micro-organisms for cloud water at the Puy de Dôme (France). *Atmospheric Environment, 41*(37), 8253–8263. https://doi.org/10.1016/j.atmosenv.2007.06.022.

Ardon-Dryer, K., Huang, Y.-W., & Cziczo, D. J. (2015). Laboratory studies of collection efficiency of sub-micrometer aerosol particles by cloud droplets on a single-droplet basis. *Atmospheric Chemistry and Physics, 15*(16), 9159–9171. https://doi.org/10.5194/acp-15-9159-2015.

Bauer, H., Heinrich, G., Regina, H., Kasper-Giebl, A., Reischl, G., Zibuschka, F., et al. (2003). Airborne bacteria as cloud condensation nuclei: Bacteria as cloud condensation nuclei. *Journal of Geophysical Research: Atmospheres*. https://doi.org/10.1029/2003JD003545.

Bauer, H, Kasper-Giebl, A. Löflund, M. Giebl, H. Hitzenberger, R., Zibuschka, F. & Puxbaum, H. (2002). The contribution of bacteria and fungal spores to the organic carbon content of cloud water, precipitation and aerosols. *Atmospheric Research*. In 2nd international conference on fog and fog collection, *64* (1–4), pp. 109–19. https://doi.org/10.1016/S0169-8095(02)00084-4.

Bianco, A., Voyard, G., Deguillaume, L., Mailhot, G., & Brigante, M. (2016). Improving the characterization of dissolved organic carbon in cloud water: Amino acids and their impact on the oxidant capacity. *Scientific Reports, 6*(November), 37420. https://doi.org/10.1038/srep37420.

Blaalid, R., Kumar, S., Nilsson, R. H., Abarenkov, K., Kirk, P. M., & Kauserud, H. (2013). ITS1 versus ITS2 as DNA metabarcodes for fungi. *Molecular Ecology Resources, 13*(2), 218–224. https://doi.org/10.1111/1755-0998.12065.

Bowers, R. M., Clements, N., Emerson, J. B., Wiedinmyer, C., Hannigan, M. P., & Fierer, N. (2013). Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. *Environmental Science and Technology, 47*(21), 12097–12106. https://doi.org/10.1021/es402970s.

Bowers, R. M., Lauber, C. L., Wiedinmyer, C., Hamady, M., Hallar, A. G., Fall, R., et al. (2009). Characterization of airborne microbial communities at a high-elevation site and their potential to act as atmospheric ice nuclei. *Applied and Environmental Microbiology, 75*(15), 5121–5130. https://doi.org/10.1128/AEM.00447-09.

Bowers, R. M., McCubbin, I. B., Hallar, A. G., & Fierer, N. (2012). Seasonal variability in airborne bacterial communities at a high-elevation site. *Atmospheric Environment, 50*(April), 41–49. https://doi.org/10.1016/j.atmosenv.2012.01.005.

Bowers, R. M., McLetchie, S., Knight, R., & Fierer, N. (2011). Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. *The ISME Journal, 5*(4), 601–612. https://doi.org/10.1038/ismej.2010.167.

Brown, J. K. M., & Hovmøller, M. S. (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science, 297*(5581), 537–541. https://doi.org/10.1126/science.1072678.

Brunet, Y., Wéry, N., & Galès, A. (2017). Short-scale transport of bioaerosols. *Microbiology of Aerosols*. https://doi.org/10.1002/9781119132318.ch2b.

Burge, H. A. (2002). An update on pollen and fungal spore aerobiology. *The Journal of Allergy and Clinical Immunology, 110*(4), 544–552.

Burrows, T., Butler, P., Jöckel, H., Tost, A., Kerkweg, U. Pöschl, & Lawrence, M. G. (2009a). Bacteria in the global atmosphere-part 2: Modeling of emissions and transport between different ecosystems. *Atmospheric Chemistry and Physics, 9*(23), 9281–9297.

Burrows, W. E., Lawrence, M. G., & Pöschl, U. (2009b). Bacteria in the global atmosphere-part 1: Review and synthesis of literature data for different ecosystems. *Atmospheric Chemistry and Physics, 9*(23), 9263–9280.

Cáliz, J., Triadó-Margarit, X., Camarero, L., & Casamayor, E. O. (2018). A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. *Proceedings of the National Academy of Sciences;*. https://doi.org/10.1073/pnas.1812826115.

Caporaso, J. Gregory, Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods, 7*(5), 335–336. https://doi.org/10.1038/nmeth.f.303.

Carotenuto, F., Georgiadis, T., Gioli, B., Leyronas, C., Morris, C. E., Nardino, M., et al. (2017). Measurements and modeling of surface-atmosphere exchange of microorganisms in Mediterranean grassland. *Atmospheric Chemistry and Physics, 17*(24), 14919–14936. https://doi.org/10.5194/acp-17-14919-2017.

Carpenter, E. J., Lin, S., & Capone, D. G. (2000). Bacterial activity in south pole snow. *Applied and Environmental Microbiology, 66*(10), 4514–4517.

Charteris, W. P., Kelly, P. M., Morelli, L., & Collins, J. K. (1998). Development and application of an in vitro methodology to determine the transit tolerance of potentially probiotic *Lactobacillus* and *Bifidobacterium* species in the upper human gastrointestinal tract. *Journal of Applied Microbiology, 84*(5), 759–768. https://doi.org/10.1046/j.1365-2672.1998.00407.x.

Christner, B. C., Morris, C. E., Foreman, C. M., Cai, R., & Sands, D. C. (2008a). Ubiquity of biological ice nucleators in snowfall. *Science, 319*(5867), 1214. https://doi.org/10.1126/science.1149757.

Christner, B. C., Rongman, C., Morris, C. E., McCarter, K., Foreman, C. M., Skidmore, M., et al. (2008b). Geographical, seasonal, and precipitation chemistry influence in the abundance and activity of biological ice nucleators in rain and snow. *Proceedings of the National Academy of Sciences of the United States of America, 105*(48), 18854–18859.

Clot, B. (2003). Trends in airborne pollen: An overview of 21 years of data in Neuchâtel (Switzerland). *Aerobiologia, 19*(3), 227–234. https://doi.org/10.1023/B:AERO.0000006572.53105.17.

Collins, R. E., Rocap, G., & Deming, J. W. (2010). Persistence of bacterial and archaeal communities in sea ice through an arctic winter. *Environmental Microbiology, 12*(7), 1828–1841. https://doi.org/10.1111/j.1462-2920.2010.02179.x.

Conen, F., Eckhardt, S., Gundersen, H., Stohl, A., & Yttri, K. E. (2017). Rainfall drives atmospheric ice-nucleating particles in the coastal climate of Southern Norway. *Atmospheric Chemistry and Physics, 17*(18), 11065–11073. https://doi.org/10.5194/acp-17-11065-2017.

Crous, P. W., Braun, U., Hunter, G. C., Wingfield, M. J., Verkley, G. J. M., Shin, H.-D., et al. (2013). Phylogenetic lineages in pseudocercospora. *Studies in Mycology, Phytopathogenic, 75*(June), 37–114. https://doi.org/10.3114/sim0005.

Cuthbertson, L., Amores-Arrocha, H., Malard, L. A., Els, N., Sattler, B., & Pearce, D. A. (2017). Characterisation of Arctic bacterial communities in the air above Svalbard. *Biology, 6*(2), 29. https://doi.org/10.3390/biology6020029.

Damialis, A., Vokou, D., Gioulekas, D., & Halley, J. M. (2015). Long-term trends in airborne fungal-spore concentrations: A comparison with pollen. *Fungal Ecology, 13*(February), 150–156. https://doi.org/10.1016/j.funeco.2014.09.010.

Deguillaume, L., Charbouillot, T., Joly, M., Vaïtilingom, M., Parazols, M., Marinoni, A., et al. (2014). Classification of clouds sampled at the Puy de Dôme (France) Based on 10 Yr of monitoring of their physicochemical properties. *Atmospheric Chemistry and Physics, 14*(3), 1485–1506. https://doi.org/10.5194/acp-14-1485-2014.

Deguillaume, L., Leriche, M., Amato, P., Ariya, P. A., Delort, A.-M., Pöschl, U., et al. (2008). Microbiology and atmospheric processes: Chemical interactions of primary biological aerosols. *Biogeosciences, 5*(4), 1073–1084. https://doi.org/10.5194/bg-5-1073-2008.

DeLeon-Rodriguez, N., Lathem, T. L., Rodriguez-R, L. M., Barazesh, J. M., Anderson, B. E., Beyersdorf, A. J., et al. (2013). Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proceedings of the National Academy of Sciences, 110*(7), 2575–2580. https://doi.org/10.1073/pnas.1212089110.

Delort, A., Delort, A. M., Vaïtilingom, M., Joly, M., Amato, P., Wirgot, N., et al. (2017). Clouds: A transient and stressing habitat for microorganisms. In C. Chénard & F. M. Lauro (Eds.), *Microbial ecology of extreme environments* (pp. 215–245). Cham: Springer. https://doi.org/10.1007/978-3-319-51686-8_10.

Di Giulio, M., Grande, R., Di Campli, E., Di Bartolomeo, S., & Cellini, L. (2010). Indoor air quality in university environments. *Environmental Monitoring and Assessment, 170*(1–4), 509–517. https://doi.org/10.1007/s10661-009-1252-7.

Dommergue, A., Amato, P., Tignat-Perrier, R., Magand, O., Thollot, A., Joly, M., et al. (2019). Methods to investigate the global atmospheric microbiome. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2019.00243.

Driks, A, & Eichenberger, P. (Eds.) 2016. Ecology of bacillaceae. In *The bacterial spore: From molecules to systems* (pp. 59–85). American Society of Microbiology. https://doi.org/10.1128/microbiolspec.TBS-0017-2013.

Du, R., Du, P., Lu, Z., Ren, W., Liang, Z., Qin, S., et al. (2017). Evidence for a missing source of efficient ice nuclei. *Scientific Reports, 7,* 1. https://doi.org/10.1038/srep39673.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics, 27*(16), 2194–2200. https://doi.org/10.1093/bioinformatics/btr381.

Elbert, W., Taylor, P. E., Andreae, M. O., & Pöschl, U. (2007). Contribution of fungi to primary biogenic aerosols in the atmosphere: Wet and dry discharged spores, carbohydrates, and inorganic ions. *Atmospheric Chemistry and Physics, 7*(17), 4569–4588. https://doi.org/10.5194/acp-7-4569-2007.

Els, N., Baumann-Stanzer, K., Larose, C., Vogel, T. M., & Sattler, B. (2019). Beyond the planetary boundary layer: Bacterial and fungal vertical biogeography at Mount Sonnblick. *Austria. Geo: Geography and Environment.*. https://doi.org/10.1002/geo2.69. **in press**.

Elster, J., Delmas, R. J., Petit, J.-R., & Reháková, K. (2007). Composition of microbial communities in aerosol, snow and ice samples from remote glaciated areas (Antarctica, Alps, Andes). *Biogeosciences Discussions, 4*(3), 1779–1813.

Evans, S. E., Elias Dueker, M., Robert Logan, J., & Weathers, K. C. (2019). The biology of fog: Results from coastal Maine and Namib Desert reveal common drivers of fog microbial composition. *Science of the Total Environment, 647*(January), 1547–1556. https://doi.org/10.1016/j.scitotenv.2018.08.045.

Fahlgren, C., Hagström, Å., Nilsson, D., & Zweifel, U. L. (2010). Annual variations in the diversity, viability, and origin of airborne bacteria. *Applied and Environment Microbiology, 76*(9), 3015–3025. https://doi.org/10.1128/AEM.02092-09.

Fang, Z., Guo, W., Zhang, J., & Lou, X. (2018). Influence of heat events on the composition of airborne bacterial communities in urban ecosystems. *International Journal of Environmental Research and Public Health, 15*(10), 2295. https://doi.org/10.3390/ijerph15102295.

Fierer, N., Liu, Z., Rodríguez-Hernández, M., Knight, R., Henn, M., & Hernandez, M. T. (2008). Short-term temporal variability in airborne bacterial and fungal populations. *Applied and Environmental Microbiology, 74*(1), 200–207. https://doi.org/10.1128/AEM.01467-07.

Franz, T. P., & Eisenreich, S. J. (1998). 'Snow scavenging of polychlorinated biphenyls and polycyclic aromatic hydrocarbons in minnesota. *Environmental Science & Technology, 10,* 1. https://doi.org/10.1021/es970601z.

Franzetti, A., Gandolfi, I., Gaspari, E., Ambrosini, R., & Bestetti, G. (2011). Seasonal variability of bacteria in fine and coarse urban air particulate matter. *Applied Microbiology and Biotechnology, 90*(2), 745–753. https://doi.org/10.1007/s00253-010-3048-7.

Fröhlich-Nowoisky, J., Burrows, S. M., Xie, Z., Engling, G., Solomon, P. A., Fraser, M. P., et al. (2012). Biogeography in the air: Fungal diversity over land and oceans.

Biogeosciences, 9(3), 1125–1136. https://doi.org/10.5194/bg-9-1125-2012.

Fuzzi, S., Mandrioli, P., & Perfetto, A. (1997). Fog droplets—an atmospheric source of secondary biological aerosol particles. *Atmospheric Environment, 31*(2), 287–290. https://doi.org/10.1016/1352-2310(96)00160-4.

Garcia, E., Hill, Thomas C. J., Prenni, A. J., DeMott, P. J., Franc, G. D., & Kreidenweis, S. M. (2012). Biogenic ice nuclei in boundary layer air over two U.S. High plains agricultural regions: Biogenic ice nuclei over two agricultural regions. *Journal of Geophysical Research: Atmospheres*. https://doi.org/10.1029/2012JD018343.

Gat, D., Mazar, Y., Cytryn, E., & Rudich, Y. (2017). Origin-dependent variations in the atmospheric microbiome community in eastern mediterranean dust storms. *Environmental Science and Technology, 51*(12), 6709–6718. https://doi.org/10.1021/acs.est.7b00362.

Geiser, D. M., Gueidan, C., Miadlikowska, J., Lutzoni, F., Kauff, F., Hofstetter, V., et al. (2006). Eurotiomycetes: Eurotiomycetidae and chaetothyriomycetidae. *Mycologia, 98*(6), 1053–1064. https://doi.org/10.1080/15572536.2006.11832633.

Gioulekas, D., Balafoutis, C., Damialis, A., Papakosta, D., Gioulekas, G., & Patakas, D. (2004). Fifteen years' record of airborne allergenic pollen and meteorological parameters in Thessaloniki, Greece. *International Journal of Biometeorology, 48*(3), 128–136. https://doi.org/10.1007/s00484-003-0190-2.

Gomes, E. A., Kasuya, M. C. M., Barros, E. G. D., Borges, A. C., & Araújo, E. F. (2002). Polymorphism in the internal transcribed spacer (ITS) of the ribosomal DNA of 26 isolates of ectomycorrhizal fungi. *Genetics and Molecular Biology, 25*(4), 477–483. https://doi.org/10.1590/S1415-47572002000400018.

Griffin, D. W. (2007). Atmospheric movement of microorganisms in clouds of desert dust and implications for human health. *Clinical Microbiology Reviews, 20*(3), 459–477. https://doi.org/10.1128/CMR.00039-06.

Hagan, M. E., Klotz, S. A., Bartholomew, W., Potter, L., & Nelson, M. (1995). A pseudoepidemic of rhodotorula rubra: A marker for microbial contamination of the bronchoscope. *Infection Control and Hospital Epidemiology, 16*(12), 727–728.

Hamaoui-Laguel, L., Vautard, R., Liu, L., Solmon, F., Viovy, N., Khvorostyanov, D., et al. (2015). Effects of climate change and seed dispersal on airborne ragweed pollen loads in Europe. *Nature Climate Change, 5*(8), 766–771. https://doi.org/10.1038/nclimate2652.

Hane, J. K., Lowe, Rohan G. T., Solomon, P. S., Tan, K.-C., Schoch, C. L., Spatafora, J. W., et al. (2007). Dothideomycete–plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen Stagonospora nodorum. *The Plant Cell, 19*(11), 3347–3368. https://doi.org/10.1105/tpc.107.052829.

Hänel, G. (1976). The properties of atmospheric aerosol particles as functions of the relative humidity at thermodynamic equilibrium with the surrounding moist air. In H. E. Landsberg & J. Van Mieghem (Eds.), *Advances in geophysics* (Vol. 19, pp. 73–188). Amsterdam: Elsevier. https://doi.org/10.1016/S0065-2687(08)60142-9.

Harding, T., Jungblut, A. D., Lovejoy, C., & Vincent, W. F. (2011). Microbes in high arctic snow and implications for the cold biosphere. *Applied and Environmental Microbiology, 77*(10), 3234–3243. https://doi.org/10.1128/AEM.02611-10.

Hermansson, J., & Thor, G. (2004). Byssoloma subdiscordans and usnea substerilis new to sweden. *Graphis Scripta, 15,* 3.

Hibbett, D. S., Bauer, R., Binder, M., Giachini, A. J., Hosaka, K., Justo, A., et al. (2014). 14 Agaricomycetes. In D. McLaughlin & J. W. Spatafora (Eds.), *Systematics and evolution. The Mycota (A comprehensive treatise on fungi as experimental systems for basic and applied research)* (Vol. 7A). Berlin: Springer. https://doi.org/10.1007/978-3-642-55318-9_14.

Honeyman, A. S., Day, M. L., & Spear, J. R. (2018). Regional fresh snowfall microbiology and chemistry are driven by geography in storm-tracked events, Colorado, USA. *Peer J, 6*, e5961. https://doi.org/10.7717/peerj.5961.

Huffman, J. A., Prenni, A. J., DeMott, P. J., Pöhlker, C., Mason, R. H., Robinson, N. H., et al. (2013). High concentrations of biological aerosol particles and ice nuclei during and after rain. *Atmospheric Chemistry and Physics, 13*(13), 6151–6164. https://doi.org/10.5194/acp-13-6151-2013.

Innocente, E., Squizzato, S., Visin, F., Facca, C., Rampazzo, G., Bertolini, V., et al. (2017). Influence of seasonality, air mass origin and particulate matter chemical composition on airborne bacterial community structure in the Po Valley, Italy. *The Science of the Total Environment, 593–594*(September), 677–687. https://doi.org/10.1016/j.scitotenv.2017.03.199.

Jäger, S. (2000). Ragweed (Ambrosia) sensitisation rates correlate withthe amount of inhaled airborne pollen. A 14-Year Studyin Vienna, Austria. *Aerobiologia, 16*(1), 149–153. https://doi.org/10.1023/A:1007603321556.

Jang, G. I., Hwang, C. Y., & Cho, B. C. (2018). Effects of heavy rainfall on the composition of airborne bacterial communities. *Frontiers of Environmental Science & Engineering, 12*(2), 12. https://doi.org/10.1007/s11783-018-1008-0.

Jeon, E. M., Kim, H. J., Jung, K., Kim, J. H., Kim, M. Y., Kim, Y. P., et al. (2011). Impact of Asian dust events on airborne bacterial community assessed by molecular analyses. *Atmospheric Environment, 45*(25), 4313–4321. https://doi.org/10.1016/j.atmosenv.2010.11.054.

Jiaxian, P., Shumin, Z., Kai, X., Junyang, Z., Yao Chuanhe, L., Senlin, Z. W., et al. (2019). Diversity of bacteria in cloud water collected at a national atmospheric monitoring station in Southern China. *Atmospheric Research, 218*(April), 176–182. https://doi.org/10.1016/j.atmosres.2018.12.004.

Jimenez-Sanchez, C., Hanlon, R., Aho, K. A., Powers, C., Morris, C. E., & Schmale, D. G. (2018). Diversity and ice nucleation activity of microorganisms collected with a small unmanned aircraft system (SUAS) in France and the United States. *Frontiers in Microbiology, 9,* 1. https://doi.org/10.3389/fmicb.2018.01667.

Joly, M., Amato, P., Deguillaume, L., Monier, M., Hoose, C., & Delort, A.-M. (2014). Direct quantification of total and biological ice nuclei in cloud water. *Atmospheric Chemistry and Physics Discussions, 14*(3), 3707–3731. https://doi.org/10.5194/acpd-14-3707-2014.

Kelley, S. T., & Gilbert, J. A. (2013). Studying the microbiology of the indoor environment. *Genome Biology, 14*(February), 202. https://doi.org/10.1186/gb-2013-14-2-202.

Khamis, A., Colson, P., Raoult, D., & La Scola, B. (2003). Usefulness of RpoB gene sequencing for identification of Afipia and Bosea Species, including a strategy for choosing discriminative partial sequences. *Applied and Environment Microbiology, 69*(11), 6740–6749. https://doi.org/10.1128/AEM.69.11.6740-6749.2003.

Kim, J. S., Shin, S. C., Hong, S. G., Lee, Y. M., Lee, H., Lee, J., et al. (2012). Genome sequence of Janthinobacterium Sp. Strain PAMC 25724, isolated from alpine glacier cryoconite. *Journal of Bacteriology, 194*(8), 2096. https://doi.org/10.1128/JB.00096-12.

Kirtsideli, I. Y., Vlasov, D. Y., Abakumov, D. V., Barantsevich, E. P., Novozhilov, Y. K., Krylenkov, V. A., et al. (2017). Airborne fungi in arctic settlement Tiksi (Russian Arctic, Coast of the Laptev Sea). *Czech Polar Reports, 7*(2), 300–310. https://doi.org/10.5817/CPR2017-2-29.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research, 41*(1), e1. https://doi.org/10.1093/nar/gks808.

Knight, C. A., & Knight, N. C. (2001). Hailstorms. In C. A. Doswell (Ed.), *Severe convective storms* (pp. 223–254)., Meteorological Monographs Boston: American Meteorological Society. https://doi.org/10.1007/978-1-935704-06-5_6.

Kõljalg, U., Henrik Nilsson, R., Abarenkov, K., Tedersoo, L., Taylor, Andy F. S., Bahram, M., et al. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology, 22*(21), 5271–5277. https://doi.org/10.1111/mec.12481.

Kourtev, P. S., Hill, K. A., Shepson, P. B., & Konopka, A. (2011). Atmospheric cloud water contains a diverse bacterial community. *Atmospheric Environment, 45*(30), 5399–5405. https://doi.org/10.1016/j.atmosenv.2011.06.041.

Krasowska, A., & Sigler, K. (2014). How microorganisms use hydrophobicity and what does this mean for human needs? *Frontiers in Cellular and Infection Microbiology*. https://doi.org/10.3389/fcimb.2014.00112.

Lallement, A., Besaury, L., Tixier, E., Sancelme, M., Amato, P., Vinatier, V., et al. (2018). Potential for phenol biodegradation in cloud waters. *Biogeosciences, 15*(18), 5733–5744. https://doi.org/10.5194/bg-15-5733-2018.

Larose, C., Berger, S., Ferrari, C., Navarro, E., Dommergue, A., Schneider, D., et al. (2010). Microbial sequences retrieved from environmental samples from seasonal arctic snow and meltwater from Svalbard, Norway. *Extremophiles, 14*(2), 205–212. https://doi.org/10.1007/s00792-009-0299-2.

Liu, X.-Z., Wang, Q.-M., Göker, M., Groenewald, M., Kachalkin, A. V., Lumbsch, H. T., et al. (2015). Towards an integrated phylogenetic classification of the tremellomycetes. *Studies in Mycology, 81*(June), 85–147. https://doi.org/10.1016/j.simyco.2015.12.001.

Lotteraner, C., & Piringer, M. (2016). Mixing-height time series from operational ceilometer aerosol-layer heights. *Boundary-Layer Meteorology, 161*(2), 265–287. https://doi.org/10.1007/s10546-016-0169-2.

Maharachchikumbura, Sajeewa S. N., Hyde, K. D., Gareth Jones, E. B., McKenzie, E. H. C., Bhat, J. D., Dayarathne, M. C., et al. (2016). Families of sordariomycetes. *Fungal Diversity, 79*(1), 1–317. https://doi.org/10.1007/s13225-016-0369-6.

Maki, L. R., Galyan, E. L., Chang-Chien, M.-M., & Caldwell, D. R. (1974). Ice nucleation induced by pseudomonas syringae. *Applied Microbiology, 28*(3), 456–459.

Maki, T., Hara, K., Iwata, A., Lee, K. C., Kawai, K., Kai, K., 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. Discuss., 2017*(January), 1–32. https://doi.org/10.5194/acp-2016-1095.

Maria, S. F., & Russell, L. M. (2005). Organic and inorganic aerosol below-cloud scavenging by suburban New Jersey precipitation. *Environmental Science and Technology, 39*(13), 4793–4800.

Martinez Arbizu, P. (2017). pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version 0.0.1. https://github.com/pmartinezarbizu/pairwiseAdonis.

Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G., & Neufeld, J. D. (2012). PANDAseq: Paired-End assembler for illumina sequences. *BMC Bioinformatics, 13*(1), 31. https://doi.org/10.1186/1471-2105-13-31.

Maynaud, G., Willems, A., Soussou, S., Vidal, C., Mauré, L., Moulin, L., et al. (2012). Molecular and phenotypic characterization of strains nodulating anthyllis vulneraria in mine tailings, and proposal of *Aminobacter anthyllidis* Sp. Nov., the first definition of aminobacter as legume-nodulating bacteria. *Systematic and Applied Microbiology, 35*(2), 65–72. https://doi.org/10.1016/j.syapm.2011.11.002.

McDonald, I. R., Kämpfer, P., Topp, E., Warner, K. L., Cox, M. J., Hancock, T. L. C., et al. (2005). *Aminobacter ciceronei* Sp. Nov. and *Aminobacter lissarensis* Sp. Nov., isolated from various terrestrial environments. *International Journal of Systematic and Evolutionary Microbiology, 55*(5), 1827–1832. https://doi.org/10.1099/ijs.0.63716-0.

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE, 8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217.

Miadlikowska, J., Kauff, F., Högnabba, F., Oliver, J. C., Molnár, K., Fraker, E., et al. (2014). A multigene phylogenetic synthesis for the class lecanoromycetes (Ascomycota): 1307 fungi representing 1139 infrageneric taxa, 317 genera and 66 families. *Molecular Phylogenetics and Evolution*. https://doi.org/10.1016/j.ympev.2014.04.003.

Michaud, A. B., Dore, J. E., Deborah Leslie, W., Lyons, B., Sands, D. C., & Priscu, J. C. (2014). Biological ice nucleation initiates hailstone formation. *Journal of Geophysical Research: Atmospheres, 119*(21), 12186–12197. https://doi.org/10.1002/2014JD022004.

Möhler, O., Demott, P. J., Vali, G., & Levin, Z. (2007). Microbiology and atmospheric processes: The role of biological particles in cloud physics. *Biogeosciences, 4*(6), 1059–1071.

Morris, C. E., Franz Conen, J., Huffman, A., Phillips, V., Pöschl, U., & Sands, D. C. (2014). Bioprecipitation: A feedback cycle linking earth history, ecosystem dynamics and land

use through biological ice nucleators in the atmosphere. *Global Change Biology, 20*(2), 341–351. https://doi.org/10.1111/gcb.12447.

Mühlhausen, S., & Kollmar, M. (2014). Molecular phylogeny of sequenced saccharomycetes reveals polyphyly of the alternative yeast codon usage. *Genome Biology and Evolution, 6*(12), 3222–3237. https://doi.org/10.1093/gbe/evu152.

Munday, C., De Deckker, P., Tapper, N., O'Loingsigh, T., & Allison, G. (2016). Characterizing bacterial assemblages in sediments and aerosols at a dry lake bed in australia using high-throughput sequencing. *Aerobiologia, 32*(4), 581–593. https://doi.org/10.1007/s10453-015-9407-1.

Nicholson, W. L., Munakata, N., Horneck, G., Melosh, H. J., & Setlow, P. (2000). Resistance of bacillus endospores to extreme terrestrial and extraterrestrial environments. *Microbiology and Molecular Biology Reviews, 64*(3), 548–572. https://doi.org/10.1128/MMBR.64.3.548-572.2000.

Oksanen, J., Blanchet, G, Friendly, F., Kindt, M., Legendre, R., McGlinn, P., Dan Minchin, P. R. et al. (2018). *Vegan: Community ecology package* (version 2.5-1). https://CRAN.R-project.org/package=vegan.

Øvreås, L., & Torsvik, V. (1998). Microbial diversity and community structure in two different agricultural soil communities. *Microbial Ecology, 36*(3), 303–315.

Paramonov, M., Grönholm, T., & Virkkula, A. (2011). Below-cloud scavenging of aerosol particles by snow at an urban site in Finland. *Boreal Environment Research, 16,* 18.

Park, H. K., Han, J.-H., Joung, Y., Cho, S.-H., Kim, S.-A., & Kim, S. B. (2014). Bacterial diversity in the indoor air of pharmaceutical environment. *Journal of Applied Microbiology, 116*(3), 718–727. https://doi.org/10.1111/jam.12416.

Petters, M. D., & Wright, T. P. (2015). Revisiting ice nucleation from precipitation samples. *Geophysical Research Letters, 42*(20), 8758–8766. https://doi.org/10.1002/2015GL065733.

Pickersgill, D. A., Wehking, J., Paulsen, H., Thines, E., Pöschl, U., Fröhlich-Nowoisky, J., et al. (2017). Lifestyle dependent occurrence of airborne fungi. *Biogeosciences Discuss., 2017*(November), 1–20. https://doi.org/10.5194/bg-2017-452.

Polymenakou, P. N. (2012). Atmosphere: A source of pathogenic or beneficial microbes? *Atmosphere, 3*(1), 87–102. https://doi.org/10.3390/atmos3010087.

Pouzet, G., Peghaire, E., Aguès, M., Baray, J.-L., Conen, F., & Amato, P. (2017). Atmospheric processing and variability of biological ice nucleating particles in precipitation at Opme, France. *Atmosphere, 8*(11), 229. https://doi.org/10.3390/atmos8110229.

Prasher, I. B. (2015). *Wood-rotting non-gilled Agaricomycetes of Himalayas*. Berlin: Springer.

Prévost-Bouré, C., Nicolas, R. C., Dequiedt, S., Mougel, C., Lelièvre, M., Jolivet, C., et al. (2011). Validation and application of a PCR primer set to quantify fungal communities in the soil environment by real-time quantitative PCR. *PLoS ONE, 6*(9), e24166. https://doi.org/10.1371/journal.pone.0024166.

Pringle, A., Patek, S. N., Fischer, M., Stolze, J., & Money, N. P. (2005). The captured launch of a ballistospore. *Mycologia, 97*(4), 866–871.

R Core Team. (2015). R: A language and environment for statistical computing. *R foundation for statistical compution. R: A language and environment for statistical computing*. https://www.R-project.org.

Reynolds, C. S. (2007). Variability in the provision and function of mucilage in phytoplankton: Facultative responses to the environment. *Hydrobiologia, 578*(1), 37–45. https://doi.org/10.1007/s10750-006-0431-6.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ, 4*(October), e2584. https://doi.org/10.7717/peerj.2584.

Rotach, M. W., Gohm, A., Lang, M. N., Leukauf, D., Stiperski, I., & Wagner, J. S. (2015). On the vertical exchange of heat, mass, and momentum over complex, mountainous terrain. *Frontiers in Earth Science*. https://doi.org/10.3389/feart.2015.00076.

Rubio, A. O., Kupferberg, S. J., García, V. V., Ttito, A., Shepack, A., & Catenazzi, A. (2018). Widespread occurrence of the antifungal cutaneous bacterium janthinobacterium lividum on andean water frogs threatened by fungal disease. *Diseases of Aquatic Organisms, 131*(3), 233–238. https://doi.org/10.3354/dao03298.

Šantl-Temkiv, T., Finster, K., Dittmar, T., Hansen, B. M., Thyrhaug, R., Nielsen, N. W., et al. (2013). Hailstones: A window into the microbial and chemical inventory of a storm cloud. *PLoS ONE, 8*(1), e53550. https://doi.org/10.1371/journal.pone.0053550.

Šantl-Temkiv, T., Gosewinkel, U., Starnawski, P., Lever, M., & Finster, K. (2018). Aeolian dispersal of bacteria in southwest greenland: Their sources, abundance, diversity and physiological states. *FEMS Microbiology Ecology*. https://doi.org/10.1093/femsec/fiy031.

Sattler, B., Puxbaum, H., & Psenner, R. (2001). Bacterial growth in supercooled cloud droplets. *Geophysical Research Letters, 28*(2), 239–242. https://doi.org/10.1029/2000GL011684.

Schmale, D. G., & Ross, S. D. (2015). Highways in the sky: Scales of atmospheric transport of plant pathogens. *Annual review of Phytopathology, 53*(1), 591–611. https://doi.org/10.1146/annurev-phyto-080614-115942.

Schoch, C. L., Crous, P. W., Groenewald, J. Z., Boehm, E. W. A., Burgess, T. I., de Gruyter, J., et al. (2009). A class-wide phylogenetic assessment of dothideomycetes. *Studies in Mycology, A phylogenetic re-evaluation of, 64*(January), 1–15. https://doi.org/10.3114/sim.2009.64.01.

Schultz-Jensen, N., Knudsen, B. E., Frkova, Z., Aamand, J., Johansen, T., Thykaer, J., et al. (2014). Large-scale bioreactor production of the herbicide-degrading *Aminobacter* sp. strain MSH1. *Applied Microbiology and Biotechnology, 98*(5), 2335–2344. https://doi.org/10.1007/s00253-013-5202-5.

Scola, L., Bernard, I. B., Greub, G., Khamis, A., Martin, C., & Raoult, D. (2003). Amoeba-resisting bacteria and ventilator-associated pneumonia. *Emerging Infectious Diseases, 9*(7), 815–821. https://doi.org/10.3201/eid0907.030065.

Segawa, T., Miyamoto, K., Ushida, K., Agata, K., Okada, N., & Kohshima, S. (2005). Seasonal change in bacterial flora

and biomass in mountain snow from the Tateyama Mountains, Japan, analyzed by 16S RRNA gene sequencing and real-time PCR. *Applied and Environmental Microbiology, 71*(1), 123–130. https://doi.org/10.1128/AEM.71.1.123-130.2005.

Setlow, P. (2006). Spores of Bacillus subtilis: Their resistance to and killing by radiation, heat and chemicals. *Journal of Applied Microbiology, 101*(3), 514–525. https://doi.org/10.1111/j.1365-2672.2005.02736.x.

Sharma, A., Dhar, S. K., Prakash, O., Vemuluri, V. R., Thite, V., & Shouche, Y. S. (2014). Description of *Domibacillus indicus* Sp. Nov., isolated from ocean sediments and emended description of the genus *Domibacillus*. *International Journal of Systematic and Evolutionary Microbiology, 64*(Pt 9), 3010–3015. https://doi.org/10.1099/ijs.0.064295-0.

Sharma, N. K., & Singh, S. (2010). Differential aerosolization of algal and cyanobacterial particles in the atmosphere. *Indian Journal of Microbiology, 50*(4), 468–473. https://doi.org/10.1007/s12088-011-0146-x.

Slinn, W. G. N. (1983). Precipitation scavenging. In D. Raderson (Ed.), *Atmospheric sciences and power production* (Chap. 11). Washington, DC: Division of Biomedical Environmental Research, U.S. Department of Energy.

Smith, D. J. (2012). *Long range transport of microorganisms in the upper atmosphere*. Seattle: University of Washington.

Smith, D. J., Griffin, D. W., & Jaffe, D. A. (2011). The high life: Transport of microbes in the atmosphere. *Eos, Transactions American Geophysical Union, 92*(30), 249–250. https://doi.org/10.1029/2011EO300001.

Smith, D. J., Jaffe, D. A., Birmele, M. N., Griffin, D. W., Schuerger, A. C., Hee, J., et al. (2012). Free tropospheric transport of microorganisms from Asia to North America. *Microbial Ecology, 64*(4), 973–985. https://doi.org/10.1007/s00248-012-0088-9.

Smith, D. J., Ravichandar, J. D., Jain, S., Griffin, D. W., Hongbin, Yu., Tan, Q., 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). *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2018.01752.

Smith, D. J., Timonen, H. J., Jaffe, D. A., Griffin, D. W., Birmele, M. N., Perry, K. D., et al. (2013). Intercontinental dispersal of bacteria and archaea by transpacific winds. *Applied and Environmental Microbiology, 79*(4), 1134–1139. https://doi.org/10.1128/AEM.03029-12.

Sofiev, M., Belmonte, J., Gehrig, R., zquierdo, R., Smith, M., Dahl, Å., et al. (2013). Airborne pollen transport. In M. Sofiev & K.-C. Bergmann (Eds.), *Allergenic pollen: A review of the production, release, distribution and health impacts* (pp. 127–159). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-4881-1_5.

Stephanie, & Waturangi, D. E. (2011). Distribution of ice nucleation-active (INA) bacteria from rain-water and air. *HAYATI Journal of Biosciences, 18*(3), 108–112. https://doi.org/10.4308/hjb.18.3.108.

Steven, B., Chen, M. Q., Greer, C. W., Whyte, L. G., & Niederberger, T. D. (2008). Tumebacillus permanentifrigoris gen. Nov., Sp. Nov., an aerobic, spore-forming bacterium isolated from canadian high arctic permafrost. *International Journal of Systematic and Evolutionary Microbiology, 58*(6), 1497–1501. https://doi.org/10.1099/ijs.0.65101-0.

Stohl, A., Eckhardt, S., Forster, C., James, P., Spichtinger, N., & Seibert, P. (2002). A replacement for simple back trajectory calculations in the interpretation of atmospheric trace substance measurements. *Atmospheric Environment, 36*(29), 4635–4648. https://doi.org/10.1016/S1352-2310(02)00416-8.

Stopelli, E., Conen, F., Morris, C. E., Herrmann, E., Bukowiecki, N., & Alewell, C. (2015). Ice nucleation active particles are efficiently removed by precipitating clouds. *Scientific Reports*. https://doi.org/10.1038/srep16433.

Sun, Q.-L., & Sun, L. (2016). Description of *Domibacillus iocasae* Sp. Nov., Isolated from Deep-Sea sediment, and emended description of the genus *Domibacillus*. *International Journal of Systematic and Evolutionary Microbiology, 66*(2), 982–987. https://doi.org/10.1099/ijsem.0.000823.

Taylor, D. Lee, Walters, W. A., Lennon, N. J., Bochicchio, J., Andrew Krohn, J., Caporaso, G., et al. (2016). Accurate estimation of fungal diversity and abundance through improved lineage-specific primers optimized for illumina amplicon sequencing. *Applied and Environmental Microbiology, 82*(24), 7217–7226. https://doi.org/10.1128/AEM.02576-16.

Techy, L., Schmale, D. G., & Woolsey, C. A. (2010). Coordinated aerobiological sampling of a plant pathogen in the lower atmosphere using two autonomous unmanned aerial vehicles. *Journal of Field Robotics, 27*(3), 335–343. https://doi.org/10.1002/rob.20335.

Temkiv, T. Š., Finster, K., Hansen, B. M., Nielsen, N. W., & Karlson, U. G. (2012). The microbial diversity of a storm cloud as assessed by hailstones. *FEMS Microbiology Ecology, 81*(3), 684–695. https://doi.org/10.1111/j.1574-6941.2012.01402.x.

Thomas-Hall, S. R., Turchetti, B., Buzzini, P., Branda, E., Boekhout, T., Theelen, B., et al. (2010). Cold-adapted yeasts from Antarctica and the Italian Alps—Description of three novel species: *Mrakia robertii* Sp. Nov., *Mrakia blollopis* Sp. Nov. and *Mrakiella niccombsii* Sp. Nov. *Extremophiles, 14*(1), 47–59. https://doi.org/10.1007/s00792-009-0286-7.

Urbano, R., Palenik, B., Gaston, C. J., & Prather, K. A. (2011). Detection and phylogenetic analysis of coastal bioaerosols using culture dependent and independent techniques. *Biogeosciences, 8*(2), 301–309. https://doi.org/10.5194/bg-8-301-2011.

Vali, G. (1971). Freezing nucleus content of hail and rain in alberta. *Journal of Applied Meteorology, 10*(1), 73–78. https://doi.org/10.1175/1520-0450(1971)010%3c0073:FNCOHA%3e2.0.CO;2.

Verma, A., Ojha, A. K., Dastager, S. G., Natarajan, R., Mayilraj, S., & Krishnamurthi, S. (2017). *Domibacillus mangrovi* Sp. Nov. and *Domibacillus epiphyticus* Sp. Nov., isolated from marine habitats of the central West Coast of India. *International Journal of Systematic and Evolutionary Microbiology, 67*(8), 3063–3070. https://doi.org/10.1099/ijsem.0.002085.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive bayesian classifier for rapid assignment of RRNA sequences into the new bacterial taxonomy. *Applied and*

Environmental Microbiology, 73(16), 5261–5267. https://doi.org/10.1128/AEM.00062-07.

Wang, Z., Johnston, P. R., Takamatsu, S., Spatafora, J. W., & Hibbett, D. S. (2006). Toward a phylogenetic classification of the leotiomycetes based on RDNA data. *Mycologia, 98*(6), 1065–1075. https://doi.org/10.1080/15572536.2006.11832634.

Wang, X., Zhang, L., & Moran, M. D. (2014). Development of a new semi-empirical parameterization for below-cloud scavenging of size-resolved aerosol particles by both rain and snow. *Geoscientific Model Development, 7*(3), 799–819. https://doi.org/10.5194/gmd-7-799-2014.

Weil, T., De Filippo, C., Albanese, D., Donati, C., Pindo, M., Pavarini, L., et al. (2017). Legal immigrants: Invasion of alien microbial communities during winter occurring desert dust storms. *Microbiome, 5*(March), 32. https://doi.org/10.1186/s40168-017-0249-7.

Weiss, S., Zhenjiang Zech, X., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome, 5*(March), 27. https://doi.org/10.1186/s40168-017-0237-y.

Wekker, D., Kossmann, M., & Kossmann, M. (2015). Convective boundary layer heights over mountainous terrain—A review of concepts. *Frontiers in Earth Science*. https://doi.org/10.3389/feart.2015.00077.

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. New-York: Springer. http://ggplot2.org.

Woo, C., An, C., Siyu, X., Yi, S.-M., & Yamamoto, N. (2018). Taxonomic diversity of fungi deposited from the atmosphere. *The ISME Journal, 12*(8), 2051. https://doi.org/10.1038/s41396-018-0160-7.

Wösten, H. A. (2001). Hydrophobins: Multipurpose proteins. *Annual Review of Microbiology, 55,* 625–646. https://doi.org/10.1146/annurev.micro.55.1.625.

Wösten, Han A. B., van Wetter, M.-A., Lugones, L. G., van der Mei, H. C., Busscher, H. J., & Wessels, Joseph G. H. (1999). How a fungus escapes the water to grow into the air. *Current Biology, 9*(2), 85–88. https://doi.org/10.1016/S0960-9822(99)80019-0.

Xia, Y., Conen, F., & Alewell, C. (2013). Total bacterial number concentration in free tropospheric air above the alps. *Aerobiologia, 29*(1), 153–159. https://doi.org/10.1007/s10453-012-9259-x.

Yan, D., Zhang, T., Jing, S., Zhao, L.-L., Wang, H., Fang, X.-M., et al. (2018). Structural variation in the bacterial community associated with airborne particulate matter in Beijing, China, during Hazy and Nonhazy Days. *Applied and Environment Microbiology, 84*(9), e00004–e00018. https://doi.org/10.1128/AEM.00004-18.

Yu, T.-T., Zhou, E.-M., Yin, Y.-R., Yao, J.-C., Ming, H., Dong, L., et al. (2013). *Vulcaniibacterium tengchongense* Gen. Nov., Sp. Nov. Isolated from a geothermally heated soil sample, and reclassification of lysobacter thermophilus Wei et Al. 2012 as Vulcaniibacteriumthermophilum Comb. Nov. *Antonie van Leeuwenhoek, 104*(3), 369–376. https://doi.org/10.1007/s10482-013-9959-4.

Yuan, Y., Si, G., Wang, J., Luo, T., & Zhang, G. (2014). Bacterial community in alpine grasslands along an altitudinal gradient on the Tibetan Plateau. *FEMS Microbiology Ecology, 87*(1), 121–132. https://doi.org/10.1111/1574-6941.12197.

Zalar, P., Sybren de Hoog, G., Schroers, H.-J., Frank, J. M., & Gunde-Cimerman, N. (2005). Taxonomy and phylogeny of the xerophilic genus *Wallemia* (Wallemiomycetes and Wallemiales, Cl. et Ord. Nov.). *Antonie van Leeuwenhoek, 87*(4), 311–328. https://doi.org/10.1007/s10482-004-6783-x.

Zeigler, D. R. (2014). The geobacillus paradox: Why is a thermophilic bacterial genus so prevalent on a mesophilic planet? *Microbiology, 160*(Pt_1), 1–11. https://doi.org/10.1099/mic.0.071696-0.

Zhang, Y.-B., Wang, Y.-L., Li, W.-H., Bao, L.-N., Huang, X.-H., Huang, B., et al. (2018). Biogas emission from an anaerobic reactor. *Aerosol and Air Quality Research, 18*(6), 1493–1502. https://doi.org/10.4209/aaqr.2018.05.0169.

Zhang, L., Wang, X., Moran, M. D., & Feng, J. (2013). Review and uncertainty assessment of size-resolved scavenging coefficient formulations for below-cloud snow scavenging of atmospheric aerosols. *Atmospheric Chemistry and Physics, 13*(19), 10005–10025. https://doi.org/10.5194/acp-13-10005-2013.

Zhu, C., Chen, J., Wang, X., Li, J., Wei, M., Caihong, X., et al. (2018). Chemical composition and bacterial community in size-resolved cloud water at the summit of Mt. Tai, China. *Aerosol and Air Quality Research, 18*(1), 1–14. https://doi.org/10.4209/aaqr.2016.11.0493.

Ziello, C., Sparks, T. H., Estrella, N., Belmonte, J., Bergmann, K. C., Bucher, E., et al. (2012). Changes to airborne pollen counts across Europe. *PLoS ONE, 7*(4), e34076. https://doi.org/10.1371/journal.pone.0034076.

Zikova, N., & Zdimal, V. (2016). Precipitation scavenging of aerosol particles at a rural site in the Czech Republic. *Tellus B: Chemical and Physical Meteorology, 68*(1), 27343. https://doi.org/10.3402/tellusb.v68.27343.

Zweifel, U. L., Hagström, Å., Holmfeldt, K., Thyrhaug, R., Geels, C., Frohn, L. M., et al. (2012). High bacterial 16S RRNA gene diversity above the atmospheric boundary layer. *Aerobiologia, 28*(4), 481–498. https://doi.org/10.1007/s10453-012-9250-6.