The authors would like to thank the Associate Editor for handling and reviewing our manuscript, as well as the Referees for reviewing our paper and their feedbacks.

We took into account all Referees and Associate Editor comments and modified the text and SI file accordingly. The changes are highlighted in grey.

Below our additional responses regarding Associate Editor specific comments:

-We increased the size of the numbers in the Figure 2 in the way that they do not appear blurry while being not too large (as they are just here to detail the y axis). We could increase again the size if necessary.

-We completed the Figure 3 legend.

-We added information on the air samplers used in the Material and Methods section, and added a column in the Table S1 to indicate which sampler was used for each sample. We have photos of the samplers that we could add in the SI file.

-We removed the word "preliminary" in the Abstract. We wanted to say that it is the first metagenomic study comparing air metagenomes with metagenomes coming from other ecosystems, but we agree that our sentence might have been confusing.

-We modified the text regarding the qPCR data and purpose in the different sections. We hope that we made it clearer.

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

Correspondence to: Romie Tignat-Perrier (romie.tignat-perrier@univ-grenoble-alpes.fr)

Abstract

Microorganisms are ubiquitous in the atmosphere and 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 cultivable 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. 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 an atmosphere-specific functional potential signature as compared to other ecosystems (*i.e.* soil, sediment, snow, feces, surface seawater *etc*.). In absence of a specific atmospheric signature, the atmospheric samples collected at nine sites around the world were similar to their underlying ecosystems. In addition, atmospheric samples were characterized by a relatively high proportion of fungi. 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*.) resulted in part from the high concentrations of fungi that might resist and survive atmospheric physical stress better than bacteria.

Keywords: atmospheric microorganisms, airborne microbial communities, planetary boundary layer, metagenomic sequencing, comparative metagenomics, selective processes

1 Introduction

Microorganisms are ubiquitous in the atmosphere and reach concentrations of up to 10^6 microbial cells per cubic meter of air (Tignat-Perrier et al., 2019). Due to their important roles in public health and meteorological processes (Ariya et al., 2009; Aylor, 2003; Brown and Hovmøller, 2002; Delort et al., 2010; Griffin, 2007), understanding how airborne microbial communities are distributed over time and space is critical. While the concentration and taxonomic diversity of airborne microbial communities in the planetary boundary layer have recently been described (Els et al., 2019; Innocente et al., 2017; Tignat-Perrier et al., 2019), the functional potential of airborne microbial communities remains unknown. Most studies have focused on laboratory cultivation to identify possible metabolic functions of microbial strains of atmospheric origin, mainly from cloud water (Amato et al., 2007; Ariya et al., 2002; Hill et al., 2007; Vaïtilingom et al., 2010, 2013). Given that cultivatable organisms represent about 1 % of the entire microbial community (Vartoukian et al., 2010), culture-independent techniques and especially metagenomic studies applied to atmospheric microbiology have the potential to provide additional information on the selection and genetic adaptation of airborne

microorganisms. However, to our knowledge, only five metagenomic studies on airborne microbial communities at one or two specific sites per study exist (Aalismail et al., 2019; Amato et al., 2019; Cao et al., 2014; Gusareva et al., 2019; Yooseph et al., 2013). Metagenomic investigations of complex microbial communities in many ecosystems (for example, soil, seawater, lakes, feces, sludge) have provided evidence that microorganism functional signatures reflect the abiotic conditions of their environment, with different relative abundances of specific microbial functional classes (Delmont et al., 2011; Li et al., 2019; Tringe et al., 2005; Xie et al., 2011). This observed correlation of microbial community functional potential and the physical and chemical characteristics of their environments could have resulted from genetic modifications (microbial adaptation) (Brune et al., 2000; Hindré et al., 2012; Rey et al., 2016; Yooseph et al., 2010) and/or physical selection. The latter refers to the death of sensitive cells and the survival of resistant or previously adapted cells. This physical selection can occur when microorganisms are exposed to physiologically adverse conditions.

The presence of a specific 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 cloud water (*i.e.* high concentrations of H_2O_2, typical cloud carbonaceous sources, UV radiation *etc*.) (Amato et al., 2007; Joly et al., 2015; Vaïtilingom et al., 2013). While atmospheric chemicals might lead to some microbial adaptation, physical and unfavorable conditions 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's 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 (Huang and Hull, 2017) and might behave differently than bacterial cells. Still, the proportion and nature (*i.e.* fungi versus bacteria) of microbial cells that are resistant to the harsh atmospheric conditions within airborne microbial communities are unknown.

Our objective was to determine whether airborne microorganisms in the planetary boundary layer possess a specific functional signature as compared to other ecosystems since this might indicate that microorganisms with specific functions tend to be more aerosolized and/or undergo a higher survival in this environment. Our previous study showed that airborne microbial taxonomy mainly depends on the underlying ecosystems, indicating that the local environments are the main source of airborne microorganisms (Tignat-Perrier et al., 2019). Still, we do not know if airborne microbial communities result from random or specific aerosolization of the underlying ecosystems' 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 the atmosphere 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 and to capture the between and within-site variability in atmospheric microbial functional potential.

2 Material and Methods

2.1 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 sub-Antarctica; **Fig 1**), elevations from sea level (from 59 m to 5230 m; **Fig 1**) and environment type (from marine for Amsterdam-Island or AMS, to coastal for Cape Point or CAP, polar for Station Nord or STN and terrestrial for Grenoble or GRE, Chacaltaya or CHC, puy de Dôme or PDD, Pic-du-Midi or PDM, Storm-Peak or STP and Namco or NAM - **Table S1**). The number of samples collected per site varied from seven to

sixteen (**Table S1**). We collected particulate matter smaller than 10 µm (PM10) on quartz fiber filters (5.9'' round filter and 8'' × 10'' rectangular types) using high volume air samplers installed on roof tops or terraces (roughly 10 m above ground level). A TISCH TE-5170V sampler, DIGITEL DA77/DA80 sampler, Chinese 2131 Laowin sampler or a custom made sampler (*i.e.* high-volume pump connected to a DIGITEL PM10 head and airflow meter) was used depending on the site (**Table S1**). 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 h) and UV radiation, respectively as detailed in (Dommergue et al., 2019). 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 including negative control filters are presented in Dommergue et al. 2019. MODIS (Moderate resolution imaging spectroradiometer) land cover approach (5' x 5' resolution) (Friedl et al., 2002; Shannan et al., 2014) was used to quantify landscapes in the 50 km diameter area of our nine sampling sites (**Fig S1**).

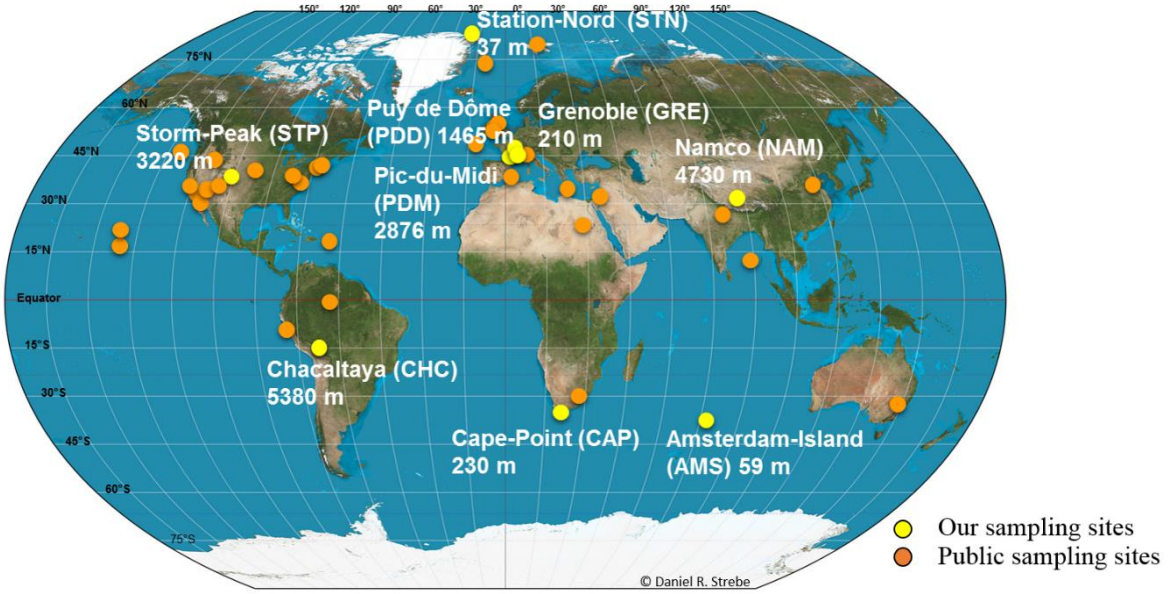


Fig 1. **Sample collection locations**. Map showing the geographical location and elevation from sea level of our nine sampling sites (in yellow), and the geographical position of whose public metagenomes come from (in orange). Abbreviations of our nine sampling sites are indicated in brackets.

2.2 Molecular biology analyses

2.2.1 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). During 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 by centrifugation at 1000 rcf for 4 min. From this step on, we followed the DNeasy PowerWater protocol. We conducted additional extractions on French agricultural soil samples collected at the Côte Saint André (that is part of the sample collection locations). We used 250 mg of soil

on which the same DNA extraction methodology as for air samples was applied. DNA concentration eluted in 100 µL of buffer was measured using the Qubit Fluorometric Quantification kit (Thermo Fisher Scientific). DNA was stored at -20 °C.

2.2.2 Real-Time qPCR analyses

16S rRNA gene and 18rRNA gene qPCR. The 16S and 18SS rRNA gene copy numbers were calculated per cubic meter of air (for air samples) and per gram of soil (for soil samples). Standards, primers and methodology are presented in Tignat-Perrier et al., 2019.

2.2.3 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 with DNA concentrations below 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 merged 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 from the dataset.

2.2.4 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 (*i.e.* MiSeq, HiSeq or 454) are summarized in **Table S2**. The sampling sites are positioned on the map in **Fig 1**.

2.3 Data analyses

All graphical and multivariate statistical analyses were carried out using the vegan (Oksanen et al., 2019), ggplot2 (Hadley and Winston, 2019) and reshape2 (Wickham, 2017) packages in the R environment (version 3.5.1).

2.3.1 Annotation of the metagenomic reads

Firstly, to access the overall functional potential of each sample, all the filtered sequences per sample were functionally annotated using Diamond and the nr database, then the gene-annotated sequences were grouped in the different SEED functional classes (around 7000 functional classes, referred simply to as functions) using MEGAN version 6 (Huson et al., 2009). Functional classes that were present ≤ 2 times in a sample were removed of this sample. In parallel, the Kraken software (Wood and Salzberg, 2014) was used to retrieve the bacterial and fungal sequences separately from the filtered sequences using the Kraken bacterial database and FindFungi (Donovan et al., 2018) fungal database (both databases included complete genomes), respectively (and using two different runs of Kraken). Separately, both the bacterial and fungal sequences were also functionally annotated using Diamond and MEGAN version 6 (number of sequences functionally annotated in **Table S3**).

Secondly, for specific metabolic and stress-related functions, we annotated the sequences using eggNOG-Mapper version 1 (Diamond option), then examined specific GO (Gene Ontology) terms chosen based on their importance for microbial resistance to atmospheric-like conditions. 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: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, bacterial sequences and fungal sequences for each sample. The ratio was given per 10000 annotated sequences and not 100 annotated sequences to get hit numbers superior to 1. The number of sequences annotated by eggNOG-Mapper (Huerta-Cepas et al., 2017) was also evaluated (**Table S3**). 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 proteins associated to this function (or functional class).

2.3.2 Statistical analyses

Observed functional richness and evenness were calculated per sample after rarefaction on all sequences (rarefaction at 2000 sequences), bacterial sequences (rarefaction at 500 sequences) and fungal sequences (rarefaction at 500 sequences). Rarefaction was used only for the purpose of calculating the diversity metrics (richness and evenness). For the following analyses, no rarefaction was applied on the number of reads per sample. The distribution of the samples was analyzed based on the SEED functional classes (using all sequences). PCoA and hierarchical clustering analysis (average method) were carried out on the Bray-Curtis dissimilarity matrix based on the relative abundances of the different SEED functional classes. SIMPER analyses were used to identify the functions responsible for the clustering of samples in groups. Because of the non-normality of the data, Kruskal-Wallis analyses (non-parametric version of ANOVA) and Dunn's post-hoc tests were used to test the difference between the percentage of fungal 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.

3 Results

3.1 Percentage of fungal sequences

The percentage of sequences annotated as belonging to fungal genomes (or fungal sequences, as opposed to bacterial sequences) was on average higher in air samples compared to soil ($P<10^{-5}$), snow ($P=10^{-3}$), seawater ($P=0.03$) and sediment samples ($P=10^{-3}$; **Fig 2** and **Table S4**). Among the air samples, NAM (19%), STN (24%) and CHC (27%) showed the lowest percentages of fungal sequences on average 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, were not evaluated by the Kruskal-Wallis test, we observed average percentages of fungal 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 fungal sequences while other samples had less than 50%. The approximated number of fungal and bacterial cells in air and soil was also estimated using 16S rRNA and 18S rRNA gene copy numbers per cubic meter of air and gram of soil, respectively. Air samples showed ratios between 16S and 18S rRNA gene copy numbers from around 4.5 times up to 160 times lower than soil samples (**Table S4**; some qPCR data have already been published in Tignat-Perrier et al., 2020). The qPCR data were used to see if similar results, or the same trend, on the estimated ratio between fungi and bacteria in air compared to the planetary-bound ecosystems could be obtained from metagenomic data and qPCR.

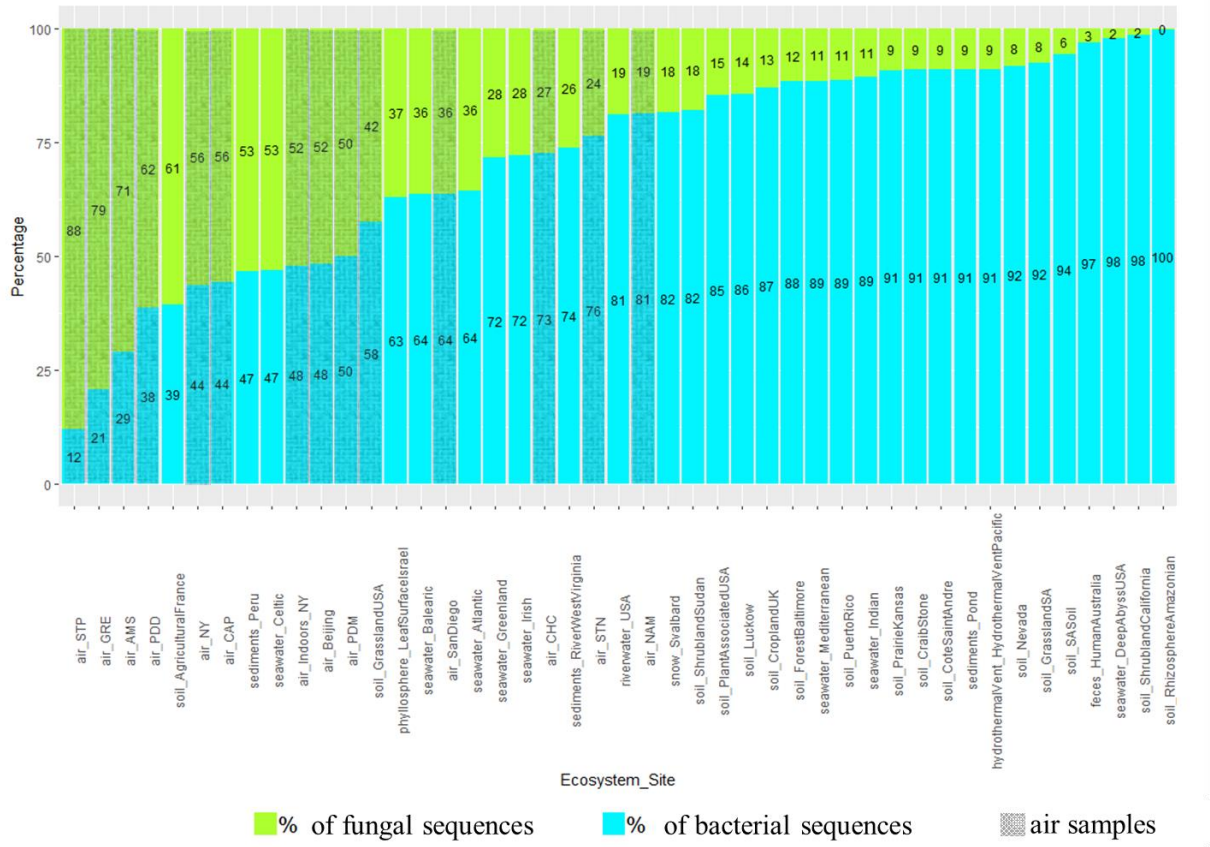


Fig 2. **Percentage of fungal and bacterial sequences in the metagenomes.** The percentages are established as the number of sequences annotated as belonging to fungal and bacterial genomes over the sum of bacterial and fungal sequences in the metagenomes. The mean was calculated for the sampling sites including several metagenomes. Air sites (*i.e.* our 9 sites + 5 sites where public air metagenomes come from) are distinguished by grey hatching lines.

3.2 Airborne microbial functional profiles

The fifty most abundant SEED functional classes represented in atmospheric samples are listed in **Table S5**. The 5-FCL-like protein, the long chain fatty acid CoA ligase and the TonB-dependent receptor were the top three functions based on number of annotated reads observed when including all the sequences (**Table S5**). The atmospheric 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 PCo multivariate analysis to visualize differences and similarities. The different samples (sampled during sequential weeks) from the same site did not cluster tightly together on the PCo multivariate analysis. In order to incorporate weekly variation when comparing sites, we used the microbial functional profile averaged per site in the subsequent multivariate analyses done with the data from other ecosystems (**Fig 3**). The PCo multivariate analysis showed that terrestrial atmospheric sites (GRE, NAM, STP, PDD, PDM, CHC, New York) grouped with the soil, sediment and snow samples while the marine and coastal atmospheric sites (AMS, CAP, San Diego) were situated between the datasets from soil, seawater and river water (**Fig 3**). The polar site STN did not group with the other sites. When considering only the bacterial sequences (*i.e.*, excluding the fungal sequences), the distribution of the terrestrial atmospheric sites did not change, while the marine Amsterdam-Island, coastal Cape Point and polar Station Nord

atmospheric sites were further from the seawater and river water datasets than when the fungal sequences were included (**Fig S2**). The distribution of the different datasets underwent further changes when considering only the fungal sequences. We observed an absence of a clear separation between soil and seawater since they (for the majority) grouped closely together, and terrestrial atmospheric datasets did not group with the other non-atmospheric datasets from soil, sediment and snow (**Fig S2**).

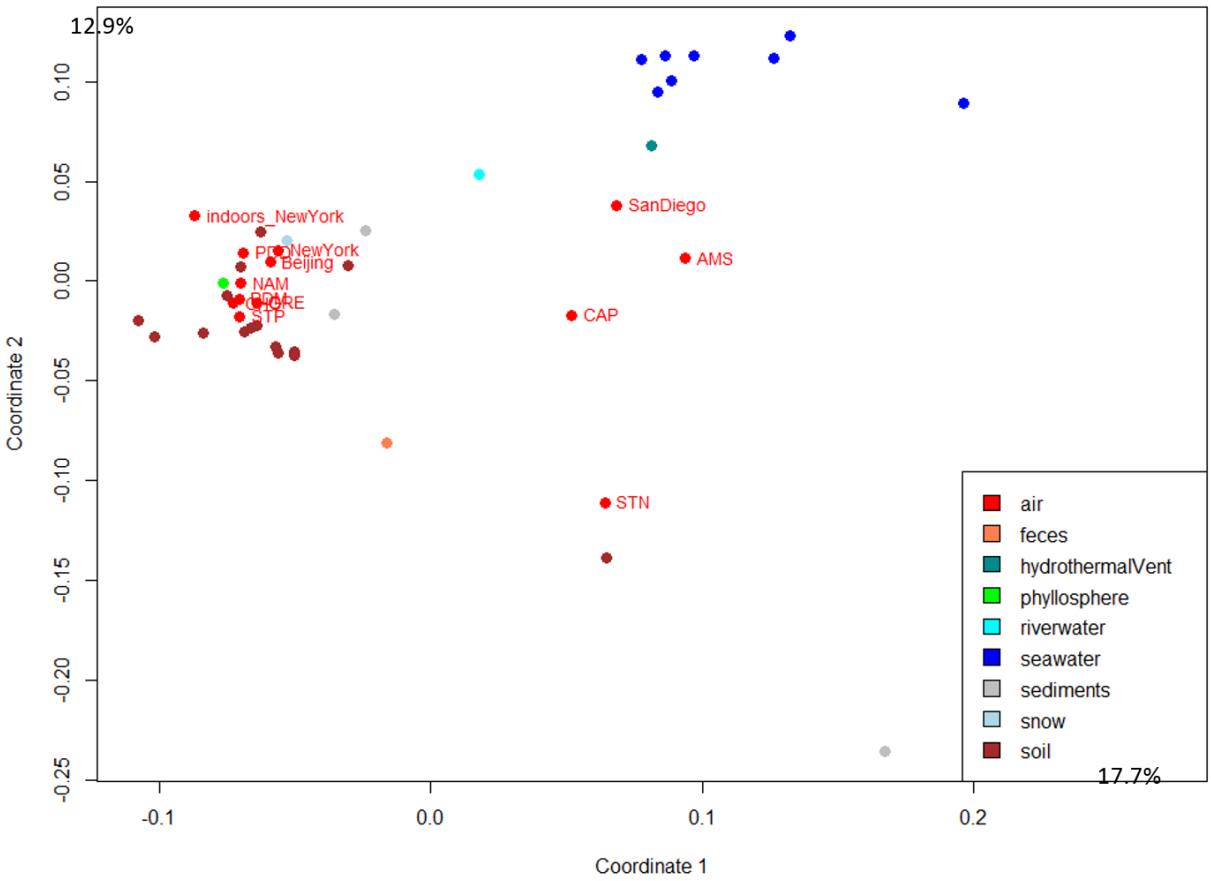


Fig 3. Distribution of the samples based on the microbial functional profile. The PCo analysis of the Bray-Curtis dissimilarity matrix is based on the functional potential structure of each site. For the site including several metagenomes, the average profile was calculated. Colors indicate the ecosystems in which the sites belong to. The percentages (17.7 % and 12.9 %) indicate the proportion of the variance explained by the first and second axis, respectively.

3.3 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 the PBL was lower than the average functional class richness in soil, surface seawater, hydrothermal vents, river water, phyllosphere and feces (P<0.05) (**Table S3**). Among the different atmospheric 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 Station Nord (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**). 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 Amsterdam-Island and 1142 +/- 461 functional classes in Chacaltaya. For fungal sequences, the functional class richness in the atmosphere was lower than the functional class richness in soil, surface seawater, feces, hydrothermal vents, river water and phyllosphere ($P<0.05$) (**Table S3**). Within air samples, the functional class richness based on fungal 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 ($P<10^{-5}$) whose values ranged from 66 +/- 58 functional classes in Amsterdam-Island and 392 +/- 131 functional classes in Storm Peak (**Table S3**). 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 bacterial and fungal sequences separately, the functional class evenness in air was on average higher than in soil, feces, phyllosphere and riverwater ($P<0.05$) (**Table S3**).

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

Two metabolic functions associated with abundant atmospheric chemicals (H_2O_2 and CH_4) were examined, hydrogen catabolism and methane monoxygenase activity. The concentration of sequences annotated as hydrogen peroxide catabolic related functional proteins per 10000 sequences varied between air sites ($P=2\times10^{-5}$) with highest values for Amsterdam-Island (27 +/- 1) and Grenoble (27 +/- 1) (**Fig S3**). It was on average higher in air compared to soil ($P=10^{-4}$) and surface seawater ($P=10^{-4}$). The French agricultural soil showed the highest relative abundance (133 +/- 4). When considering the fungal and bacterial sequences separately, this concentration was not different between air and the other ecosystems ($P>0.05$) (**Fig S3**). The number of sequences annotated as methane monooxygenase-related functional proteins per 10000 sequences was only detectable when considering all the sequences (*i.e.* bacterial and fungal sequences). The number of sequences annotated as methane monooxygenase-related functional proteins did not vary between air sites ($P>0.05$) while we observed a high variability between sampling periods within sites, but on average it was not different from the ecosystems ($P>0.05$).

Different stress response functions (sporulation, UV response, oxidative stress cell response, desiccation response, chromosome plasmid partitioning protein ParA and lipoate synthase) were examined. The concentration of sequences annotated as sporulation-related functional proteins per 10000 annotated sequences largely varied between air sites ($P=2\times10^{-9}$), with the lowest values observed for Station Nord (7 +/- 9), San Diego (9 +/- 6), Namco (17 +/- 15) and Chacaltaya (26 +/-13), and the highest values observed for Storm Peak (120 +/- 18), Beijing (126 +/- 22), Grenoble (131 +/- 21) and New York (141 +/- 98) (**Fig 4**). It was on average higher in air compared to soil ($P<10^{-5}$), sediments ($P<10^{-5}$) and surface seawater ($P=4\times10^{-4}$) although the Celtic seawater sample presented a very high concentration (127). Snow showed a relatively high average concentration (*i.e.* 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 Kruskal-Wallis tests), feces showed a relatively high concentration of sequences annotated as sporulation-related functional proteins (*i.e.* 41) while hydrothermal vent, phyllosphere and river water showed relatively low concentrations compared to air (<10). When considering the fungal sequences separately from the bacterial sequences, the same trend was observed, *i.e.* the concentration of sequences annotated as sporulation-related functional proteins in air was on average higher compared to soil ($P<10^{-5}$), sediments ($P<10^{-5}$), surface seawater ($P=7\times10^{-4}$) 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=3×10^{-5}). When considering the bacterial sequences only, this concentration in air was on average higher compared to soil (P=0.02), sediments (P=4×10^{-3}) and snow (P=0.01), and showed a smaller variability between air sites. Two samples, the phyllosphere (*i.e.* 35) and the shrubland soil from Sudan (*i.e.* 32) showed high numbers of sequences annotated as sporulation-related functional proteins per 10000 annotated sequences (**Fig 4**).

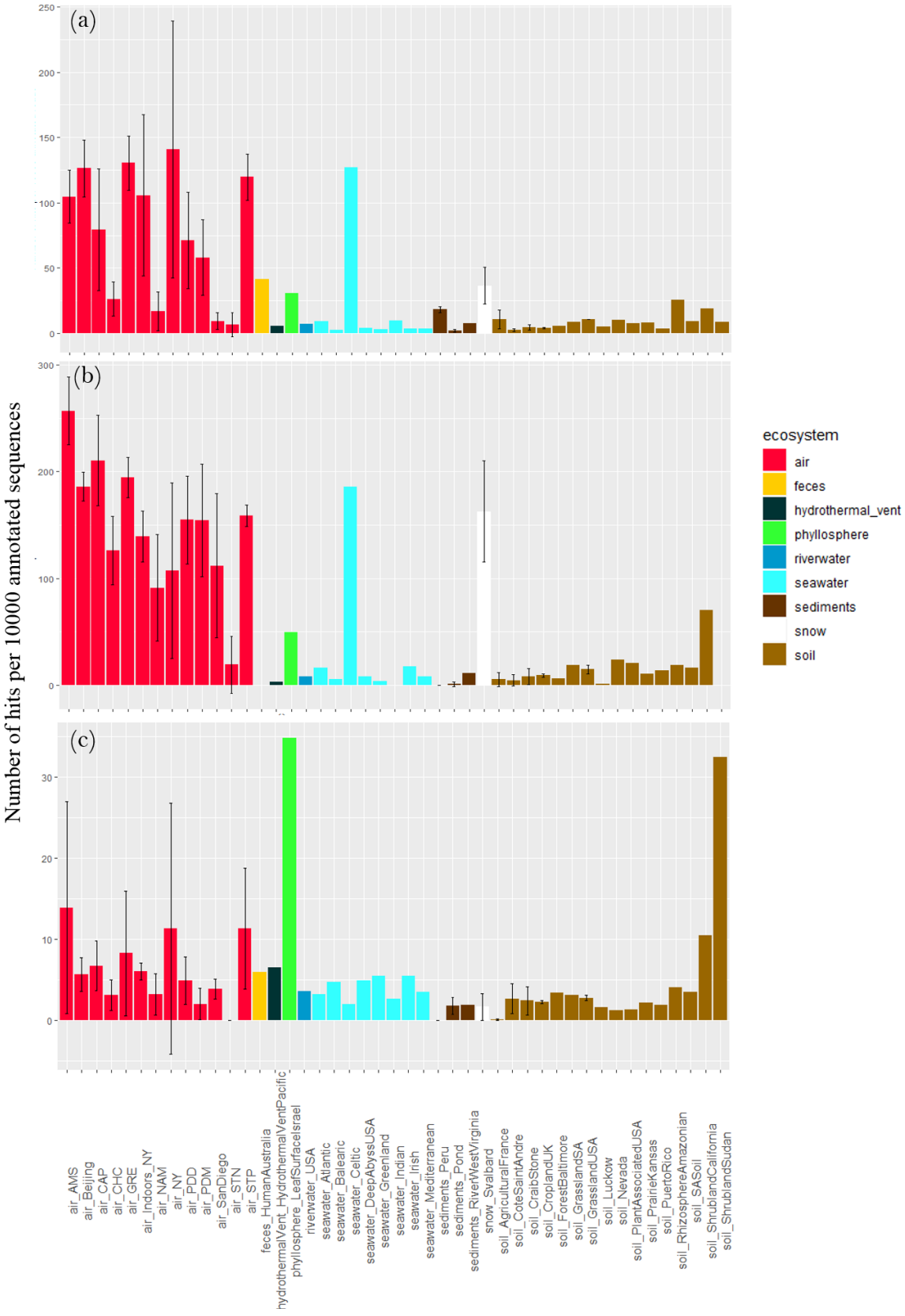
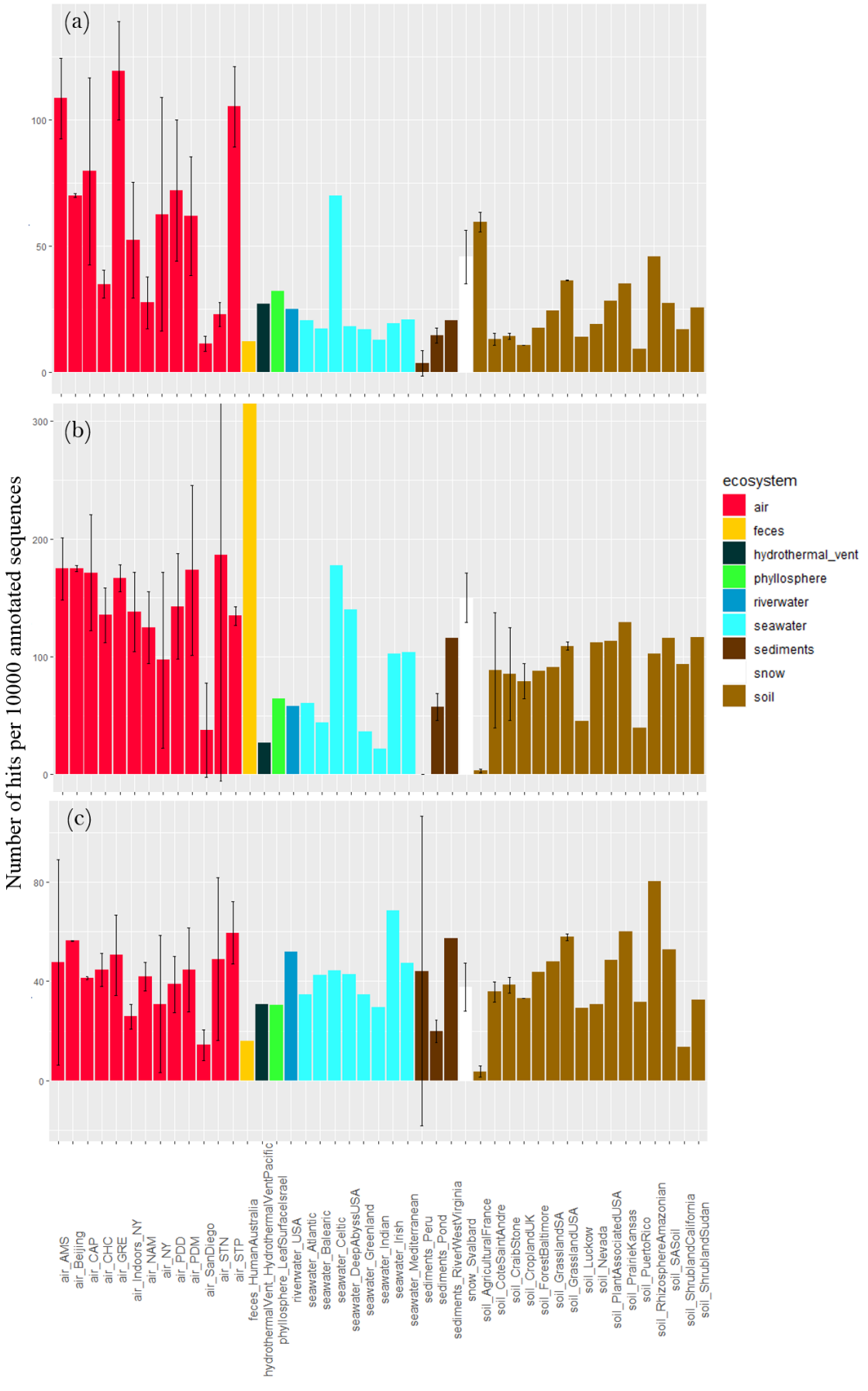


Fig 4. Proportion of sequences annotated as sporulation related functional proteins in the metagenomes. Average number of sequences annotated as proteins implicated in sporulation per 10000 annotated sequences from (**a**) all sequences, (**b**) fungal sequences and (**c**) bacterial

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 sequences annotated as UV response related functional proteins per 10000 annotated sequences varied between air sites ($P=10^{-5}$), with values ranging from 16 +/- 2 in Namco and 19 +/- 4 in STN to 29 +/- 3 in Storm Peak and 36 +/- 6 in Amsterdam-Island (**Fig S4**). The concentration was on average higher in air compared to sediments ($P<10^{-5}$), soil ($P<10^{-5}$) and comparable to snow and surface 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 fungal sequences separately, the concentration of sequences annotated as UV response related functional proteins was higher in air compared to soil ($P=9\times10^{-4}$), and comparable to the other ecosystems ($P>0.05$). When considering the bacterial sequences only, this concentration in air was on average higher compared to seawater ($P=3\times10^{-3}$) and sediments ($P=6\times10^{-3}$).
The concentration of sequences annotated as oxidative stress cell response related functional proteins per 10000 annotated sequences varied largely between air sites ($P=5\times10^{-7}$), with the lowest values observed for Station Nord (23 +/- 5), San Diego (11 +/- 3) and Namco (28 +/- 10), and the highest values observed for Storm Peak (105 +/- 16), Amsterdam-Island (108 +/- 16) and Grenoble (119 +/- 19) (**Fig 5**). The concentration was on average higher in air compared to soil ($P<10^{-5}$), sediments ($P<10^{-5}$) and surface seawater ($P=2\times10^{-3}$). 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 fungal sequences separately, the concentration of sequences annotated as oxidative stress related functional proteins per 10000 sequences was on average higher in air compared to soil ($P<10^{-5}$), sediments ($P<10^{-5}$) and surface seawater ($P=10^{-3}$). Feces showed a very high average value (2237). When considering bacterial sequences separately, this concentration was not different between air and the other ecosystems ($P>0.05$). When considering both fungal and bacterial sequences separately, the variability in the concentration of sequences annotated as oxidative stress cell response related functional proteins between air sites diminished and their difference was not detected anymore ($P>0.05$).


(a)
(b)
(c)
Number of hits per 10000 annotated sequences
ecosystem
air
feces
hydrothermal_vent
phyllosphere
riverwater
seawater
sediments
snow
soil
air_AMS
air_Beijing
air_CAP
air_CHC
air_GRE
air_Indoors_NY
air_NAM
air_NY
air_PDD
air_PDM
air_SanDiego
air_STN
air_STP
feces_HumanAustralia
hydrothermalVent_HydrothermalVentPacific
phyllosphere_LeafSurfaceIsrael
riverwater_USA
seawater_Atlantic
seawater_Balearic
seawater_Celtic
seawater_DeepAbyssUSA
seawater_Greenland
seawater_Indian
seawater_Irish
seawater_Mediterranean
sediments_Peru
sediments_Pond
sediments_RiverWestVirginia
snow_Svalbard
soil_AgriculturalFrance
soil_CoteSaintAndre
soil_CraibStone
soil_CroplandUK
soil_ForestBaltimore
soil_GrasslandSA
soil_GrasslandUSA
soil_Luckow
soil_Nevada
soil_PlantAssociatedUSA
soil_PrairieKansas
soil_PuertoRico
soil_RhizosphereAmazonian
soil_SASoil
soil_ShrublandCalifornia
soil_ShrublandSudan

377

Fig 5. **Proportion of sequences as oxidative stress cell response related functional proteins in the metagenomes**. Average number of sequences annotated as proteins implicated in oxidative stress cell response per 10000 annotated sequences from (**a**) all sequences, (**b**) fungal sequences and (**c**) bacterial 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 sequences annotated as desiccation response related functional proteins per 10000 sequences varied between air sites ($P=2\times10^{-5}$), with the highest values in Grenoble (4 +/- 1), Storm Peak (4 +/- 1) and Amsterdam-Island (3 +/- 3), and the lowest values in Station Nord (0.5 +/- 1) and San Diego (0.1 +/- 0.1) (**Fig S4**). It was on average higher in air compared to the other ecosystems ($P=4\times10^{-9}$). Still Svalbard snow and French agricultural soil showed high values (2 +/- 1 and 3 +/- 1, respectively) (**Fig S4**). When considering fungal sequences only, the concentration in air was higher compared to soil ($P>10^{-5}$), sediments ($P>10^{-5}$) and surface seawater ($P=10^{-3}$). No difference between the ecosystems was observed when considering bacterial sequences separately ($P=0.62$).
Two proteins (lipoate synthase and chromosome plasmid partitioning protein ParA) related to stress response showed high relative concentrations in bacterial sequences of a few air samples compared to the other ecosystems (**Fig S3**), although the number of sequences related to these proteins was on average not higher in the atmosphere than other ecosystems ($P>0.05$).

4 Discussion

Metagenomic investigations of different ecosystems revealed a specific functional potential signature of their associated microbial communities (Delmont et al., 2011; Tringe et al., 2005). These specific signatures are thought to result from microbial adaptation and/or physical selection to the environmental abiotic conditions (Hindré et al., 2012; Li et al., 2019; Rey et al., 2016) 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), that is an organosulfur compound produced by phytoplankton (Delmont et al., 2011). Our results showed a clear separation between surface 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 PCo analysis based on the microbial functional potential (**Fig 3**). For air microbiomes, the PCo analyses showed that the individual air samples did not group for each site 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 (**Fig S1**). While terrestrial air samples (GRE, NAM, CHC, STP, PDD, PDM) grouped with snow, soil and sediment samples, the marine (Amsterdam-Island), coastal (Cape Point) and arctic (Station Nord) air samples were closer to surface seawater and river water samples. 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 bacterial sequences are mainly responsible for the distribution of the samples on the PCo analysis (as observed when comparing the PCoA to that carried out with the fungal sequences only) although they were in smaller numbers compared to fungal sequences for many of the air samples (*i.e.* STP, GRE, AMS, PDD, CAP, Beijing *etc*.). The low statistical weight of fungal sequences relative to the overall sequences might be related to

their low richness in terms of functional genes that might have resulted in the spreading of the samples on the PCoA based on the fungal sequences (**Table S3**).

Metagenomes extracted from atmospheric samples taken around the planet were characterized by a relatively high percentage of fungal sequences as compared to other ecosystems even though bacterial sequences still dominated. 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**). 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 fungal sequences was also reported for air samples from Beijing, New York and San Diego and validates our DNA extraction method set-up specifically for quartz fiber filter (Dommergue et al., 2019). Similarly, the sequencing technology (Illumina MiSeq) could not have been responsible for the larger percentage of fungal sequences observed in our datasets as the Beijing and New York/San Diego air sample datasets originated from 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 agricultural soil samples evidenced that the ratio between fungal and bacterial cell number might be much higher (from 4.5 to 160 times higher for the most vegetated site Grenoble) in air than in soil (**Table S4**; see Tignat-Perrier et al., 2020 for more qPCR data on air samples). The ratio between fungal and bacterial cell number might be higher in the planetary boundary layer (PBL) than in other environments like soil (Malik et al., 2016), and thus, would explain the relatively higher percentage of fungal sequences observed in air metagenomes. High throughput sequencing allows the sequencing of a small part of the metagenomic DNA (with large fungal genomes likely to be sequenced first) and might explain why the values of the ratio between 16S and 18S rRNA gene copy numbers obtained by qPCR does not match exactly those obtained by the metagenomic sequencing approach, while they show the same trend. Our study is a preliminary metagenomic investigation of the air environment with a limited number of sequences per sample, and further studies are needed to confirm our results.

Fungi in the atmosphere are expected to be found mostly as fungal spores. While some spore and hyphae concentrations have been measured in air (Després et al., 2012), the relative concentration of fungal spores and fungal hyphae fragments and its temporal dynamics at the same site remains unknown. Our results showed that the number of sporulation-related functions was higher in air than the other ecosystems (with the exception of snow and phyllosphere). 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 (Huang and Hull, 2017). 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. **Fig 6** 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 compared to their non-atmospheric origins (*i.e.* the surrounding ecosystems) (**Fig 6**).

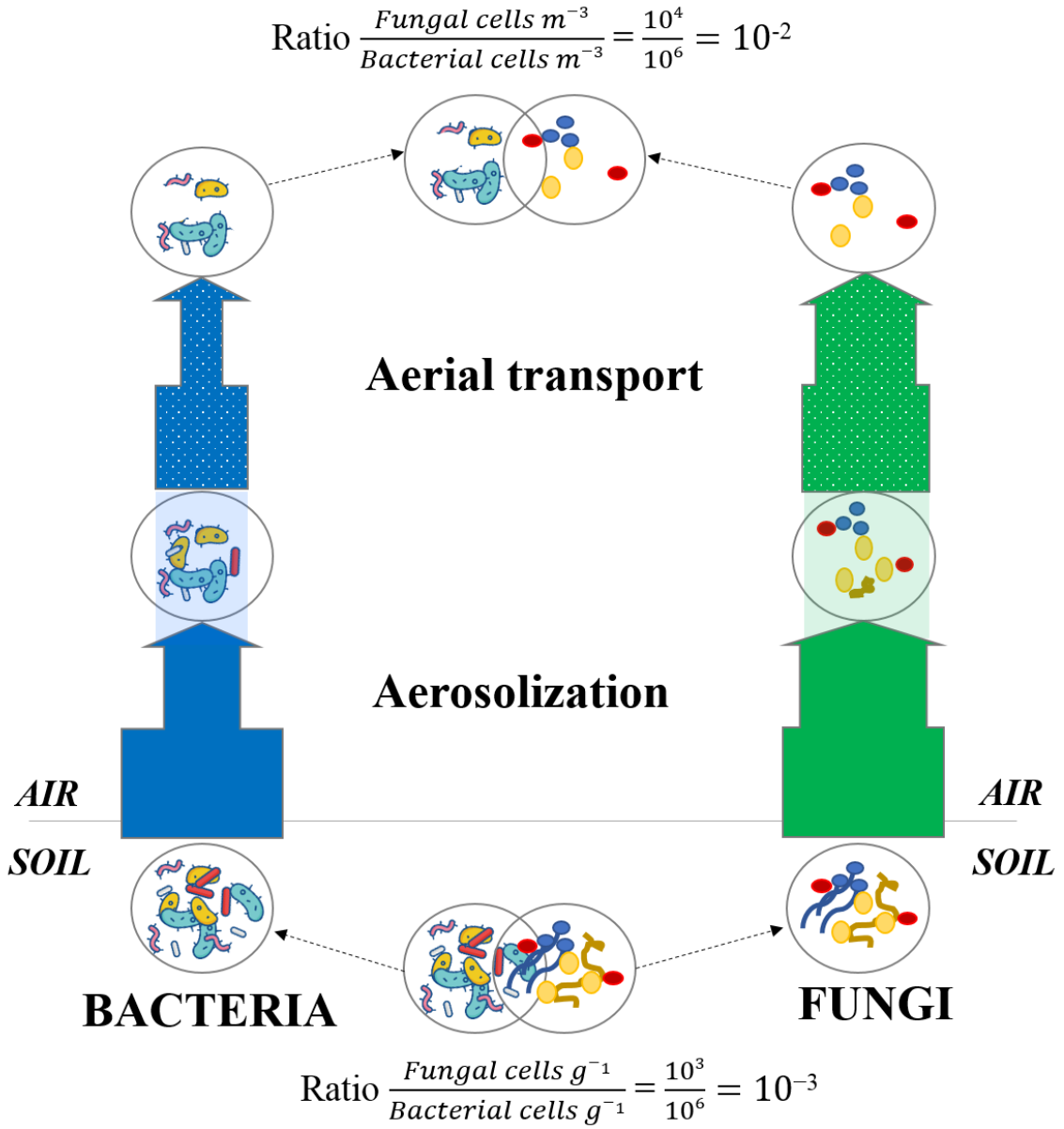


Fig 6. Microbial cell loss due to atmospheric physical stress. Conceptual model on the microbial cell loss occurring during the aerosolization and aerial transport steps due to physical selection. The thickness of the arrows represents the impact of the physical selection on both bacterial and fungal cell loss (the more microbial cells survive the physical selection, the thicker becomes the arrow). Approximate ratios are indicative and result from 16S rRNA and 18S rRNA gene qPCR data on Côte Saint André soil samples (crop soil, France) and puy de Dôme air samples (France; puy de Dôme landscape is mainly composed of croplands as shown in **Fig S1**).

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 (Tignat-Perrier et al., 2019). 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 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 microbial strains from cloud water origin have been shown to metabolize and grow on culture medium in the presence of H_2O_2 (Vaïtilingom et al., 2013), radical species and their precursors are reactive compounds and might not easily serve as energy and carbon sources for microorganisms (Imlay, 2013). 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 S3**). 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) (Joly et al., 2015; Yang et al., 2008). 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 and oxidative stress response related functions) than the other ecosystems due to the higher concentration of fungi (relatively to bacteria) in air. Thus, when the annotated sequences were separated between sequences belonging to fungal and bacterial genomes, the bacterial and fungal sequences from air samples did not show a significantly higher concentration of stress-related functions compared to the samples coming from other ecosystems (**Fig 4, 5, Fig S4**).

Fungal genomes are expected to carry genes associated to global stress-related functions (*i.e.* UV radiation, desiccation, oxidative stress), because of the innate resistance of fungi especially fungal spores. These genes associated to global stress-related functions are likely acquired during sporulation formation and certainly do not result from adaptation of fungi in air. When studying genes coding more specific proteins that are not associated to spore resistance, such as lipoate synthase and chromosome plasmid partitioning protein ParA, that might play a role in oxidative stress (Allary et al., 2007; Bunik, 2003) and are more generally found in stress resistance and adaptability of microorganisms (Shoeb et al., 2012; Zhang et al., 2018), they were occasionally found in relatively high concentration in air samples (**Fig S3**). The detection of metagenomic sequences annotated as genes coding 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 input of microbial cells to the planetary boundary layer and their relatively short residence time (a few hours to a few days based on a model assuming that microbial cells behave like non biological aerosols (Jaenicke, 1980)) might have hindered the observation of the potential adaptation (physical selection and/or microbial adaptation) of airborne microorganisms to the stressful atmospheric conditions and 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 planetary 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 6**). Another explanation might be due to the metagenomic approach that allows to sample both living and dead cells. Aerosolization has been shown to be particularly stressful and even lethal for microorganisms (Alsved et al., 2018; Thomas et al., 2011). 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 fungal 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. Meta-omics investigations on air with a deeper sequencing are needed to confirm our results and explore the functionality of atmospheric microorganisms further.

References

Aalismail, N. A., Ngugi, D. K., Díaz-Rúa, R., Alam, I., Cusack, M. and Duarte, C. M.: Functional metagenomic analysis of dust-associated microbiomes above the Red Sea, Sci Rep, 9(1), 1–12, doi:10.1038/s41598-019-50194-0, 2019.

Allary, M., Lu, J. Z., Zhu, L. and Prigge, S. T.: Scavenging of the cofactor lipoate is essential for the survival of the malaria parasite Plasmodium falciparum, Mol Microbiol, 63(5), 1331–1344, doi:10.1111/j.1365-2958.2007.05592.x, 2007.

Alsved, M., Holm, S., Christiansen, S., Smidt, M., Ling, M., Boesen, T., Finster, K., Bilde, M., Löndahl, J. and Šantl-Temkiv, T.: Effect of Aerosolization and Drying on the Viability of Pseudomonas syringae Cells, Front Microbiol, 9, 3086, doi:10.3389/fmicb.2018.03086, 2018.

Amato, P., Demeer, F., Melaouhi, A., Fontanella, S., Martin-Biesse, A.-S., Sancelme, M., Laj, P. and Delort, A.-M.: A fate for organic acids, formaldehyde and methanol in cloud water: their biotransformation by micro-organisms, Atmospheric Chemistry and Physics, 7(15), 4159–4169, doi:https://doi.org/10.5194/acp-7-4159-2007, 2007.

Amato, P., Besaury, L., Joly, M., Penaud, B., Deguillaume, L. and Delort, A.-M.: Metatranscriptomic exploration of microbial functioning in clouds, Sci Rep, 9(1), 1–12, doi:10.1038/s41598-019-41032-4, 2019.

591 Ariya, P., Sun, J., Eltouny, N., Hudson, E., Hayes, C. and Kos, G.: Physical and chemical
592 characterization of bioaerosols--Implications for nucleation processes, International Reviews in
593 Physical Chemistry, 28, 1–32, doi:10.1080/01442350802597438, 2009.

594 Ariya, P. A., Nepotchatykh, O., Ignatova, O. and Amyot, M.: Microbiological degradation of
595 atmospheric organic compounds, Geophysical Research Letters, 29(22), 34-1-34–4,
596 doi:10.1029/2002GL015637, 2002.

597 Aylor, D. E.: Spread of Plant Disease on a Continental Scale: Role of Aerial Dispersal of Pathogens,
598 Ecology, 84(8), 1989–1997, 2003.

599 Brown, J. K. M. and Hovmøller, M. S.: Aerial dispersal of pathogens on the global and continental
600 scales and its impact on plant disease, Science, 297(5581), 537–541, doi:10.1126/science.1072678,
601 2002.

602 Brune, A., Frenzel, P. and Cypionka, H.: Life at the oxic–anoxic interface: microbial activities and
603 adaptations, FEMS Microbiol Rev, 24(5), 691–710, doi:10.1111/j.1574-6976.2000.tb00567.x, 2000.

604 Bunik, V. I.: 2-Oxo acid dehydrogenase complexes in redox regulation, Eur. J. Biochem., 270(6),
605 1036–1042, doi:10.1046/j.1432-1033.2003.03470.x, 2003.

606 Cao, C., Jiang, W., Wang, B., Fang, J., Lang, J., Tian, G., Jiang, J. and Zhu, T. F.: Inhalable
607 Microorganisms in Beijing's PM2.5 and PM10 Pollutants during a Severe Smog Event, Environ. Sci.
608 Technol., 48(3), 1499–1507, doi:10.1021/es4048472, 2014.

609 Delmont, T. O., Malandain, C., Prestat, E., Larose, C., Monier, J.-M., Simonet, P. and Vogel, T. M.:
610 Metagenomic mining for microbiologists, ISME J, 5(12), 1837–1843, doi:10.1038/ismej.2011.61,
611 2011.

612 Delort, A.-M., Vaïtilingom, M., Amato, P., Sancelme, M., Parazols, M., Mailhot, G., Laj, P. and
613 Deguillaume, L.: A short overview of the microbial population in clouds: Potential roles in
614 atmospheric chemistry and nucleation processes, Atmospheric Research, 98(2), 249–260,
615 doi:10.1016/j.atmosres.2010.07.004, 2010.

616 Després, V., Huffman, J. A., Burrows, S. M., Hoose, C., Safatov, A., Buryak, G., Fröhlich-Nowoisky,
617 J., Elbert, W., Andreae, M., Pöschl, U. and Jaenicke, R.: Primary biological aerosol particles in the
618 atmosphere: a review, Tellus B: Chemical and Physical Meteorology, 64(1), 15598,
619 doi:10.3402/tellusb.v64i0.15598, 2012.

620 Dommergue, A., Amato, P., Tignat-Perrier, R., Magand, O., Thollot, A., Joly, M., Bouvier, L.,
621 Sellegri, K., Vogel, T., Sonke, J. E., Jaffrezo, J.-L., Andrade, M., Moreno, I., Labuschagne, C., Martin,
622 L., Zhang, Q. and Larose, C.: Methods to investigate the global atmospheric microbiome, Front.
623 Microbiol., 10, doi:10.3389/fmicb.2019.00243, 2019.

624 Donovan, P. D., Gonzalez, G., Higgins, D. G., Butler, G. and Ito, K.: Identification of fungi in
625 shotgun metagenomics datasets, PLOS ONE, 13(2), e0192898, doi:10.1371/journal.pone.0192898,
626 2018.

627 Els, N., Larose, C., Baumann-Stanzer, K., Tignat-Perrier, R., Keuschnig, C., Vogel, T. M. and Sattler,
628 B.: Microbial composition in seasonal time series of free tropospheric air and precipitation reveals
629 community separation, Aerobiologia, doi:10.1007/s10453-019-09606-x, 2019.

630 Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H.,
631 Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F. and Schaaf, C.: Global land
632 cover mapping from MODIS: algorithms and early results, Remote Sensing of Environment, 83(1),
633 287–302, doi:10.1016/S0034-4257(02)00078-0, 2002.

Griffin, D. W.: Atmospheric Movement of Microorganisms in Clouds of Desert Dust and Implications for Human Health, Clin Microbiol Rev, 20(3), 459–477, doi:10.1128/CMR.00039-06, 2007.

Gusareva, E. S., Acerbi, E., Lau, K. J. X., Luhung, I., Premkrishnan, B. N. V., Kolundžija, S., Purbojati, R. W., Wong, A., Houghton, J. N. I., Miller, D., Gaultier, N. E., Heinle, C. E., Clare, M. E., Vettath, V. K., Kee, C., Lim, S. B. Y., Chénard, C., Phung, W. J., Kushwaha, K. K., Nee, A. P., Putra, A., Panicker, D., Yanqing, K., Hwee, Y. Z., Lohar, S. R., Kuwata, M., Kim, H. L., Yang, L., Uchida, A., Drautz-Moses, D. I., Junqueira, A. C. M. and Schuster, S. C.: Microbial communities in the tropical air ecosystem follow a precise diel cycle, PNAS, 116(46), 23299–23308, doi:10.1073/pnas.1908493116, 2019.

Hadley, W. and Winston, C.: Create Elegant Data Visualisations Using the Grammar of Graphics, [online] Available from: https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf, 2019.

Hill, K. A., Shepson, P. B., Galbavy, E. S., Anastasio, C., Kourtev, P. S., Konopka, A. and Stirm, B. H.: Processing of atmospheric nitrogen by clouds above a forest environment, Journal of Geophysical Research: Atmospheres, 112(D11), doi:10.1029/2006JD008002, 2007.

Hindré, T., Knibbe, C., Beslon, G. and Schneider, D.: New insights into bacterial adaptation through *in vivo* and *in silico* experimental evolution, Nature Reviews Microbiology, 10(5), 352–365, doi:10.1038/nrmicro2750, 2012.

Huang, M. and Hull, C. M.: Sporulation: How to survive on planet Earth (and beyond), Curr Genet, 63(5), 831–838, doi:10.1007/s00294-017-0694-7, 2017.

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C. and Bork, P.: Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper, Mol. Biol. Evol., 34(8), 2115–2122, doi:10.1093/molbev/msx148, 2017.

Huson, D. H., Richter, D. C., Mitra, S., Auch, A. F. and Schuster, S. C.: Methods for comparative metagenomics, BMC Bioinformatics, 10 Suppl 1, S12, doi:10.1186/1471-2105-10-S1-S12, 2009.

Imlay, J. A.: The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium, Nat Rev Microbiol, 11(7), 443–454, doi:10.1038/nrmicro3032, 2013.

Innocente, E., Squizzato, S., Visin, F., Facca, C., Rampazzo, G., Bertolini, V., Gandolfi, I., Franzetti, A., Ambrosini, R. and Bestetti, G.: 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, doi:10.1016/j.scitotenv.2017.03.199, 2017.

Jaenicke, R.: Atmospheric aerosols and global climate, Journal of Aerosol Science, 11(5), 577–588, doi:10.1016/0021-8502(80)90131-7, 1980.

Joly, M., Amato, P., Sancelme, M., Vinatier, V., Abrantes, M., Deguillaume, L. and Delort, A.-M.: Survival of microbial isolates from clouds toward simulated atmospheric stress factors, Atmospheric Environment, 117, 92–98, doi:10.1016/j.atmosenv.2015.07.009, 2015.

Li, Y., Zheng, L., Zhang, Y., Liu, H. and Jing, H.: Comparative metagenomics study reveals pollution induced changes of microbial genes in mangrove sediments, Scientific Reports, 9(1), 5739, doi:10.1038/s41598-019-42260-4, 2019.

Louca, S., Doebeli, M. and Parfrey, L. W.: Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem, Microbiome, 6, doi:10.1186/s40168-018-0420-9, 2018.

676 Malik, A. A., Chowdhury, S., Schlager, V., Oliver, A., Puissant, J., Vazquez, P. G. M., Jehmlich, N.,
677 von Bergen, M., Griffiths, R. I. and Gleixner, G.: Soil Fungal:Bacterial Ratios Are Linked to Altered
678 Carbon Cycling, Front. Microbiol., 7, doi:10.3389/fmicb.2016.01247, 2016.

679 Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.
680 R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E. and Wagner, H.:
681 Community Ecology Package, [online] Available from: https://github.com/vegandevs/vegan, 2019.

682 Rey, O., Danchin, E., Mirouze, M., Loot, C. and Blanchet, S.: Adaptation to Global Change: A
683 Transposable Element–Epigenetics Perspective, Trends in Ecology & Evolution, 31(7), 514–526,
684 doi:10.1016/j.tree.2016.03.013, 2016.

685 Shannan, S., Collins, K. and Emanuel, W. R.: Global mosaics of the standard MODIS land cover type
686 data, 2014.

687 Shoeb, E., Badar, U., Akhter, J., Shams, H., Sultana, M. and Ansari, M. A.: Horizontal gene transfer
688 of stress resistance genes through plasmid transport, World J. Microbiol. Biotechnol., 28(3), 1021–
689 1025, doi:10.1007/s11274-011-0900-6, 2012.

690 Thomas, R. J., Webber, D., Hopkins, R., Frost, A., Laws, T., Jayasekera, P. N. and Atkins, T.: The
691 Cell Membrane as a Major Site of Damage during Aerosolization of Escherichia coli, Appl. Environ.
692 Microbiol., 77(3), 920–925, doi:10.1128/AEM.01116-10, 2011.

693 Tignat-Perrier, R., Dommergue, A., Thollot, A., Keuschnig, C., Magand, O., Vogel, T. M. and
694 Larose, C.: Global airborne microbial communities controlled by surrounding landscapes and wind
695 conditions, Sci Rep, 9(1), 1–11, doi:10.1038/s41598-019-51073-4, 2019.

696 Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., Podar, M.,
697 Short, J. M., Mathur, E. J., Detter, J. C., Bork, P., Hugenholtz, P. and Rubin, E. M.: Comparative
698 metagenomics of microbial communities, Science, 308(5721), 554–557, doi:10.1126/science.1107851,
699 2005.

700 Vaïtilingom, M., Amato, P., Sancelme, M., Laj, P., Leriche, M. and Delort, A.-M.: Contribution of
701 Microbial Activity to Carbon Chemistry in Clouds, Appl Environ Microbiol, 76(1), 23–29,
702 doi:10.1128/AEM.01127-09, 2010.

703 Vaïtilingom, M., Deguillaume, L., Vinatier, V., Sancelme, M., Amato, P., Chaumerliac, N. and Delort,
704 A.-M.: Potential impact of microbial activity on the oxidant capacity and organic carbon budget in
705 clouds, PNAS, 110(2), 559–564, doi:10.1073/pnas.1205743110, 2013.

706 Vartoukian, S. R., Palmer, R. M. and Wade, W. G.: Strategies for culture of 'unculturable' bacteria,
707 FEMS Microbiology Letters, 309(1), 1–7, doi:10.1111/j.1574-6968.2010.02000.x, 2010.

708 Wickham, H.: Flexibly Reshape Data: A Reboot of the Reshape Packa, [online] Available from:
709 https://cran.r-project.org/web/packages/reshape2/reshape2.pdf, 2017.

710 Wood, D. E. and Salzberg, S. L.: Kraken: ultrafast metagenomic sequence classification using exact
711 alignments, Genome Biology, 15(3), R46, doi:10.1186/gb-2014-15-3-r46, 2014.

712 Xie, W., Wang, F., Guo, L., Chen, Z., Sievert, S. M., Meng, J., Huang, G., Li, Y., Yan, Q., Wu, S.,
713 Wang, X., Chen, S., He, G., Xiao, X. and Xu, A.: Comparative metagenomics of microbial
714 communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries, The
715 ISME Journal, 5(3), 414–426, doi:10.1038/ismej.2010.144, 2011.

716 Yang, Y., Yokobori, S. and Yamagishi, A.: UV-resistant bacteria isolated from upper troposphere and
717 lower stratosphere, Biol.Sci.Space, 22, doi:10.2187/bss.22.18, 2008.

718 Yooseph, S., Nealson, K. H., Rusch, D. B., McCrow, J. P., Dupont, C. L., Kim, M., Johnson, J.,
719 Montgomery, R., Ferriera, S., Beeson, K., Williamson, S. J., Tovchigrechko, A., Allen, A. E., Zeigler,
720 L. A., Sutton, G., Eisenstadt, E., Rogers, Y.-H., Friedman, R., Frazier, M. and Venter, J. C.: Genomic
721 and functional adaptation in surface ocean planktonic prokaryotes, Nature, 468(7320), 60–66,
722 doi:10.1038/nature09530, 2010.

723 Yooseph, S., Andrews-Pfannkoch, C., Tenney, A., McQuaid, J., Williamson, S., Thiagarajan, M.,
724 Brami, D., Zeigler-Allen, L., Hoffman, J., Goll, J. B., Fadrosh, D., Glass, J., Adams, M. D., Friedman,
725 R. and Venter, J. C.: A Metagenomic Framework for the Study of Airborne Microbial Communities,
726 PLOS ONE, 8(12), e81862, doi:10.1371/journal.pone.0081862, 2013.

727 Zhang, H., Hu, Y., Zhou, C., Yang, Z., Wu, L., Zhu, M., Bao, H., Zhou, Y., Pang, M., Wang, R. and
728 Zhou, X.: Stress resistance, motility and biofilm formation mediated by a 25kb plasmid pLMSZ08 in
729 Listeria monocytogenes, Food Control, 94, 345–352, doi:10.1016/j.foodcont.2018.07.002, 2018.

730
731 **Competing interests.** The authors declare that they have no conflict of interest.
732
733 **Financial support**. This work was supported by the Agence Nationale de la Recherche
734 [ANR-15-CE01-0002–03 INHALE]; Région Auvergne-Rhône Alpes [ARC3 2016];
735 CAMPUS France [program XU GUANGQI] and the French Polar Institute IPEV [program
736 1028 and 399].
737
738 **Author contributions**. AD, CL and TMV designed the experiment. RTP, AD, AT and OM
739 conducted the sampling field campaign. RTP did the molecular biology, bioinformatics and
740 statistical analyses. RTP, AD, CL and TMV analyzed the results. RTP, TM, AD and CL wrote
741 the manuscript. All authors reviewed the manuscript.
742
743 **Acknowledgements**. The chemical analyses were performed at the IGE AirOSol platform. This
744 work was hosted by the following stations: Chacaltaya, Namco, puy de Dôme, Cape-Point, Pic-
745 du-Midi, Amsterdam-Island, Storm-Peak, Villum RS and we thank I.Jouvie, G.Hallar,
746 I.McCubbin, Benny and Jesper, B.Jensen, A.Nicosia, M.Ribeiro, L.Besaury, L.Bouvier,
747 M.Joly, I.Moreno, M.Rocca, F.Velarde for sampling and station management. We thank our
748 project partners: K.Sellegri, P.Amato, M.Andrade, Q.Zhang, C.Labuschagne and L.Martin, J.
749 Sonke. We thank R.Edwards, J. Schauer and C.Worley for lending their HV sampler. We thank
750 L.Pouilloux for computing assistance and maintenance of the Newton server.
751
752 **Data availability**. Sequences reported in this paper have been deposited in ftp://ftp-adn.ec-
753 lyon.fr/Tignat-Perrier_2020_air_metagen_INHALE/. A file has been attached explaining the
754 correspondence between file names and samples.

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

Correspondence to: Romie Tignat-Perrier (romie.tignat-perrier@univ-grenoble-alpes.fr)

1 Supplementary Materials

2 Supplementary Figures

3 Supplementary tables

1 Supplementary Materials

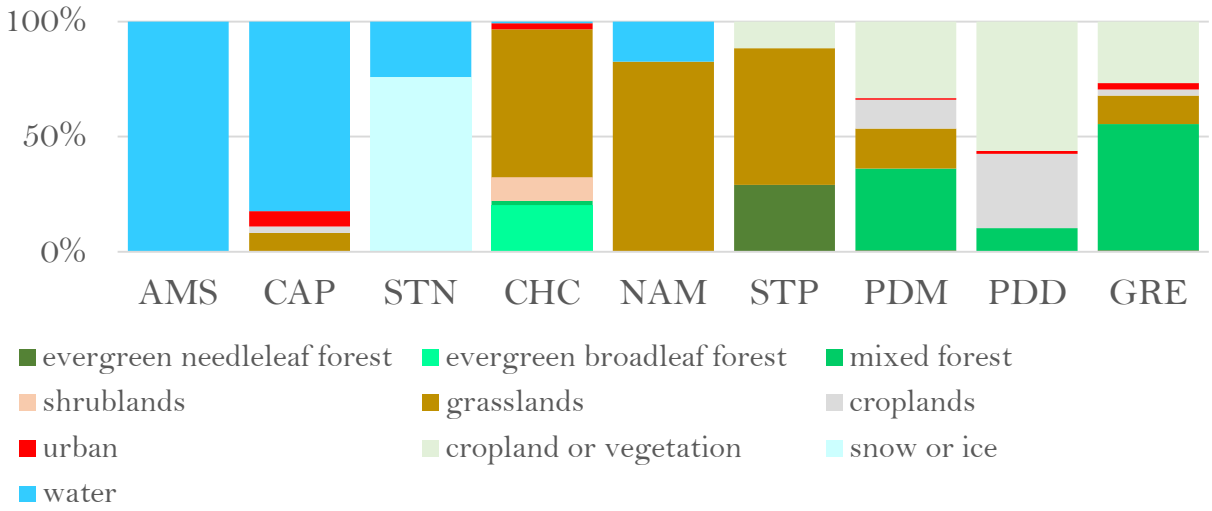


Fig S1. **Surrounding landscapes of the air sampling sites.** The proportions of different landscapes within a perimeter of 50 km have been calculated based on the land cover MODIS approach. AMS: Amsterdam-Island, CAP: Cape-Point, STN: Station Nord, CHC: Chacaltaya, NAM: Namco, STP: Storm Peak, PDM: Pic-du-Midi, PDD: puy de Dôme, GRE: Grenoble.

2 Supplementary Figures

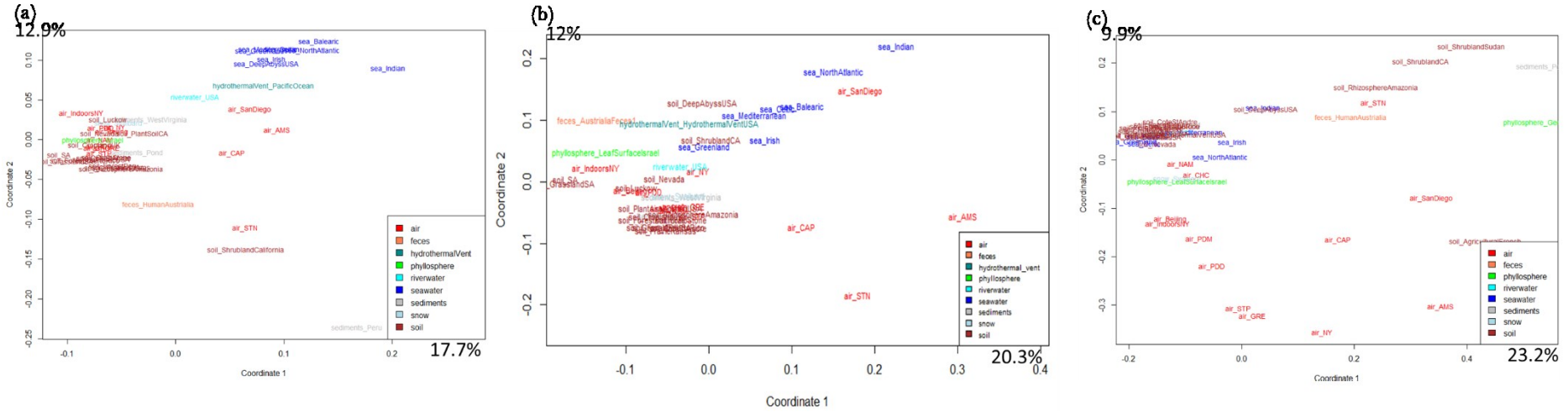


Fig S2. **Distribution of the samples based on the microbial functional profile when considering all, bacterial or fungal sequences.** PCoA analysis of the Bray-Curtis dissimilarity matrix based on the functional potential structure of each site. All sequences (**a**), bacterial sequences (**b**) and fungal 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.

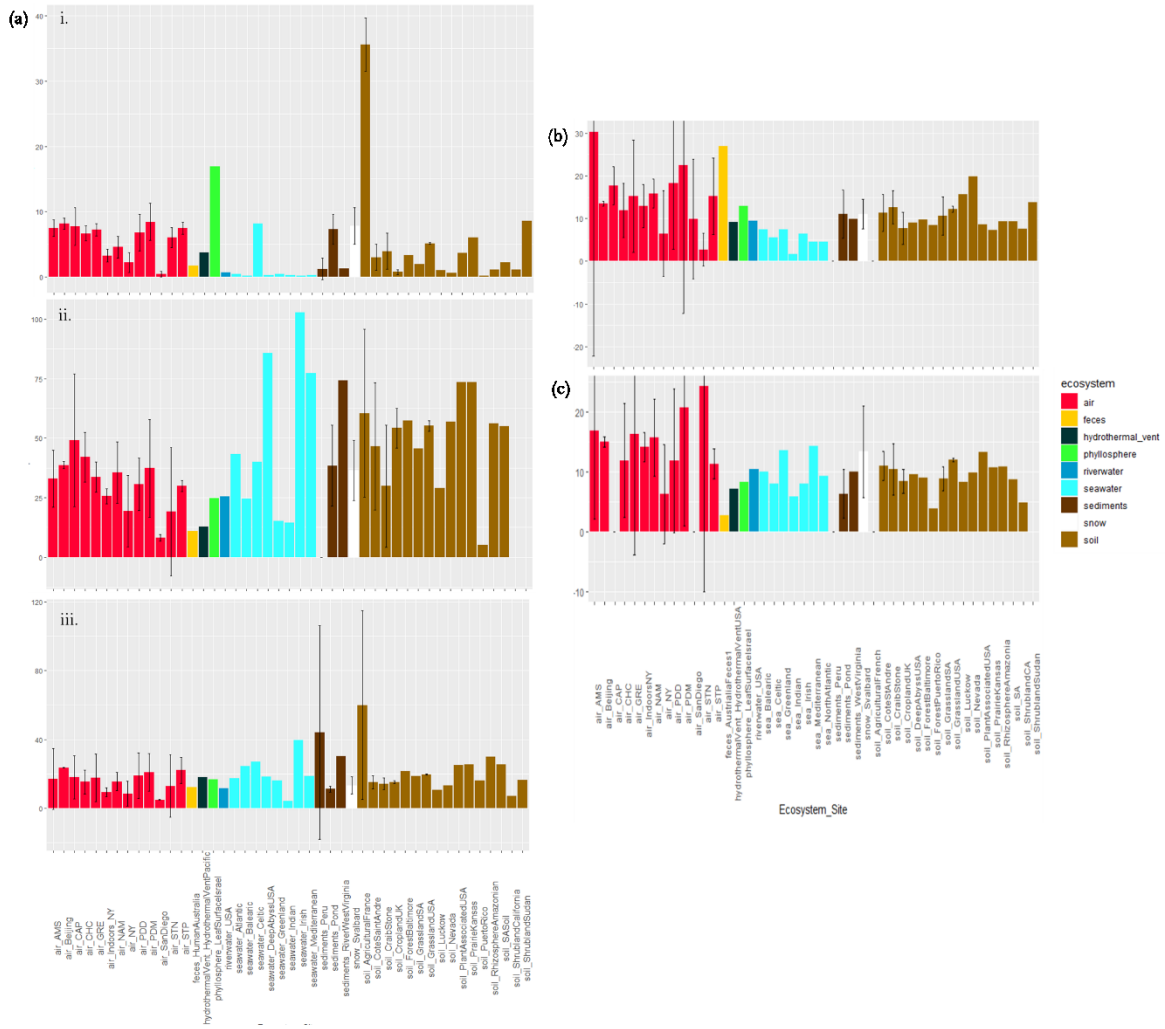


Fig S3. **Proportion of sequences annotated as hydrogen peroxide catabolic process related functional proteins as well as proteins potentially implicated in stress resistance in the metagenomes**. Average number of sequences annotated as proteins implicated in the hydrogen peroxide catabolic process (**a**, left) per 10000 annotated sequences from (**i**) all sequences, (**ii**) fungal sequences and (**iii**) bacterial sequences per site, and average number of hits of lipoate synthase (**b**) and chromosome plasmid partitioning protein ParA (**c**) 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.

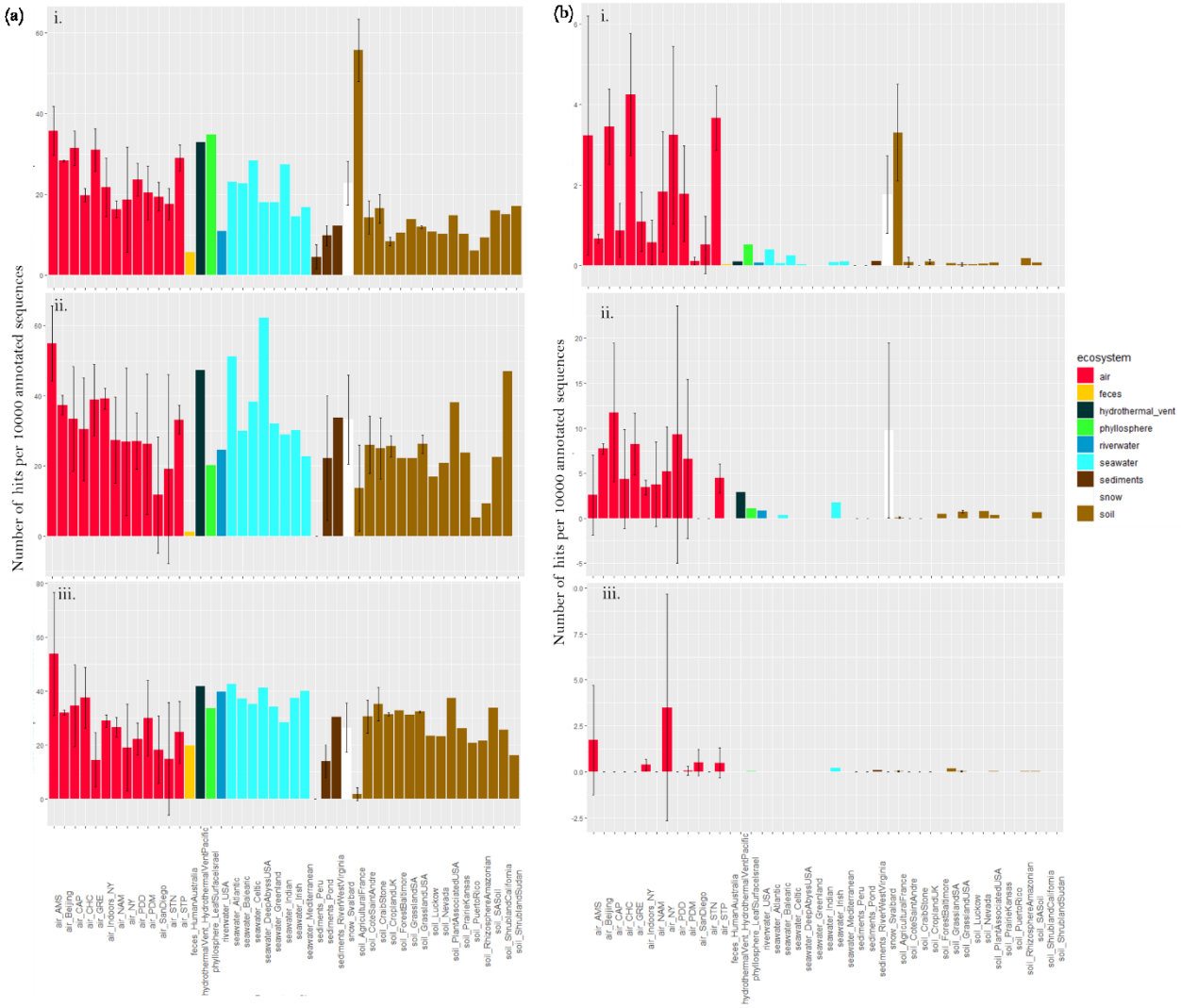


Fig S4. **Proportion of sequences annotated as UV protection and desiccation response related functional proteins in the metagenomes**. Average number of sequences annotated as proteins implicated in the UV protection (**a**, left) and desiccation response (**b**, right) per 10000 annotated sequences from (**i**) all sequences, (**ii**) fungal sequences and (**iii**) bacterial sequences per site. Colors indicate the ecosystems in which the sites belong to. For the sites including several metagenomes, the standard deviation was added.

3 Supplementary tables

Table S1. **Air sample collection characteristics.** Standardized collected air volume and sampling starting date of each air sample that we collected for this study. AMS: Amsterdam Island, CAP: Cape Point, CHC: Chacaltaya, GRE: Grenoble, NAM: Namco, PDD: Puy de Dôme, PDM: Pic du Midi, STN: Station Nord, STP: Storm Peak

Site	Sample name	Standardized collected air volume (m^3)	High volume air sampler	Sampling starting date (ending date: exactly 7 days after) (month/day/year)
AMS	AMS_10/09/2016	5232	Digitel PM10 head + pump	10/09/16
AMS	AMS_08/10/2016	4449	Digitel PM10 head + pump	08/10/16
AMS	AMS_21/10/2016	5059	Digitel PM10 head + pump	10/21/16
CAP	CAP_21/10/2016	4679	Digitel DA77 sampler	10/21/16
CAP	CAP_28/10/2016	545	Digitel DA77 sampler	10/28/16
CHC	CHC_23/09/2016	1148	Digitel PM10 head + pump	09/23/16
CHC	CHC_21/10/2016	1154	Digitel PM10 head + pump	10/21/16
CHC	CHC_28/10/2016	1220	Digitel PM10 head + pump	10/28/16
CHC	CHC_01/07/2016	1284	Digitel PM10 head + pump	01/07/16
CHC	CHC_08/07/2016	1284	Digitel PM10 head + pump	08/07/16
CHC	CHC_15/07/2016	1289	Digitel PM10 head + pump	07/15/16
CHC	CHC_12/08/2016	1160	Digitel PM10 head + pump	12/08/16
CHC	CHC_19/08/2016	1158	Digitel PM10 head + pump	08/19/16
CHC	CHC_02/09/2016	1156	Digitel PM10 head + pump	02/09/16
GRE	GRE_03/07/2017	4688	Digitel PM10 head + pump	03/07/17
GRE	GRE_10/07/2017	4717	Digitel PM10 head + pump	10/07/17
GRE	GRE_17/07/2017	4677	Digitel PM10 head + pump	07/17/17
GRE	GRE_24/07/2017	4718	Digitel PM10 head + pump	07/24/17
GRE	GRE_31/07/2017	4665	Digitel PM10 head + pump	07/31/17
GRE	GRE_07/08/2017	4762	Digitel PM10 head + pump	07/08/17
GRE	GRE_14/08/2017	4729	Digitel PM10 head + pump	08/14/17
GRE	GRE_21/08/2017	4707	Digitel PM10 head + pump	08/21/17
GRE	GRE_04/09/2017	4742	Digitel PM10 head + pump	04/09/17
NAM	NAM_17/05/2017	5511	2131 Laowin chinese sampler	05/17/17
NAM	NAM_25/05/2017	5503	2131 Laowin chinese sampler	05/25/17
NAM	NAM_02/06/2017	5513	2131 Laowin chinese sampler	02/06/17
NAM	NAM_13/06/2017	4218	2131 Laowin chinese sampler	06/13/17
NAM	NAM_20/06/2017	5418	2131 Laowin chinese sampler	06/20/17

NAM	NAM_29/06/2017	5415	2131 Laowin chinese sampler	06/29/17
NAM	NAM_07/07/2017	5483	2131 Laowin chinese sampler	07/07/17
NAM	NAM_14/07/2017	5413	2131 Laowin chinese sampler	07/14/17
NAM	NAM_21/07/2017	5465	2131 Laowin chinese sampler	07/21/17
PDD	PDD_07/06/2017	8761	Digitel PM10 head + pump	06/07/17
PDD	PDD_14/06/2017	8360	Digitel PM10 head + pump	06/14/17
PDD	PDD_21/06/2017	8672	Digitel PM10 head + pump	06/21/17
PDD	PDD_28/06/2017	9012	Digitel PM10 head + pump	06/28/17
PDD	PDD_02/08/2017	7399	Digitel PM10 head + pump	08/02/17
PDD	PDD_09/08/2017	9926	Digitel PM10 head + pump	08/09/17
PDD	PDD_30/05/2017	10232	Digitel PM10 head + pump	05/30/17
PDD	PDD_12/07/2017	8713	Digitel PM10 head + pump	07/12/17
PDD	PDD_19/07/2017	8620	Digitel PM10 head + pump	07/19/17
PDD	PDD_26/07/2017	8664	Digitel PM10 head + pump	07/26/17
PDM	PDM_20/06/2016	9664	TISCH TE-5170V sampler	06/20/16
PDM	PDM_29/06/2016	6803	TISCH TE-5170V sampler	06/29/16
PDM	PDM_12/07/2016	7550	TISCH TE-5170V sampler	12/07/16
PDM	PDM_19/07/2016	8040	TISCH TE-5170V sampler	07/19/16
PDM	PDM_26/07/2016	7794	TISCH TE-5170V sampler	07/26/16
PDM	PDM_02/08/2016	8103	TISCH TE-5170V sampler	02/08/16
PDM	PDM_09/08/2016	7747	TISCH TE-5170V sampler	08/09/16
PDM	PDM_16/08/2016	8100	TISCH TE-5170V sampler	08/16/16
PDM	PDM_23/08/2016	7956	TISCH TE-5170V sampler	08/23/16
PDM	PDM_13/09/2016	7931	TISCH TE-5170V sampler	09/13/16
PDM	PDM_20/09/2016	7853	TISCH TE-5170V sampler	09/20/16
PDM	PDM_06/09/2016	7867	TISCH TE-5170V sampler	06/09/16
PDM	PDM_16/08/2016	8100	TISCH TE-5170V sampler	08/16/16
STN	STN_27/03/2017	5153	Digitel DA80 sampler	03/27/17
STN	STN_15/05/2017	5246	Digitel DA80 sampler	05/15/17
STP	STP_14/07/2017	11213	TISCH PM10 sampler	07/14/17
STP	STP_21/07/2017	9333	TISCH PM10 sampler	07/21/17
STP	STP_28/07/2017	5702	TISCH PM10 sampler	07/28/17
STP	STP_11/08/2017	5702	TISCH PM10 sampler	08/11/17
STP	STP_18/08/2017	5702	TISCH PM10 sampler	08/18/17
STP	STP_25/08/2017	5702	TISCH PM10 sampler	08/25/17

Table S2. **Characteristics of the metagenomes.** Number of samples, ecosystem, sequencing technology, database and accession number, number of sequences per sample (mean + standard deviation), percentage of fungal and bacterial 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	present 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	present 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	present 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	present 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	present 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	present 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	present 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	present 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	present 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, Easterm 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.*, under review	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.*, under review	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 S3. **Functional richness and evenness averaged per site.** 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.

		All sequences				**Fungal sequences**				**Bacterial sequences**			
Site	Ecosystem	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. **Ratio between 16S and 18S rRNA gene copy numbers in air and soil.** qPCR on the 16s rRNA gene and on the 18S rRNA gene on air and soil samples, and ratio between these qPCRs. 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., 2019.

	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. **Dominant SEED functions in the metagenomes.** Top 50 of the SEED functions observed in the air samples (mean +/- standard deviation) considering all the sequences (*i.e.* bacterial and fungal 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

1 References

2 Tignat-Perrier, R., Dommergue, A., Thollot, A., Keuschnig, C., Magand, O., Vogel, T. M. and
3 Larose, C.: Global airborne microbial communities controlled by surrounding landscapes and wind
4 conditions, Sci Rep, 9(1), 1–11, doi:10.1038/s41598-019-51073-4, 2019.

5