Rainfalls sprinkle cloud bacterial diversity while scavenging biomass

Raphaëlle Péguilhan, Ludovic Besaury, Florent Rossi, François Enault, Jean-Luc Baray, Laurent Deguillaume, Pierre Amato

Rainfalls sprinkle cloud bacterial diversity while scavenging biomass

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

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

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

[3] Université Clermont Auvergne, CNRS, Observatoire de Physique du Globe de Clermont-Ferrand, UMS 833, F-63000 CLERMONT-FERRAND, France.

[4] Université Clermont Auvergne, CNRS, Laboratoire de Météorologie Physique, UMR 6016, F-63000 CLERMONT-FERRAND, France.

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

Keywords: Bacterial diversity, Atmosphere, Cloud, Precipitation, Scavenging, Bioaerosol dispersal

Abstract

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

Introduction

Earth' surface continuously exchanges biological material with the atmosphere. It was estimated that, globally, ~10^{24} bacteria cells are aerosolized from surface ecosystems each year (Burrows *et al.* 2009). With a modelled atmospheric residence time of up to several days and a half-life of several hours (Amato *et al.* 2015), airborne bacteria are then prone to be dispersed over regional and continental scales (Hervàs *et al.* 2009; Smith *et al.* 2013; Barberán *et al.* 2015; Griffin *et al.* 2017; Weil *et al.* 2017), carried up to high altitudes (Smith *et al.* 2018), and integrate clouds and the atmospheric water cycle (e.g., (Bauer *et al.* 2002; Amato *et al.* 2007)).

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

The composition of water reaching the surface with precipitation does not directly reflect that in the cloud. On their path to the ground, falling raindrops collect aerosol particles by below-cloud scavenging (Jaffrezo and Colin 1988; Bourcier *et al.* 2012). This is a complex phenomenon involving physical processes such as Brownian diffusion, inertial impaction and interception, whose efficiency largely depends on aerosol particle size and composition, rain drop size and rainfall intensity (Willis and Tattelman 1989; Mircea, Stefan and Fuzzi 2000; Hou *et al.* 2018; Sonwani and Kulshrestha 2019). Heavy rains (> ~10 mm h^{-1}) are in general much more efficient in scavenging aerosols particles <2.5 µm in diameter (PM2.5) than light rains (< ~1.0 mm h^{-1}), with efficiencies of ~50% versus ~5%, respectively (Luan *et al.* 2019). This discrimination leads to the enrichment or depletion of certain solutes and particles in rain water with respect to cloud water or to aerosol (Jaffrezo and Colin 1988). However, very little is known currently concerning the possibility of differential scavenging among bacteria aerosols. Interestingly for microbial ecology, a gap (Greenfield gap) in the scavenging efficiency of aerosol particles by rain was identified for particles between ~0.2 µm and 1.0 µm, i.e. the usual size of bacteria cells (Radke, Hobbs and Eltgroth 1980; Ladino *et al.* 2011; Blanco-Alegre *et al.* 2018).

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

discriminating between bacteria deposited from clouds from those that entered the water cycle after being washed-out from the air column underneath. The data help understanding bacteria dispersal and fate in the atmosphere and atmosphere-surface exchanges.

Material and methods

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

Sample collection

Samples were collected in the rural area of puy de Dôme Mountain, France, located ~400 km east from the Atlantic Ocean, and ~300 km north from the Mediterranean Sea. The surrounding landscape comprises mainly deciduous forests and pastures. Two sites separated by 12 km and 785 m in elevation were prospected for cloud water and rain sampling: puy de Dôme Mountain' summit (PUY; 1465 m a.s.l., 45.772° N, 2.9655° E) and Opme station (OPME; 680 m a.s.l., 45.7125° N, 3.090278° E), respectively (**Supplementary Figure 1**). These meteorological stations are part of the Cézeaux-Aulnat-Opme-Puy de Dôme (CO-PDD) instrumented platforms for atmospheric research (Baray *et al.* 2020). In total, 4 cloud water and 1 fresh snow samples were collected from PUY, and 10 rain samples from OPME.

Cloud water was collected during spring 2017 and autumn 2019 over periods of ~2 to 7.5 consecutive hours (**Table 1**). In 2017, only cloud droplet impactors sterilized by autoclave were used for all biological and chemical analyses as in (Amato *et al.* 2019). In 2019, three additional high-flow-rate (HFR) impingers (DS6, Kärcher SAS, Bonneuil sur Marne, France) (air-flow rate of 2 m^3/min) (Šantl-Temkiv *et al.* 2017) were deployed specifically for nucleic acid analyses (**Supplementary Figure 1 C**); other analyses were carried out from the cloud droplet impactors sampling in parallel. HFR impingers were filled with 850 mL of GF/F-filtered autoclaved nucleic acid preservation (NAP) buffer solution (Camacho-Sanchez *et al.*

2013; Menke *et al.* 2017) as the collection liquid. This contains 0.019 M of ethylenediaminetetra-acetic acid (EDTA) disodium salt dihydrate, 0.018 M of sodium citrate trisodium salt dihydrate, 3.8 M of ammonium sulfate, and H_2SO_4 to adjust the pH at 5.2. Total volumes of 309 to 1300 mL of cloud water were collected and processed immediately after sampling using the station's microbiology facility, within a laminar flow hood previously exposed to UV for 15 min.

Rain water was collected over 24-hours periods from the plateau underneath the mountain summit in November 2016, March 2017 and October-November 2019 using an automated refrigerated (4°C) rain collector NSA 181/KHS (47.4 cm diameter; surface area = 1764 cm²) (Eigenbrodt; Königsmoor, Germany) as in (Pouzet *et al.* 2017) (**Supplementary Figure 1 D**). Total volumes of water of 20 to 1124 mL were collected and processed in a laminar flow hood within 24 hours following collection.

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

Meteorological data and backward trajectory plots

Meteorological variables were monitored by the meteorological stations installed at PUY and OPME. In addition, vertical profiles of cloud liquid and ice water contents (LWC and IWC) and boundary layer height (BLH) were extracted from the ECMWF ERA5 global reanalysis (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5) (Hoffmann *et al.* 2019). The geographical origin of the air masses were obtained from 72-hours backward

trajectory plots computed using the CAT trajectory model (Baray *et al.* 2020). The model uses dynamical fields extracted from the ERA-5 meteorological data archive with, for the present work, a spatial resolution of 0.5° in latitude and longitude. This tool allowed to compute: (i) air mass backward trajectories starting from Opme (ground and cloud level) and PUY summit; (ii) air masses history, as the density of trajectory points below the BLH and the percentage of trajectory points above and below the BLH, over land and seas; (iii) the percentage of trajectory points near the CO-PDD observatory (distance < 50 km) and in each of 8 direction sectors (Renard *et al.* 2020).

Chemical analyses

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

Cell counts and ATP quantification

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

DNA extraction, amplification and sequencing

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

Bioinformatics and data analysis

Illumina reads were demultiplexed using a custom Python 3.0 script. All the 360,567 reads (average size of 395 bp) were pre-processed using Mothur software v 1.41.3 (Schloss *et al.* 2009) with the Miseq standard operating procedure (Kozich *et al.* 2013). Sequences were filtered for ambiguous bases and to a minimum length of 350 bp. A total of 329,986 sequences with an average length of 394 bp remained for further analysis. Then, the pipeline FROGS (Find Rapidly OTUs with Galaxy Solution) (Escudié *et al.* 2018) was used through the Galaxy v 3.1 environment deployed by the AuBi (Auvergne BioInformatique) network and the regional calculation cluster Mesocentre Clermont Auvergne. The tools "FROGS Clustering swarm" and "FROGS Remove chimera" were used with default parameters to cluster the sequences at 97% identity and remove possible chimera. This step removed 10.9% of the OTUs (Operational Taxonomic Units), representing 4% of the total sequences, and resulted in 82,267 OTUs. Then, "FROGS Filters" was used to select only OTUs represented by at least 3 sequences: 4,601 OTUs were kept after this step, corresponding to 6.3% of the

clusters and to 22.7% of all sequences. The "FROGS Clusters stat" tool was used at each step to obtain metrics on the clusters. Taxonomic affiliation of each OTU's seed was carried out using the "FROGS Affiliation OTU" tool with "Silva_132_16S" as the reference database (Quast *et al.* 2013). Both BLAST and RDP assignations were performed. Data were rarefied to 11,300 sequences per sample (corresponding to the sample with the lowest number of sequences) using "FROGS Abundance normalization" tool.

Abundance table was checked manually for accurate OTU affiliations. Multi-affiliations at family level, percentages of identity <95% and percentages of query coverage <98% with BLAST assignations from SILVA 132 database were verified using RDP and EzBioCloud 16S rRNA gene-based ID database (https://www.ezbiocloud.net/, update 2020.02.25) (Yoon *et al.* 2017). Thirty-eight OTUs affiliated with *Mitochondria* and 5 unaffiliated OTUs were deleted, leaving 4,510 OTUs for analysis.

OTU abundance data were then centered log-ratio (CLR)-transformed, as recommended by Gloor and colleagues to account for their compositional nature (Gloor *et al.* 2017). Data were analyzed and represented using the *R* environment 3.6.0 [4]. First, *zCompositions* package v 1.3.4 (Palarea-Albaladejo and Martín-Fernández 2015) was used for imputing zeros in our compositional count data set based on a Bayesian-multiplicative replacement with *cmultRepl* function using CZM method (count zero multiplicative) and an output format in p-counts (pseudo-counts). Then, the abundance table was transformed in centered log-ratio using *clr* function. Principal component analyses (PCA) were done based on total clustered biodiversity (4,510 OTUs) using *factoextra* v 1.0.7 (Kassambara and Mundt 2019). ANOSIM test was used to do multivariate comparison on microbial communities (*anosim* function from *vegan* v 2.5-6 (Oksanen *et al.* 2020)). Heatmaps were done using *pheatmap* v 1.0.12 (Kolde 2019) and *ggdendro* v 0.1-20 (de Vries and Ripley 2020) packages) and Venn diagrams were obtained using *VennDiagram* package v 1.6.0 (Chen and Boutros 2011). Other univariate and

multivariate statistics were performed using PAST v 3.07 (Hammer, Ryan and Harper 2001) and SIMCA 16.0 (Sartorius Stedim Biotech).

Results

Meteorological context

The main meteorological characteristics pertaining to sample collection are summarized in **Table 1** and **Supplementary Figure 2**. Sampling was conducted during fall and winter; at 3 occasions designated as events *a*, *b*, and *c*, clouds and precipitation could be collected simultaneously or within 1 day.

Samples were all collected at positive ambient temperature as liquids, at the exception of the snow sample collected at nearly -6°C. Due to the difference of elevation of almost 800 m between the two sampling sites, ambient meteorological conditions were colder and windier during cloud sampling (5.3 ±2.1°C and 9.6 ±6.1 m s^{-1}) than during rain collection (10.6°C ±4.8°C and 2.1 ±1.8 m s^{-1}). Rainfall events of different intensities were sampled, from light and moderate with average rates < 2.0 mm h^{-1}, such as event *a*, to heavy rains with maxima up to 24.0 mm h^{-1}, such as events *b* and *c*.

Based on the meteorological model ERA5, the boundary layer (BL) top altitude ranged from 441 m to 1,815 m a.s.l. on the days of rain collection and from 511 to 1,728 m a.s.l. during cloud sampling (**Supplementary Figure 2**). All rainfalls were thus collected at least partly from within the BL, while clouds and snow were sampled in the free troposphere.

Backward trajectory plots of the air masses associated with the samples indicate a wide range of geographical origins, with a general predominance of a large Western area (Atlantic Ocean) (**Supplementary Table 1; Supplementary Figure 3**). The geographical zones over which the air masses travelled at low altitude within the planetary boundary layer can be considered probable source areas of the material collected; these are shown in

Supplementary Figure 3. In most cases, the air masses at cloud altitude spent most of the last 72 hours preceding sampling in the free troposphere over marine areas (**Figure 1**). In event *a*, the air masses travelled almost exclusively at high altitude over the Atlantic Ocean before reaching the sampling sites. Event *b*'s air masses travelled at low altitude over the Channel Sea and south Great Britain Island, 500 km from there. Contrasting with other situations, event *c*'s air masses originated from South-East and were issued from continental regions in northern Africa and the Saharan desert (**Figure 2**).

Chemical signature

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

Oceanic sources-related ions were in average all significantly more concentrated in cloud water than in rain (Mann-Whitney test, $p < 0.05$, **Figure 3 A**), as observed in the past at the same sampling sites (Pouzet *et al.* 2017), by median factors of ~47, ~26 and ~12 for Na^+, Cl^- and K^+ respectively. However, there were large variations depending on samples, in relation with rainfall intensity: in event *a* (light rainfall), Na^+ concentrations in rain and cloud water were similar, while in events *b* and *c* (heavy and moderate rain), this was diluted in rain compared with cloud water, by factors of 51 and 25, respectively (see **Figure 3 B**). In turn, rain samples were characterized by relatively high contributions of Ca^{2+}, NH_4^+ and NO_3^-

attesting of continental influence. In particular, NH_4^+ and to a lesser extent Ca^{2+} tended to be more concentrated in rain than in clouds by median factors of 3.6 and 1.2 respectively, supporting consequent inputs from below cloud scavenging for these compounds. Here again, event *a*'s cloud and rain samples were more even than in events *b* and *c*, indicating lower scavenging.

In our dataset, pH ranged from 4.62 to 6.82. Acidity increased with decreasing oceanic influence (Spearman's correlations between pH and Na^+ or Cl^-, $p < 0.05$; **Supplementary Table 3 A**), so rain water tended to be more acidic than clouds.

Biomass and ATP contents

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

The raw ATP concentration varied from 340 to 842 pmol L^{-1} in cloud water and from 155 to 1,405 pmol L^{-1} in rain (**Table 2**), with no statistical difference. These values are consistent with previous reports in clouds at PUY, averaging 410 pmol L^{-1} over 28 samples (Vaïtilingom *et al.* 2012). In rainfall, ATP and cell concentrations were strongly positively linked (Spearman's correlation, p-value = 0.01, Spearman's r = 0.89; **Supplementary Table 3 B**), suggesting a large proportion of viable cells. However, the ATP-to-cell ratio was significantly higher in cloud water than in rain, by a factor of ~8.2 [(4.1 to 10.6) $\times 10^{-6}$ pmol ATP $cell^{-1}$ in rain *versus* (12.2 to 127.2) $\times 10^{-6}$ pmol ATP $cell^{-1}$ in cloudwater] (Mann-Whitney test, p-value < 0.01) (**Figure 3 A**).

Bacterial diversity

A total of 4,510 distinct prokaryotic OTUs (4,507 bacteria and 3 archaea) were detected: 246 to 600 in rain, 174 to 1,173 in clouds, and 154 in snow (**Table 2** and see complete list in **Supplementary Table 5**). These were distributed over 23 distinct phyla, 88 orders, and 435 genera whose respective distributions among clouds (363 genera) and rain samples (277 genera) are shown in **Supplementary Figure 4**.

Bacteria species richness (number of distinct OTUs) tended to be higher in clouds than in precipitation (**Table 2; Figure 3; Supplementary Figure 5)**, except for event *a* which exhibited particularly low richness likely in relation with air mass history (**Figure 1**). Richness varied independently from biomass (Spearman's correlation; p-values > 0.05; **Supplementary Table 3**). Rather, this increased with the time spent by air masses at low altitude (PLS model coefficients, **Supplementary Table 4**) and in rain, this decreased with the time spent by the air mass nearby the sampling sites (< 50 km) before sampling, supporting a distant origin of a large fraction of the richness.

The most represented phyla in the dataset (in terms of read numbers) were Proteobacteria (in particular the orders Betaproteobacteriales [eq. Burkholderiales], Pseudomonadales, Sphingomonadales and Rhizobiales), Actinobacteria (Micrococcales, Corynebacteriales and Frankiales), Firmicutes (Bacillales and Lactobacillales), Bacteroidetes and Deinococcus-Thermus. These microorganisms are commonly observed in atmospheric samples (e.g., (Amato *et al.* 2017a) and references therein). Bacteria community composition varied from one sample to another, as shown at in **Figure 5** for the whole bacteria community and in **Supplementary Figure 6** for Proteobacteria, Firmicutes and Actinobacteria. Hierarchical clustering and multivariate analyses (PCA) based on community structure (**Supplementary Figure 7**) allowed distinguishing clouds from precipitation, although 20170308CLOUD[a] tended to resemble precipitation compared to other cloud samples.

In clouds, the common core of bacteria (i.e. the taxa detected in each cloud) was composed of 20 genera (**Supplementary Table 6**). These comprised a small fraction of total cloud's richness (5% of the 363 genera) but large proportions of the reads in each sample (42% to 79%). Most taxa (215 genera, i.e., 59%) were common to at least 2 samples (**Supplementary Figure 4**). There were, logically, more specific taxa in the richest sample (20190925CLOUD, 129 specific genera) than in the poorest (20170308CLOUD[b]; 4 specific genera) (**Supplementary Figure 8**).

In rain, although more samples were collected, still numerous (33%) of the 277 genera detected were sample-specific (**Supplementary Figure 4**). This reflects the high biological variability in the atmospheric environment. As for clouds, the bacterial common core of rain samples was relatively limited in richness, with 15 genera that largely predominated (**Supplementary Table 6**; 59% to 92% of the reads).

At the genus level, about half of the bacteria detected in the study were common to clouds and rain samples (i.e. these were present in at least 1 cloud and 1 rain sample), and 155

(36%) and 69 (16%) were specific of either clouds or rain, respectively (**Figure 4**). A large proportion of rain's richness was thus contained in clouds (208 genera out of 277, 75%), whereas cloud's richness was not fully retrieved (57%) in the rain samples (**Supplementary Table 7**).

In the associated cloud-rain samples events *a* and *b*, ~21% of the total genera were common to the two types of environment (**Figure 4**). In event *c*, even at the OTU level (i.e. species) cloud and rain samples were remarkably similar, with as high as 42 % of rain's OTU comprised in the corresponding cloud, versus 4%-11% for samples collected at 1-d interval from same air masses (events *a* and *c(2)*), and 3% when the air mass origin in addition slightly differed (event *b*) (**Supplementary Table 7**). Six identified genera, among the most abundant, were detected in all cloud and rain samples (see **Supplementary Table 6**): *Massilia*, *Noviherbaspirillum*, *Pseudomonas*, *Sphingomonas*, *Undibacterium* (Proteobacteria) and *Rhodococcus* (Actinobacteria); with the exception of the latter, these were also all present in snow and thus composed the common bacterial core of this study. In turn, numerous less abundant or rare taxa were specific of clouds or rains, such as *Aerococcus*, *Oceanobacillus* and *Ammoniphilus* in clouds, and *Hymenobacter*, *Granulicella* and *Variovorax* in rain.

Certain microbial taxa were significantly more represented in clouds than in rain samples, and conversely (Kruskal-Wallis, $p \leq 0.05$). Those included in the common core of clouds and/or of rain samples are indicated in **Figure 6** (complete list in **Supplementary Table 8**). In particular, *Undibacterium* (Proteobacteria), *Bacillus* and *Staphylococcus* (Firmicutes) were more represented in clouds than in precipitation samples, while notably *Massilia*, *Sphingomonas* (Proteobacteria), as well as *Rhodococcus*, *Curtobacterium* and *Frondihabitans* (Actinobacteria) were more represented in rain; *Pseudomonas*, the genus gathering most ice nucleation active bacteria, tended to be more present in precipitation. In particular, in the events *a*, *b* and *c*, rainfall was mostly characterized by strong enrichments of *Pseudomonas*

and *Sphingomonas*, *Massilia* and *Rhodococcus*, and *Massilia* and *Sphingomonas*, respectively, as compared with the corresponding source clouds (**Supplementary Table 5**).

Identified probable marine bacteria, yet at low abundance, were more often present in clouds than in rain, such as CL500.29 marine group, *Demequina* (Park *et al.* 2016), *Marinactinospora* (Tian *et al.* 2009), *Nitratireductor* (Kang, Yang and Lee 2009), or again *Oceanobacillus*, confirming the influence of distant sources.

Discussion

Sources of biological material in atmospheric waters

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

Modeling and experimental studies have showed that bacteria and other microorganisms can travel thousands of kilometers from their emission source and connect distant ecosystems (e.g., (Burrows *et al.* 2009; Griffin *et al.* 2017; Leyronas *et al.* 2018)). In our dataset, biomass and biodiversity were overall disconnected from each other, but these could be both explained for large parts by air mass history over the 3 days preceding sampling: the fraction of time spent over continental areas for biomass, and the fraction spent at low attitude for diversity,

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

Clouds host higher metabolic activity than precipitation

ATP content in bacteria is typically in the order of ~10^{-6} pmol $cell^{-1}$ (e.g., (Amato and Christner 2009)). Here, we observed higher ATP-per-cell content in clouds than in precipitation. This could indicate higher proportions of viable cells in clouds, and/or relatively higher metabolic activity supporting clouds as microbial habitats (Sattler, Puxbaum and Psenner 2001; Ervens and Amato 2020), with potential implication for atmospheric chemistry (Khaled *et al.* 2020). The stressful conditions existing in clouds for microbial cells, such as low temperatures and high H_2O_2 concentrations, could also be responsible for increased ATP contents in cells as reported from laboratory studies (Napolitano and Shain 2004; Amato and Christner 2009; Wirgot *et al.* 2017). Many of the bacteria detected in this study were reported earlier to be active in cloud water by examining their rRNA content (Amato *et al.* 2017b). These included both abundant and rare genera, such as *Acidiphilium*, *Sphingomonas*, *Pseudomonas*, *Rickettsia*, *Curtobacterium*, *Deinococcus* and many others. Although bacteria abundance and metabolic activity are often correlated in communities, rare bacteria can be

even more active at the individual level and greatly contribute to the whole microbial activity in the ecosystem (Campbell *et al.* 2011).

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

Wet deposition fluxes of ~10^7 to nearly ~10^9 bacteria cells m^{-2} h^{-1} were quantified during rainfall periods. These are about one order of magnitude higher than the dry deposition fluxes reported on a daily basis above the boundary layer (Reche *et al.* 2018), which clearly confirms rainfall as major routes for airborne bacteria redeposition (Woo and Yamamoto 2020).

The biological similarity between cloud and rain samples was remarkably high, with ~75% of the bacteria genera present in rain also detected in clouds. This was noticeable even at deep taxonomic level in samples collected simultaneously, illustrating the strong connectivity between these consecutive steps of the water cycle. Beta-diversity can thus overall be interpreted through the ecological concepts of nestedness, with precipitation carrying a proportion of cloud's richness as one can expect, , and spatial turnover of taxa (Baselga 2010; Baselga and Leprieur 2015; Aho *et al.* 2020), illustrated by the numerous sample specific taxa. . The high proportion of sample specific taxa in particular in rain is likely related at least in part with the boundary layer/free troposphere localization of the sampling locations, and with the influence of multiple local sources acting alternately in relation with air mass movements. It is well documented indeed that the planetary boundary layer carries more material per unit volume than the free troposphere, and is more variable at small scale due to proximity with sources (Patton, Sullivan and Moeng 2005; Sasakawa *et al.* 2013).

Na^+ is emitted by marine sources and its relative contribution to the pool of dissolved ions can be used for tracking marine inputs to an air mass (Xiao *et al.* 2018). Assuming an origin

in precipitation exclusively sourced in clouds at our sampling sites, we used the $[Na^+]_{cloud}$ to $[Na^+]_{rain}$ concentration ratio to track the proportion of material originating from the source cloud in rain, and normalize biomass and richness data. We infer that 25.5%, 0.2% and 0.8% of the bacteria cells in rain originated from the corresponding source cloud in events *a*, *b* and *c* respectively, so that by far that the largest fraction of the bacterial biomass was scavenged from the air column, in particular during intense rain events (*b* and *c*). In turn, the relative bacteria richness in rain water tended to decrease with the proportion of material originating from the source cloud, pointing these latter as sources of diversity.

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

Biodiversity is not evenly distributed between clouds and precipitation

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

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

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

The possibility that bacteria could have avoided precipitation from cloud due to particular unidentified trait seems unreasonable. On the contrary, certain bacteria enriched in precipitation may have properties that could have favoured their wet deposition. Phenotypic

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

Concluding remarks

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

stability of ecosystems and tends to improve microbial fitness by spreading potentially beneficial and compatible biological innovations (Jalasvuori 2020).

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

Fundings

This work was supported by the French National Research Agency MOBIDIC project [ANR-17-MPGA-0013] to [PA] and the French National Research Agency – German Research Foundation CHLOROFILTER project [DFG KE 884/10-1, DFG KO 2912/10-1, ANR-14-CE35-0005-01] to [PA].

Acknowledgements

We thank M. Brissy and C. Ghaffar for help in the field, L. Nauton for managing computer environment, and T. Mas and V. Darbot for help with bioinformatics. We are grateful to the Mésocentre Clermont Auvergne University and AuBi platform for providing

support, computing and storage resources, and to ECMWF's computing and archive facilities. We thank also F. Conen for the loan of the automated rain collector. This study has been performed using CO-PDD instrumented site of the OPGC observatory and LaMP laboratory, under support of Université Clermont Auvergne, CNRS-INSU, and CNES. Cloud sampling was performed in the frame of PUYCLOUD observation service.

Conflicts of interest. None declared

References

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

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

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

Amato P, Brisebois E, Draghi M *et al.* Main Biological Aerosols, Specificities, Abundance, and Diversity. In: Delort A-M, Amato P (eds.). *Microbiology of Aerosols*. John Wiley & Sons, Inc., 2017a, 1–21.

Amato P, Christner BC. Energy Metabolism Response to Low-Temperature and Frozen Conditions in Psychrobacter cryohalolentis. *Appl Environ Microbiol* 2009;**75**:711–8.

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

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

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

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

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

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

Baselga A, Leprieur F. Comparing methods to separate components of beta diversity. *Methods in Ecology and Evolution* 2015;**6**:1069–79.

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

Blanco-Alegre C, Castro A, Calvo AI *et al.* Below-cloud scavenging of fine and coarse aerosol particles by rain: The role of raindrop size. *Quarterly Journal of the Royal Meteorological Society* 2018;**144**:2715–26.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Ervens B, Amato P. The global impact of bacterial processes on carbon mass. *Atmospheric Chemistry and Physics* 2020;**20**:1777–94.

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

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

Fierer N, Liu Z, Rodríguez-Hernández M *et al.* Short-Term Temporal Variability in Airborne Bacterial and Fungal Populations. *Appl Environ Microbiol* 2008;**74**:200–7.

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

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

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

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

Hammer Ø, Ryan P, Harper D. PAST: Paleontological Statistics software package for education and data analysis. *Palaeontologia Electronica* 2001;**4**:9.

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

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

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

Huffman JA, Pöhlker C, Prenni AJ *et al.* High concentrations of biological aerosol particles and ice nuclei during and after rain. *Atmospheric Chemistry and Physics Discussions* 2013;**13**:1767–93.

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

Jaffrezo J-L, Colin J-L. Rain-aerosol coupling in urban area: Scavenging ratio measurement and identification of some transfer processes. *Atmospheric Environment (1967)* 1988;**22**:929–35.

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

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

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

Kämpfer P, Rosselló-Mora R, Hermansson M *et al.* Undibacterium pigrum gen. nov., sp. nov., isolated from drinking water. *International Journal of Systematic and Evolutionary Microbiology,* 2007;**57**:1510–5.

Kang HS, Yang HL, Lee SD. Nitratireductor kimnyeongensis sp. nov., isolated from seaweed. *International Journal of Systematic and Evolutionary Microbiology* 2009;**59**:1036–9.

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

Khaled A, Zhang M, Amato P *et al.* Biodegradation by bacteria in clouds: An underestimated sink for some organics in the atmospheric multiphase system. *Atmospheric Chemistry and Physics Discussions* 2020:1–32.

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

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

Ladino L, Stetzer O, Hattendorf B *et al.* Experimental Study of Collection Efficiencies between Submicron Aerosols and Cloud Droplets. *J Atmos Sci* 2011;**68**:1853–64.

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

Leyronas C, Morris CE, Choufany M *et al.* Assessing the Aerial Interconnectivity of Distant Reservoirs of Sclerotinia sclerotiorum. *Front Microbiol* 2018;**9**, DOI: 10.3389/fmicb.2018.02257.

Lindow SE, Arny DC, Upper CD. Distribution of ice nucleation-active bacteria on plants in nature. *Applied and environmental microbiology* 1978;**36**:831–8.

Luan T, Guo X, Zhang T *et al.* Below-Cloud Aerosol Scavenging by Different-Intensity Rains in Beijing City. *J Meteorol Res* 2019;**33**:126–37.

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

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

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

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

Moore RA, Hanlon R, Powers C *et al.* Scavenging of Sub-Micron to Micron-Sized Microbial Aerosols during Simulated Rainfall. *Atmosphere* 2020;**11**:80.

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

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

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

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

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

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

Palarea-Albaladejo J, Martín-Fernández JA. zCompositions — R package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics and Intelligent Laboratory Systems* 2015;**143**:85–96.

Park S, Jung Y-T, Won S-M *et al.* Demequina litorisediminis sp. nov., isolated from a tidal flat, and emended description of the genus Demequina. *International Journal of Systematic and Evolutionary Microbiology*, 2016;**66**:4197–203.

Patton EG, Sullivan PP, Moeng C-H. The Influence of Idealized Heterogeneity on Wet and Dry Planetary Boundary Layers Coupled to the Land Surface. *J Atmos Sci* 2005;**62**:2078–97.

Pouzet G, Peghaire E, Aguès M *et al.* Atmospheric Processing and Variability of Biological Ice Nucleating Particles in Precipitation at Opme, France. *Atmosphere* 2017;**8**:229.

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

Radke LF, Hobbs PV, Eltgroth MW. Scavenging of Aerosol Particles by Precipitation. *J Appl Meteor* 1980;**19**:715–22.

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

Renard P, Bianco A, Baray J-L *et al.* Classification of Clouds Sampled at the Puy de Dôme Station (France) Based on Chemical Measurements and Air Mass History Matrices. *Atmosphere* 2020;**11**:732.

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

Šantl-Temkiv T, Amato P, Gosewinkel U *et al.* High-Flow-Rate Impinger for the Study of Concentration, Viability, Metabolic Activity, and Ice-Nucleation Activity of Airborne Bacteria. *Environmental Science and Technology* 2017;**51**:11224–34.

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

Sarmiento-Vizcaíno A, Espadas J, Martín J *et al.* Atmospheric Precipitations, Hailstone and Rainwater, as a Novel Source of Streptomyces Producing Bioactive Natural Products. *Front Microbiol* 2018;**9**, DOI: 10.3389/fmicb.2018.00773.

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

Sattler B, Puxbaum H, Psenner R. Bacterial growth in supercooled cloud droplets. *Geophysical Research Letters* 2001;**28**:239–42.

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

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

Smith DJ, Ravichandar JD, Jain S *et al.* Airborne Bacteria in Earth's Lower Stratosphere Resemble Taxa Detected in the Troposphere: Results From a New NASA Aircraft Bioaerosol Collector (ABC). *Front Microbiol* 2018;**9**, DOI: 10.3389/fmicb.2018.01752.

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

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

Tian X-P, Tang S-K, Dong J-D *et al.* Marinactinospora thermotolerans gen. nov., sp. nov., a marine actinomycete isolated from a sediment in the northern South China Sea. *International Journal of Systematic and Evolutionary Microbiology* 2009;**59**:948–52.

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

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

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

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

de Vries A de, Ripley BDR (author of package. *Ggdendro: Create Dendrograms and Tree Diagrams Using "Ggplot2."*, 2020.

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

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

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

Willis PT, Tattelman P. Drop-Size Distributions Associated with Intense Rainfall. *J Appl Meteor* 1989;**28**:3–15.

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

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

Xiao H-W, Xiao H-Y, Shen C-Y *et al.* Chemical Composition and Sources of Marine Aerosol over the Western North Pacific Ocean in Winter. *Atmosphere* 2018;**9**:298.

Yoon SH, Ha SM, Kwon S *et al.* Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *International Journal of Systematic and Evolutionary Microbiology* 2017;**67**:1613–7.

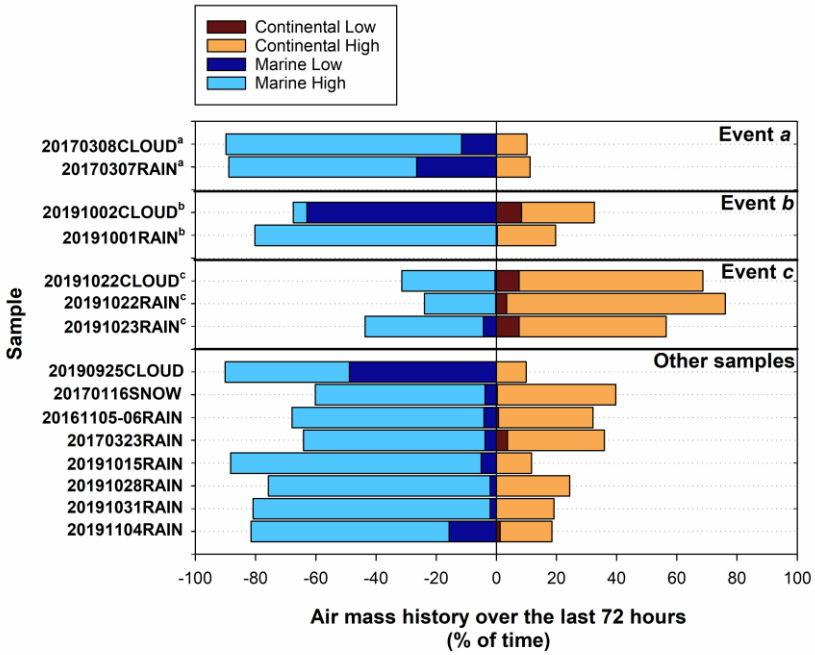

Figure 1: Air masses' history over the 72 hours preceding sampling, expressed as % of the time spent over continental (right, brown) or marine (left, blue) surfaces, in the free troposphere (High) or in the planetary boundary layer (Low). Data were extracted from ERA5 backward trajectory plots at cloud attitude for rain samples or PUY altitude for clouds and

snow. The % of time spent over marine surfaces is shown as negative values for graphical representation. *Superscripted letters indicate chronological associations between cloud and rain samples.

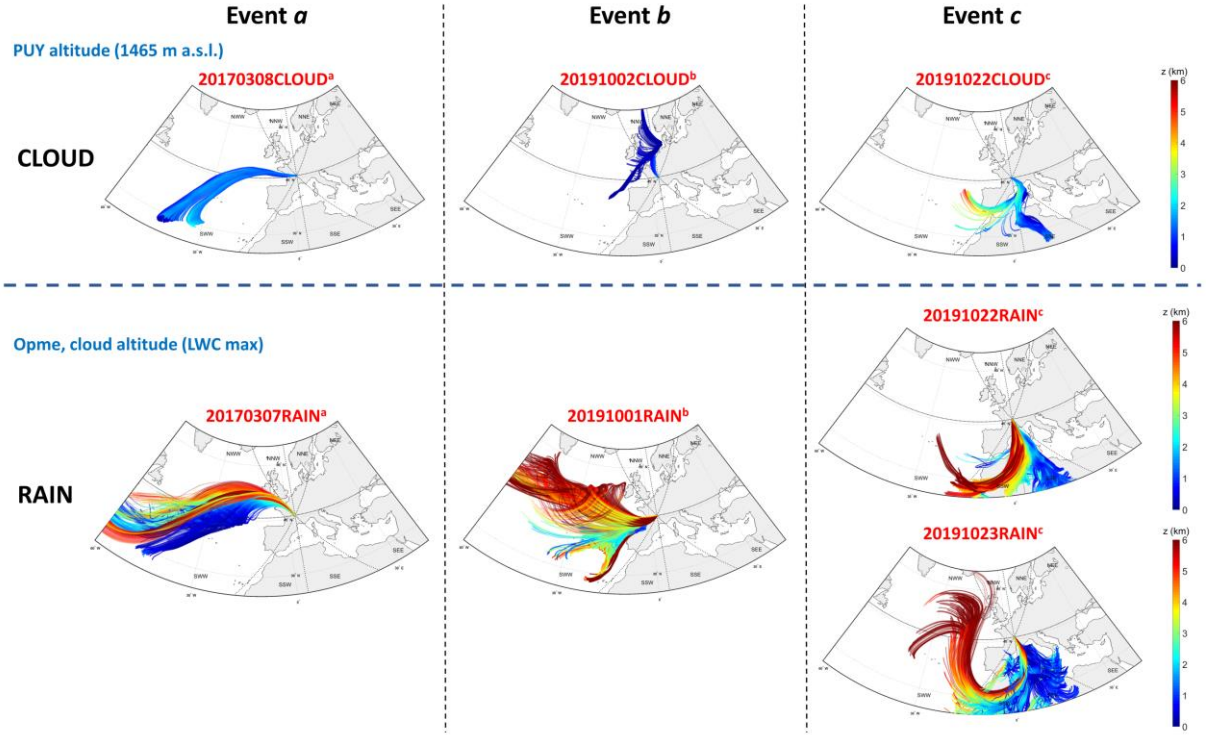

Figure 2: Seventy-two-hours backward trajectory plots at cloud level, extracted from ERA5 data reanalysis, for the associated cloud and rain samples in events *a*, *b* and *c*. See Supplementary Figure 3 for other samples and the potential corresponding source areas identified.

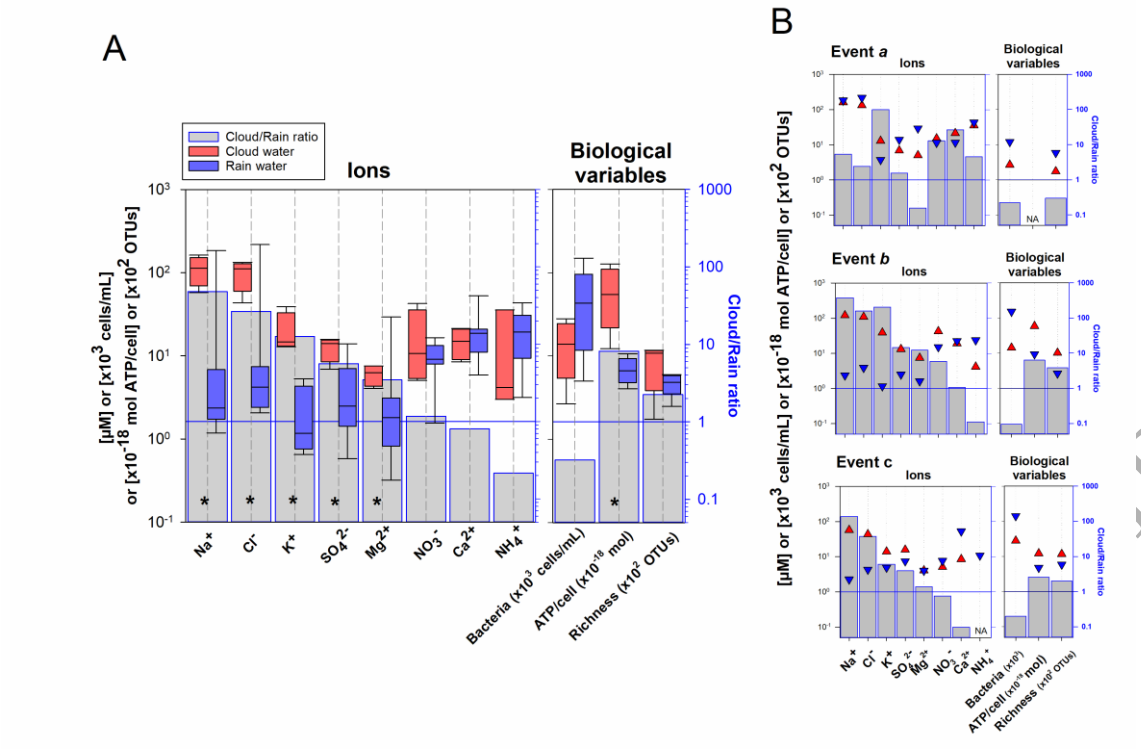

Figure 3: Absolute concentrations (left axes) and median ratios (right axes) of the main ions and biological variables in cloud and rain water samples, (as medians, 10th, 25th, 75th and 90th percentiles). **A**- all samples together; **B**- events *a*, *b* and *c*. Note the logarithmic scales on both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

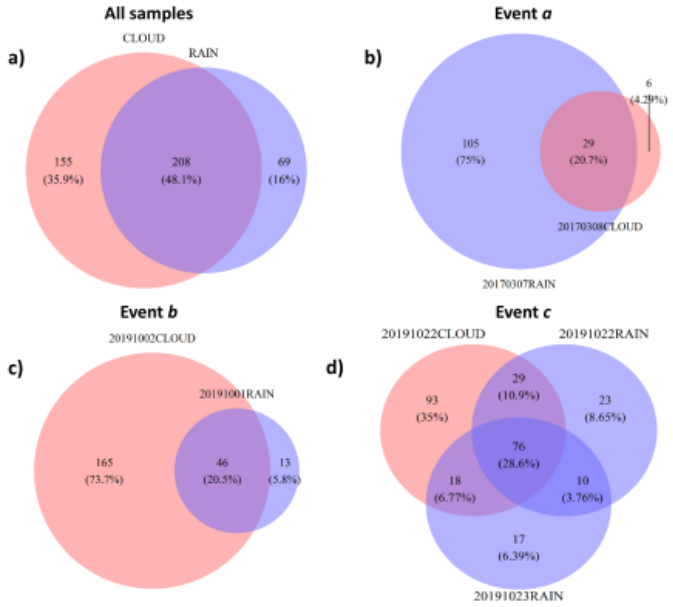

Figure 4: Venn diagrams depicting the distribution of distinct bacteria genera between cloud and rain samples: a) all samples included; b), c) and d) for the associated cloud-rain samples in events *a*, *b* and *c*.

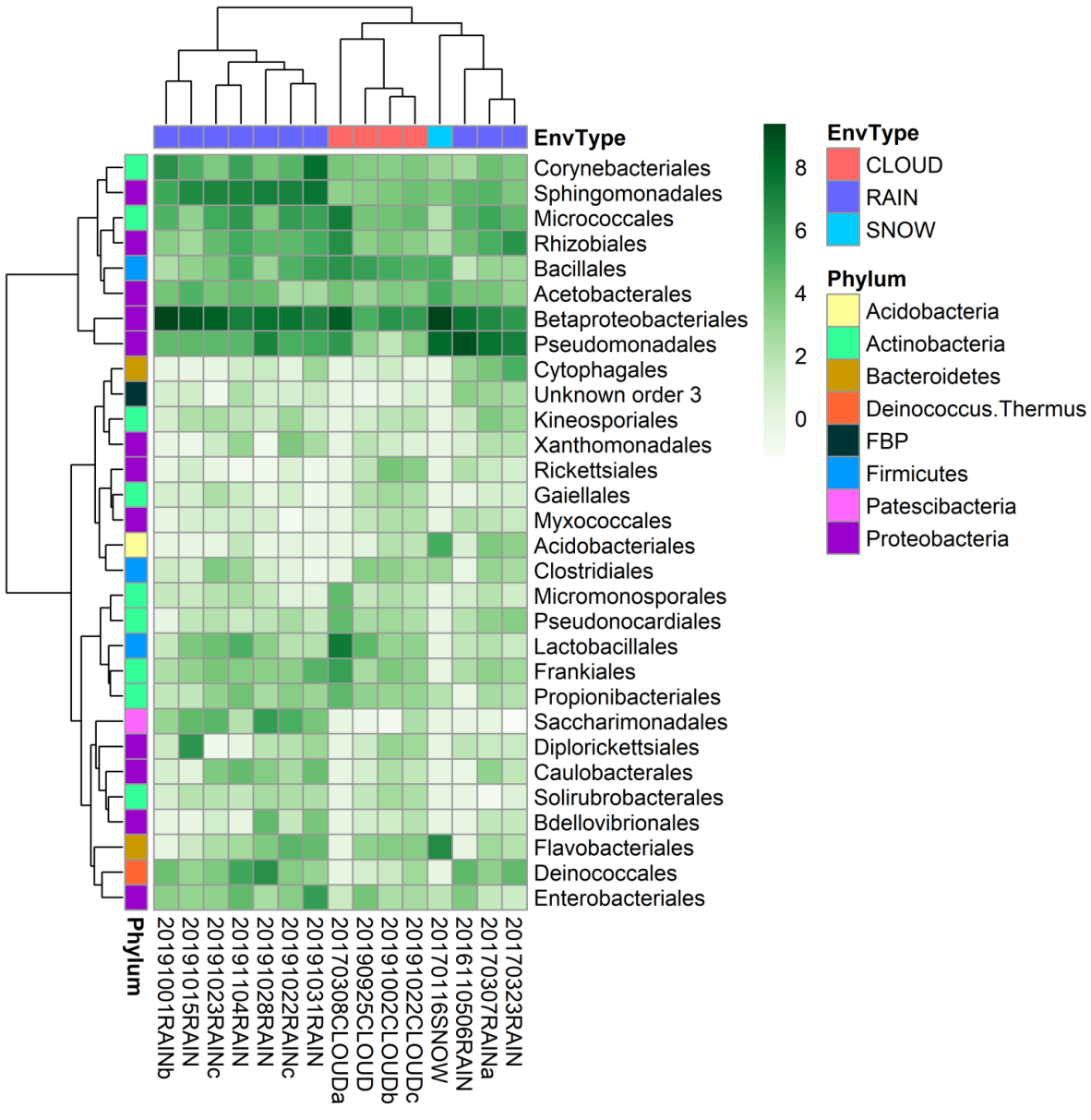

Figure 5: Distribution of the 30 most represented bacteria orders, and corresponding hierarchical clusterings (Ward's method). Intensity scale depicts centered-log ratio abundances; *a*, *b* and *c* letters indicate chronological associations between cloud and rain samples.

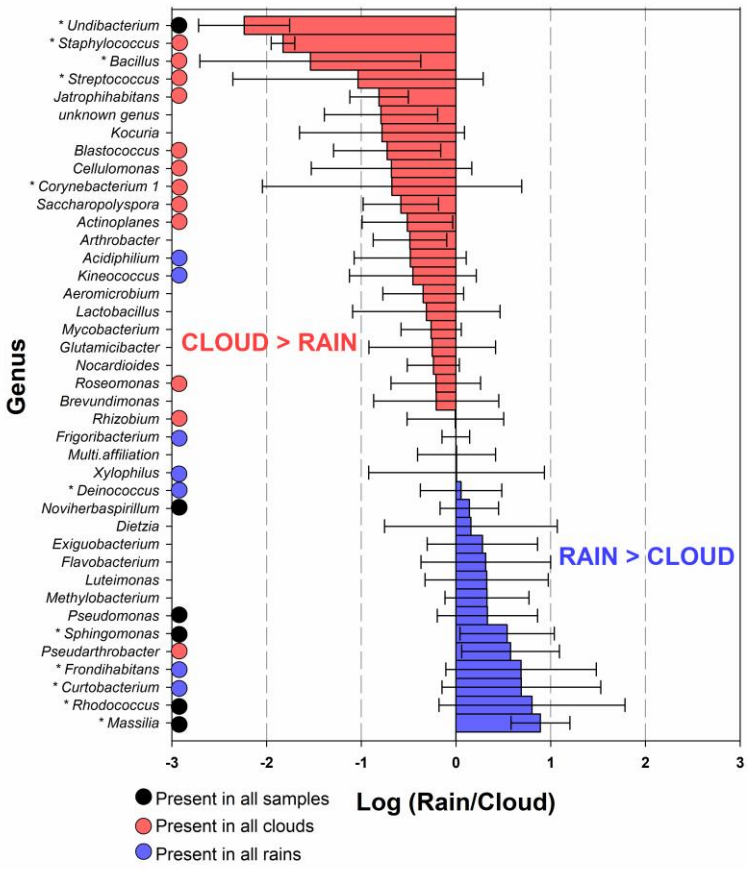

Figure 6: Average rain-to-cloud log-ratio representation in the events *a*, *b* and *c* for the 40 bacteria genera represented by >100 reads, out of the 135 distinct genera detected in total in this set of samples. The genera more represented in clouds than in rains are depicted in red,

the genera more represented in rain in blue. Error-bars represent standard deviations from the mean rain-to-cloud ratio. The genera present in all the samples (common core), in all clouds and in all rain samples of the study are indicated by black, red and blue dots respectively. Asterisks on genera names indicate those whose representation was significantly different between clouds and rains, considering all the samples of the study (Kruskal-Wallis test, p-value < 0.05).

Table 1: Meteorological contexts during sampling.

SampleID*	Sampling location – altitude (m a.s.l.)	Sampling date (mm/dd/yyyy)	Cumulated precipitation or cloud sampling duration (h)	Average temperature (°C)[#]	Wind speed (average [max]) (m s^{-1})[#]	Precipitation rate (average [max]) (mm/h)[#]	Boundary layer altitude (min-max [average]) (m)[# ‡]	Main geographical origin[¤]
SNOW								
20170116SNOW	PUY – 1465 m	1/16/2017	0.5	-5.7	3.2 [4.6]	-	844-1012 [928]	NW
RAIN								
20161105-06RAIN	Opme – 680 m	11/5 and 11/6/2016	23	7.2	NA**	2.5 [12.0]	510-1318 [925]	W
20170307RAIN[a]	Opme – 680 m	3/7/2017	3	3.4	5.2 [9.1]	1.3 [3.6]	540-1808 [1399]	W
20170323RAIN	Opme – 680 m	3/23/2017	8	6.5	1.1 [3.5]	2.9 [24.0]	557-1346 [856]	SW
20191001RAIN[b]	Opme – 680 m	10/1/2019	6	20.1	2.3 [5.0]	4.4 [24.0]	1035-1406 [1165]	W
20191015RAIN	Opme – 680 m	10/15/2019	15	NA**	NA**	4.3 [14.4]	727-1532 [1196]	W
20191022RAIN[c]	Opme – 680 m	10/22/2019	14	10.0	0.3 [2.4]	1.5 [7.2]	444-777 [569]	S
20191023RAIN[c]	Opme – 680 m	10/23/2019	13.5	11.4	1.5 [4.7]	2.7 [24.0]	481-1285 [835]	SE
20191028RAIN	Opme – 680 m	10/28/2019	12	14.2	0.3 [2.9]	1.5 [3.6]	441-864 [491]	W
20191031RAIN	Opme – 680 m	10/31/2019	9	12.4	1.6 [5.1]	1.3 [6.0]	449-1235 [760]	W
20191104RAIN	Opme – 680 m	11/4/2019	19	9.9	4.3 [7.2]	1.9 [8.4]	889-1815 [1476]	W
CLOUDS								
20170308CLOUD[a]	PUY – 1465 m	3/8/2017	7.5	2.9	14.9 [19.5]	-	798-1384 [1174]	W
20190925CLOUD	PUY – 1465 m	9/25/2019	3.3	6.6	10.8 [13.0]	-	1105-1728 [1321]	W
20191002CLOUD[b]	PUY – 1465 m	10/2/2019	2.4	6.5	3.0 [4.6]	-	1263-1318 [1296]	NW
20191022CLOUD[c]	PUY – 1465 m	10/22/2019	6.4	NA**	NA**	-	511-777 [623]	SE

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

Table 2: Biological characteristics of the samples.

SampleID*	Bacteria cell number concentration (cells mL^{-1} of water)	Bacteria wet deposition rate (cells m^{-2} h^{-1})	ATP concentration (pmol L^{-1})	Bacteria richness (number of distinct OTUs)
SNOW				
20170116SNOW	3.29×10^{4}	-	NA**	154
RAIN				
20161105-06RAIN	4.30×10^{3}	1.08×10^{7}	NA**	246
20170307RAIN[a]	1.19×10^{4}	1.54×10^{7}	NA**	582
20170323RAIN	1.12×10^{4}	3.26×10^{7}	NA**	600
20191001RAIN[b]	1.51×10^{5}	6.63×10^{8}	1405	264
20191015RAIN	1.84×10^{4}	7.91×10^{7}	155	380
20191022RAIN[c]	1.41×10^{5}	2.11×10^{8}	673	586
20191023RAIN[c]	6.70×10^{4}	1.81×10^{8}	271	531
20191028RAIN	8.14×10^{4}	1.22×10^{8}	431	433
20191031RAIN	6.64×10^{4}	8.64×10^{7}	443	393
20191104RAIN	2.01×10^{4}	3.83×10^{7}	214	529
Average rain	$\mathbf{5.72\times10^{4}}$	$\mathbf{1.44\times10^{8}}$	**513.1**	**454**
Standard error rain	$\mathbf{5.41\times10^{4}}$	$\mathbf{1.95\times10^{8}}$	**430.0**	**132**
CLOUDS				
20170308CLOUD[a]	2.67×10^{3}	-	340	174
20190925CLOUD	1.36×10^{4}	-	682	1140
20191002CLOUD[b]	1.43×10^{4}	-	842	1019
20191022CLOUD[c]	2.81×10^{4}	-	344	1173
Average clouds	$\mathbf{1.46\times10^{4}}$	-	**552.1**	**877**
Standard error clouds	$\mathbf{1.04\times10^{4}}$	-	**251.5**	**473**

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