Environment International 154 (2021) 106551

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

The aerobiome uncovered: Multi-marker metabarcoding reveals potential drivers of turn-over in the full microbial community in the air

G. Arjen de Groot [a], Stefan Geisen [b,*], E.R. Jasper Wubs [c], Liz Meulenbroek [a], Ivo Laros [a], L. Basten Snoek [b,d], Dennis R. Lammertsma [a], Lars H. Hansen [e,f], Pieter A. Slim [a]

[a] Wageningen Environmental Research, Wageningen University & Research, PO Box 47, 6700 AA Wageningen, the Netherlands
[b] Laboratory of Nematology, Wageningen University, 6700 AA Wageningen, the Netherlands
[c] Sustainable Agroecosystems Group, Institute of Agricultural Sciences, Department of Environmental Systems Science, ETH Zürich, Universitätstrasse 2, 8092 Zürich, Switzerland
[d] Theoretical Biology and Bioinformatics, Institute of Biodynamics and Biocomplexity Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands
[e] Department of Environmental Science, Aarhus University, Frederiksborgvej 399, 4000 Roskilde, Denmark
[f] Department of Plant and Environmental Science, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

ARTICLE INFO

Handling Editor: Guo-Ping Sheng

Keywords:
Aerobiome monitoring
Microbiome
Microbial dispersal
Potential pathogens
Early-warning system
Meteorology
Co-dispersal

ABSTRACT

Air is a major conduit for the dispersal of organisms at the local and the global scale. Most research has focused on the dispersal of plants, vertebrates and human disease agents. However, the air represents a key dispersal medium also for bacteria, fungi and protists. Many of those represent potential pathogens of animals and plants and have until now gone largely unrecorded. Here we studied the turnover in composition of the entire aerobiome, the collective diversity of airborne microorganisms. For that we performed daily analyses of all prokaryotes and eukaryotes (including plants) using multi-marker high-throughput sequencing for a total of three weeks. We linked the resulting communities to local weather conditions, to assess determinants of aerobiome composition and distribution. We observed hundreds of microbial taxa, mostly belonging to spore-forming organisms including fungi, but also protists. Additionally, we detected many potential human- and plant-pathogens. Community composition fluctuated on a daily basis and was linked to concurrent weather conditions, particularly air pressure and temperature. Using network analyses, we identified taxonomically diverse groups of organisms with correlated temporal dynamics. In part, this was due to co-variation with environmental conditions, while we could also detect specific host-parasite interactions. This study provides the first full inventory of the aerobiome and identifies putative drivers of its dynamics in terms of taxon composition. This knowledge can help develop early warning systems against pathogens and improve our understanding of microbial dispersal.

1. Introduction

The atmosphere is arguably the most important medium of dispersal for many terrestrial organisms (Adams et al. 2013; Muñoz et al. 2004), as well as for plant pollination (Frankel and Galun 2012). While microorganisms dominate the Earth's soils and waters in numbers, biomass (Bar-On et al. 2018) and diversity (Hawksworth and Lucking 2017), most research on capabilities for aerial dispersal has been restricted to land plants and macroscopic animals (Clobert et al., 2001). Nevertheless, understanding microbial dispersal is of vital importance. Many bacteria, fungi and protists are potential plant- and animal-pathogens or beneficial symbionts, thereby influencing agricultural yields and animal health. Besides, airborne fungal spores and bacteria can drive rainfall patterns, by acting as nuclei around which water droplets and ice crystals are formed (Möhler et al. 2007). Changing climatic conditions are altering the dispersal of microbes, which in turn affects disease epidemiology (Fisher et al. 2012) as well as local community composition. Thus, a more complete understanding of aerial dispersal of microbes, and how this depends on weather conditions, is a critical component of better predictions of pathogenic colonization and population dynamics (Brown and Hovmøller 2002; Fröhlich-Nowoisky et al. 2009), and thereby potential (agro-)economic and health risks.

Hitherto, studies of airborne microbes have focused on bacteria and/or fungi, while the entire microbial diversity in the air including protists

* Corresponding author.
E-mail address: Stefan.geisen@wur.nl (S. Geisen).

https://doi.org/10.1016/j.envint.2021.106551
Received 16 January 2021; Received in revised form 23 March 2021; Accepted 28 March 2021
Available online 12 April 2021

and micro-sized plant material and flying arthropods, which we collectively designate as the aerobiome, has not yet been targeted. Most previous studies investigated abundances and community compositions of individual groups, particularly fungi (Bowers et al. 2013; Fröhlich-Nowoisky et al. 2009; Yamamoto et al. 2012) and bacteria (Bowers et al. 2013; Bowers et al. 2011a). Integrated assessments of the full aerobiome diversity is important as taxonomically diverse microbes may disperse together, either through physical attachment or via a shared set of conditions conducive to dispersal (Bass et al. 2019). As such, one key microbiome part is always missing in contemporary aerial and most non-aerial microbiome studies: protists. So far, protists have only been targeted in a few studies focusing solely on potential human-pathogenic organisms and using low-throughput culture-based analyses (Astorga et al. 2011; Rivera et al. 1992). A key protistan group is often considered separately, as it was traditionally considered to be fungal-like: the often plant pathogenic Oomycetes (Genitsaris et al. 2014). Molecular methods have now made it possible to study protistan diversity. For example, high throughput sequencing (HTS) approaches have revealed the presence of a largely unknown diversity of protists in e.g. water (de Vargas et al. 2015) and soils (Mahé et al. 2017), while complete microbiome analyses including bacteria, fungi and protists have also recently started to be explored together (Xiong et al. 2020). However, the full range of airborne prokaryotic and eukaryotic organisms from across the tree of life (i.e. the aerobiome) including potential links between them remain unknown.

While the data are still scant, it is now clear that microbial distribution patterns do not follow the commonly cited "everything is everywhere" hypothesis (Baas-Becking 1934; Foissner 2006; Rout and Callaway 2012). The data suggest that despite their small sizes, even dispersal via air currents does not guarantee transportation of all microbes equivalently to all environments (Foissner 2008) and that arrival frequencies are sufficiently low to be outpaced by local diversification through natural selection (Hellweger et al. 2014). Next to direct uplift and transport via wind, co-dispersal with other organisms can add importantly to the dispersal capability of microorganisms. For instance, co-dispersal can enable free-living bacteria to reach suitable establishment sites when moving along with amoeba predators (Brock et al. 2011), and micro-sized plant material (fragments, pollen or seeds) may take along (facultative) endophytes (Greub and Raoult 2004; Schardl et al. 2004). Co-dispersal may be particularly important as symbiotic and antagonistic effects can be severely aggravated by co-infections with multiple micro-organisms (Morris et al. 2007). However, the abiotic and biotic factors that facilitate airborne dispersal have yet to be disentangled.

Here we applied high-throughput sequencing using a combination of multiple DNA metabarcoding markers and characterized the aerobiome targeting prokaryotes (16S rRNA gene: bacteria and archaea) and eukaryotes (18S rRNA gene to target fungi, protists and animals; rbcL cpDNA gene to target plants). We sampled airborne particles up to 10 mm in size continuously for three consecutive weeks in autumn 2015, using two automated samplers on top of a 12 m high building located amidst a mixed urban and agricultural landscape in the center of the Netherlands. Combining the sequencing data for all three markers with meteorological measurements via indicator analyses and network approaches, we aimed to gain further insights in dispersal mechanisms that govern the airborne microbiome composition.

2. Methods

2.1. Sampling

In October 2015, air sampling was conducted on the roof top (~12 m above ground level) of a building at the campus of Wageningen University & Research, Wageningen, the Netherlands (51°59′ N 5°40′ E). Within a circle of 1 km distance from the building, the direct surroundings include both urban area to the south (Wageningen campus and town) and east (Bennekom) and agricultural area to the north and west (mostly arable crop production, some cattle). A large area of deciduous forest is present further eastwards (>1 km).

Two sampling devices were placed at the western and eastern outer edge of the rooftop respectively, separated by a distance of 50 m. We used a Hirst Volumetric Spore Sampler (BS02225; Burkard Scientific Ltd.), which is collecting airborne particles by sucking in air through a 2 × 14 mm opening at a rate of 10 L / min, and trapping the particles onto a 2 cm wide adhesive tape. This tape was prepared within a flow cabinet, by applying a petroleum jelly liquefied with xylene as a thin layer on the standard transparent plastic tape provided with the spore sampler by the supplier (Lacey and West 2007). The tape was then mounted on a drum that is rotating at a fixed speed and makes one full round in 7 days. As a result, the mounted piece of tape represents a continuous air sample covering a period of 7 days. Both samplers were started at the same day, and tapes were replaced after 7 and 14 days, resulting in simultaneous measurements from two locations for a period covering 21 days. Each tape was cut into seven equal pieces, each representing a 1-day sampling period. Every piece was then divided into two equal parts; one half was mounted on a slide for inspection under a microscope for visual inspection under a binocular, while the other half was put into a sterile Eppendorf tube and stored at −4 °C until molecular analysis.

Preparation of the tapes was performed inside a laminar flow cabinet to avoid contamination as far as possible. To check for any remaining contamination, an extra piece of adhesive tape was prepared as a negative control, following the same procedure as above but being cut and stored directly after preparation without being placed in the sampler. As with the other tapes, the control tape was cut into seven equal pieces, included as separate control samples in all further analyses.

2.2. Molecular analysis

DNA extraction of all 49 tape samples was performed using a PowerPlant® DNA Isolation Pro Kit (Mobio Laboratories Inc.), using a modified protocol. Samples were inserted into the Eppendorf tube in a spiraling manner, so that the tape was wound along the tube wall with the adhesive size facing inwards. We then added 600 μL of B1 buffer solution + Nonidet P40 substitute (0.2 M sodium phosphate, pH 8.0, with 1 μL/ml Nonidet P40 and 3 μL RNase). Samples were homogenized by adding Ballotini Zirkonia/Silica beads (1.0 mm diameter; 0.6 g per tube) and shaking the tube on a Ribolyzer at 6000 rpm for two periods of 40 s with 2 min cooling on ice in between. From that point onwards, DNA extraction was then continued by following step 6–18 of the standard protocol as supplied by the kit's manufacturer.

Per DNA extract, PCR amplification was performed using three different primer sets that together cover the entire range of prokaryotic and eukaryotic diversity: 1) primers 341F and 805R (Herlemann et al. 2011) targeting the V3 + V4 subregions of the 16S rRNA gene in order to cover the bacterial community, primers Euk_1391f and EukB targeting the V4 subregion of the 18S rRNA gene in order to cover the eukaryotic community (Amaral-Zettler et al. 2009), and primers 1F (Fay et al. 1997) and 432R (Little and Barrington 2003) targeting the chloroplast rbcL gene in order to zoom in on the plant community at a finer taxonomic resolution. While we acknowledge the possibility of adding additional markers to increase the resolution also for other groups (e.g. ITS to increase species-level resolution for fungi), and we encourage their use in future studies to test hypotheses for specific groups of organisms, we did not consider this to be essential for our current explorative study. We chose to increase the resolution for plants as we observed small plant fragments by visual examination of the tapes and were interested in exploring their potential role as dispersal vectors for microbes.

All primers were flanked by the Nextera indexing adapters (Illumina, San Diego, CA, USA). PCR conditions were identical for all three markers. Reactions were performed in a final volume of 25 ul, consisting of 4 μL of undiluted DNA extract, 0.6 U Platinum Taq (Invitrogen), 1x

PCR Buffer, 2.5 mM MgCl2, 5% Trehalose, 200 ng/μl BSA, 50 μM dNTP and 250 nM of each NXT primer. Cycling conditions consisted of 94 °C for 2 min, followed by 5 cycles of 94 °C for 40 s, 55 °C for 40 s and 72 °C for 60 s, followed by 30 cycles of 94 °C for 40 s, 61 °C for 40 s and 72 °C for 60 s, and a final step of 72 °C for 5 min.

All 147 amplicons (49 samples × 3 markers) were prepared for paired-end sequencing on an Illumina MiSeq platform. PCR, index sequences and sequencing adapters were added to the first PCR products using a Nextera XT index kit (Illumina, San Diego, CA, USA). Dual index tags were added to each sample to build a library for each sample, resulting in 147 libraries with unique pairs of dual indexes. Dual indexes were added by PCR in a 28 μL system with 12 μL AccuPrime SuperMix II (Invitrogen, Eugene, OR, USA), 2 μL index primer P5, 2 μL index primer P7, 5 μL PCR product from first PCR and 7 μL PCR-grade H2O (Sigma-Aldrich). The amplification was performed under the following thermal cycler condition: 98 °C for 1 min, followed by 13 cycles of denaturation at 98 °C for 10 s, annealing at 55 °C for 20 s, and extension at 68 °C for 40 s. Cycling was completed at 68 °C for 5 min. The PCR products were purified using HighPrep™ PCR (Magbio Genomics Inc., Gaithersburg, Maryland, US) beads. Libraries were equimolarly pooled following quantification using a Qubit®2.0 fluorometer (Thermo Fisher Scientific, Waltham, USA) and sequenced at Aarhus University on an Illumina Miseq platform using the 250 bp PE MiSeq version 2 reagent kit (Illumina, San Diego, CA, USA). To avoid template termination of some clusters, the number of cycles were set to only 2 × 230.

2.3. Bioinformatics

We used the USEARCH/UPARSE Illumina paired-end pipeline (Edgar 2013). First sequences were trimmed, merged and quality filtered (maxEE 0.5). Subsequently the sequence data was dereplicated, sorted by abundance, chimeras removed, clustered (OTU radius 3%) into OTUs and singletons removed. Finally merged raw reads were mapped to a de novo OTU database at 97% similarity. OTUs for the 18S rRNA gene and rbcL gene datasets were pre-assigned to a rough taxonomic classification taking the top hit of a Basic Local Alignment Search Tool (BLAST; algorithm v 2.2.23; (Altschul et al. 1990)) search against the NCBI database using an e-value cut-off of 1e_5, an identity cut-off of 96% and a coverage cut-off of 90% of the query sequence covered in the alignments. All assigned eukaryotic 18S rRNA gene sequences were then manually verified by BLASTn searches against the NCBI GenBank nt database for correct taxonomic assignments. For that, the best 50 hits were analysed and OTUs conservatively classified if resulting hits showed consistent taxonomic patterns. OTUs for the 16S rRNA gene dataset were assigned using an automated procedure as described (Ramirez et al. 2019).

2.4. Statistical analysis

To avoid false detections due to PCR or sequencing errors, we deleted all OTUs that occupied a read count less than 0.001% of the total read count across all samples. In both the 16S and 18S rRNA gene datasets significant numbers of OTUs were observed for the control samples (fragments of a tape that was prepared in the lab but not placed in the sampler). To rigorously avoid any false positives produced by potential contamination in the lab, we used a conservative approach. For each OTU, we calculated an average relative abundance among the control samples, as well as among the 'true' samples (those actually collected from the air samplers), and excluded all OTUs for which the average abundance in the controls was more than half of the average abundance in the true samples. All analyses were based on relative abundances per OTU per sample to correct for variation in sequencing depth among samples.

2.4.1. Relations with weather conditions

Daily records of meteorological conditions were available for the entire sampling period from a weather station located a few kilometres west of the sampling locations (https://www.wur.nl/en/show/Weather-Station-De-Veenkampen.htm). Data for the following variables were obtained and used in the analyses: temperature (°C; daily average, minimum and maximum), wind speed (m/s; daily average and maximum), air humidity (% moisture content; daily average), air pressure (kPa; daily average), solar radiation (W/m^2; daily sum), rainfall (mm; daily sum), wind direction (4 categories of 90 compass degrees, 0 is due north) and sunshine duration (minutes; daily sum).

To test how the weather variables influenced aerobiome composition we conducted Multiple-Response Permutation Procedures (PERMANOVA) for each marker. All weather variables were included as predictors and their significance was assessed by testing their marginal contribution to the model (i.e. SS type 3 in ANOVA terminology). Permutations were restricted within sampling station and sampling dates were treated as a series within the sampling station using the 'how()' function from the permute v. 0.9–0 package (Simpson, 2016). For each analysis we ran a full enumeration of all possible permutations (here 440 permutations).

Sampling days were clustered into three clusters using k-means clustering with 10 random starts using the 'kmeans()' function. Subsequently we conducted indicator analysis on the identified taxa (16S rRNA gene order, 18S rRNA gene order, RbcL gene family levels respectively), using 'multipatt()' function in the indicspecies v. 1.7.6. package (De Cáceres and Legendre 2009), following (De Cáceres et al. 2010). We used the group weighted point biserial correlation coefficient on log-transformed relative abundances as our test statistic and tested for significant taxon to cluster associations using permutation tests. These permutations were constrained as above.

For each taxon that was identified as an indicator for differences in weather conditions, we modelled its responses to the weather variables. For each taxon we fitted using maximum-likelihood a linear mixed model to the log-transformed relative abundances with sampling station as a random effect and included weather variables as fixed factors. We only included those weather variables that had an absolute correlation smaller than $r = 0.8$ with the other weather variables (Additional file A). We fitted models, with the same random effects structure, with all possible combinations of the included fixed effects (all subsets regression), including an intercept-only model (the null-model), and ranked them based on AIC_c (Burnham and Anderson 2004; Grueber et al. 2011). We selected all models that were within two AIC_c units from the optimal model (the model with the lowest AIC_c score). We performed model averaging (Lukacs et al. 2009) on the selected models and evaluated the importance of each weather variable based on the standardized regression coefficients (using partial standard deviation (Bring 1994), which accounts for collinearity among predictors). Linear mixed models were fitted using nlme v3.1–131 (Pinheiro et al. 2013) and model selection and model averaging was done in package MuMIn v. 1.15.6 (Barton and Barton 2015). The statistical analyses were conducted in R v. 3.3.3 (R Core Team, 2019).

2.4.2. Co-abundance network analysis

To make the co-response network visualization, the relative abundance tables of the three datasets, 16S, rRNA, 18S rRNA and rbcL gene were combined. OTUs present in < 10 samples were discarded. Next, using R (R Core Team, 2019) the Spearman correlations were calculated for each pair of OTUs and between each OTU and the measured weather variables. For visualisation using Cytoscape (Shannon et al. 2003) pairs with a correlation ≥ 0.8 were exported, together with the relative abundance of the nodes and the correlations of each OTU with the weather measurements.

3. Results

3.1. Communities were taxonomically and functionally diverse, including many pathogens and parasites

Bacterial reads dominated the aerobiome's prokaryotic 16S rRNA gene reads (mean 95.71%), but also plant-derived chloroplast (mean 3.18%) and archaea (mean 0.06%) sequences were found. The dominant phylum was Firmicutes, representing 62.9% of all 16S rRNA gene reads. Among the most abundant orders were Clostridiales, Bacillales, Lactobacillales and Erysipelotrichales (Firmicutes) and Sphingomonadales (Proteobacteria) (Fig. 1A). The vast majority of eukaryotic 18S rRNA gene reads (Fig. 1B) were assigned as fungi (98.91%), largely composed of Basidiomycota (80.50%) and Ascomycota (18.18%), and a small fraction of insect-pathogenic Entomophthoromycota (0.21%). The remaining reads were highly diverse, being placed into 21 classes, including arthropods (0.39%; mainly a drain fly (*Clogmia* sp.), plants (0.02%) and a taxonomically and functionally diverse collection of protists (68 OTUs belonging to 19 classes): The protist diversity included potential plant pathogens, such as Oomycota (24 OTUs; mostly *Phytophthora*), and invertebrate (endo)parasites belonging to the Apicomplexa (10 OTUs) and Opalinata (1 OTU). Apart from these potential endoparasites (together representing on average 0.05% of the reads), all identified fungal and protist taxa were identified as spore-producing. The dataset for the rbcL cpDNA gene revealed 29 plant families. All were phanerogams (seed producing plants; dominant taxa shown in Fig. 1C) and the majority of reads (68.5%) were assigned to taxa that were likely releasing pollen during our sampling period: e.g. members of the Urticaceae (28.3%), Pinaceae (12.5%) and Euphorbiaceae (10.9%) (Additional file B). Nevertheless, part of the reads (27.6%) originated from taxa that cannot have been flowering during the sampling period. These were either tree species (22.3%; e.g. *Quercus robur*, *Alnus* sp.) that occurred nearby and likely released small fragments (e.g. leaf hairs) or crop species harvested from nearby fields during the time of sampling (4.3%, *Glycine* max, *Solanum tuberosum*).

3.2. Weather conditions drive presence of dominant prokaryotes, while affecting only some eukaryotes

Multivariate permutation tests (PERMANOVA) showed significant relations between the overall community composition per marker dataset and multiple meteorological variables (Table 1). Air pressure was consistently related to community composition in all three markers. This was also the case for temperature, although prokaryotic (16S) and plant (rbcL) communities mainly related to daily minimum temperature, while the eukaryotic community (18S) as a whole related to daily maximum temperature, along with levels of radiation and rainfall (Table 1). Wind direction was important for both the prokaryotic and eukaryotic communities (as determined based on 16S and 18S respectively, Table 1).

Using k-means clustering, we grouped the 21 sampling days into three main weather types (Additional file C): dry sunny days with little wind (cluster A), wet cloudy days with low temperatures and relatively little wind (B) and wet cloudy days with higher temperatures and relatively strong wind (C). We then identified indicator taxa for these clusters and tested for significant relations between the relative abundances of these taxa and specific meteorological parameters (Additional file D). Of the dominant prokaryotes (i.e. top 10 orders with highest average relative abundances depicted in Fig. 1A), Clostridiales and Erysipelotrichales were particularly abundant in cluster B. Both groups were significantly enriched on days with a high air pressure, northeasterly winds and infrequent periods of sunshine (Fig. 2). Furthermore, Erysipelotrichales were more abundant on days with high humidity. In contrast, Cytophagales were most abundant in cluster C and significantly enriched on days with low air pressure and high radiation.

The dominant eukaryotes were not associated to particular weather cluster(s), with a clear exception for the Oomycetes (cluster A and C), which were positively related to temperature and low humidity (Fig. 2).

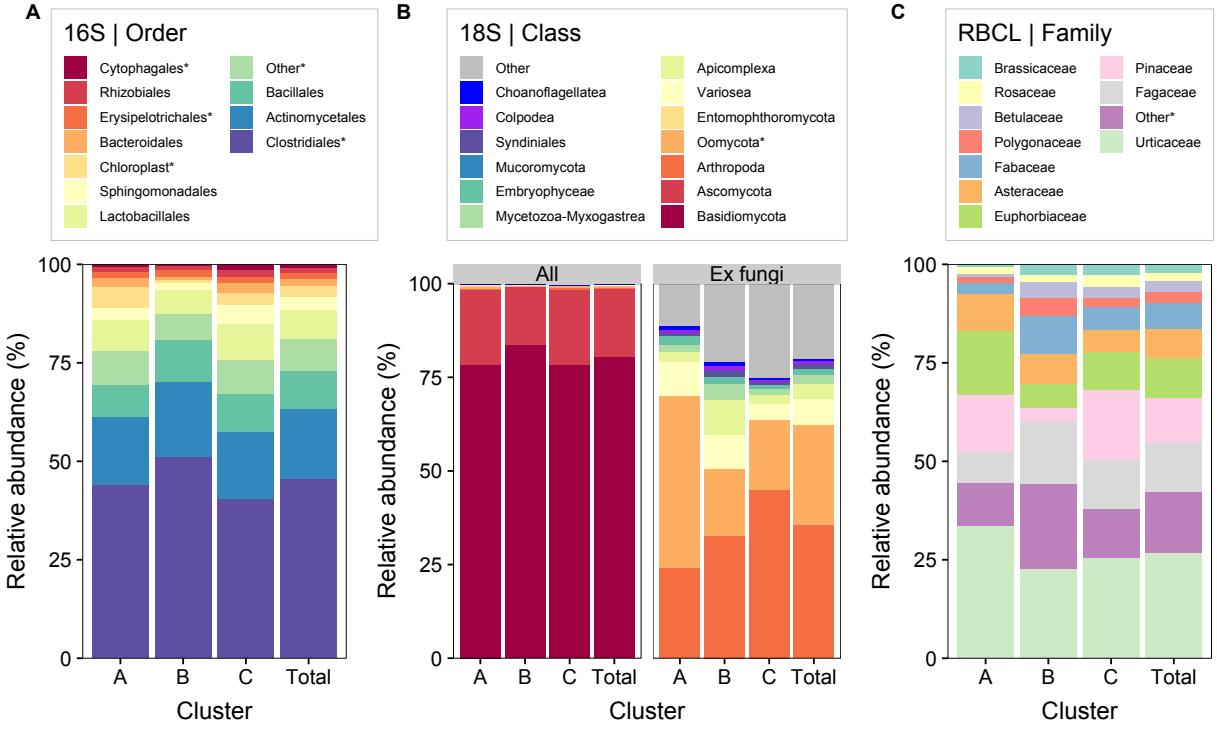

Fig. 1. Relative abundance of the ten most dominant taxonomic groups per marker averaged per weather cluster. Taxa of which the name is followed by an asterisk (*) were also identified as an indicator taxon for one or more weather clusters (Additional file D). A) Prokaryotic orders, based on 16S; B) eukaryotes classes, based on 18S, including and excluding fungi (Ex fungi); C) plant families, based on rbcL. Weather clusters are clusters of sampling days with broad similarity in weather conditions (see main text and Additional file C).

Table 1
Results of multivariate permutation tests (PERMANOVA) testing for effects of meteorological parameters on aerobiome composition.

		16S (Prokaryotes)		18S (Eukaryotes)		rbcL (Plants)	
Parameter	d.f.	Pseudo-F	P	Pseudo-F	P	Pseudo-F	P
Temperature (mean)	1	**1.73**	**0.045**	1.57	0.154	2.02	0.072
Temperature (minimum)	1	**4.12**	**0.002**	0.59	0.438	**3.15**	**0.002**
Temperature (maximum)	1	1.03	0.206	**2.52**	**0.043**	1.90	0.086
Radiation	1	0.97	0.510	**4.12**	**0.007**	1.21	0.311
Sunshine	1	**1.70**	**0.045**	0.50	0.583	0.54	0.819
Air humidity	1	1.58	0.156	1.82	0.141	0.92	0.517
Rainfall	1	1.43	0.138	**2.75**	**0.045**	1.46	0.222
Air pressure	1	**2.44**	**0.039**	**4.87**	**0.014**	**4.13**	**0.005**
Wind mean	1	1.51	0.056	0.14	0.941	1.17	0.218
Wind max	1	1.47	0.088	0.89	0.317	0.97	0.390
Wind direction (NE, SE, SW, NW)	3	**2.15**	**0.014**	**2.09**	**0.048**	1.38	0.093
Residual	28						
R^2_{adj}		38.0%		30.7%		21.0%	

The effects of meteorological predictors on the taxonomic composition of the 16S rRNA gene, 18S rRNA gene and rbcL gene datasets were analyzed separately. In order to isolate unique effects of predictors F-tests were constructed using the marginal sums of squares. See methods for further details. Bold face is used to highlight significant predictors ($p < 0.05$).

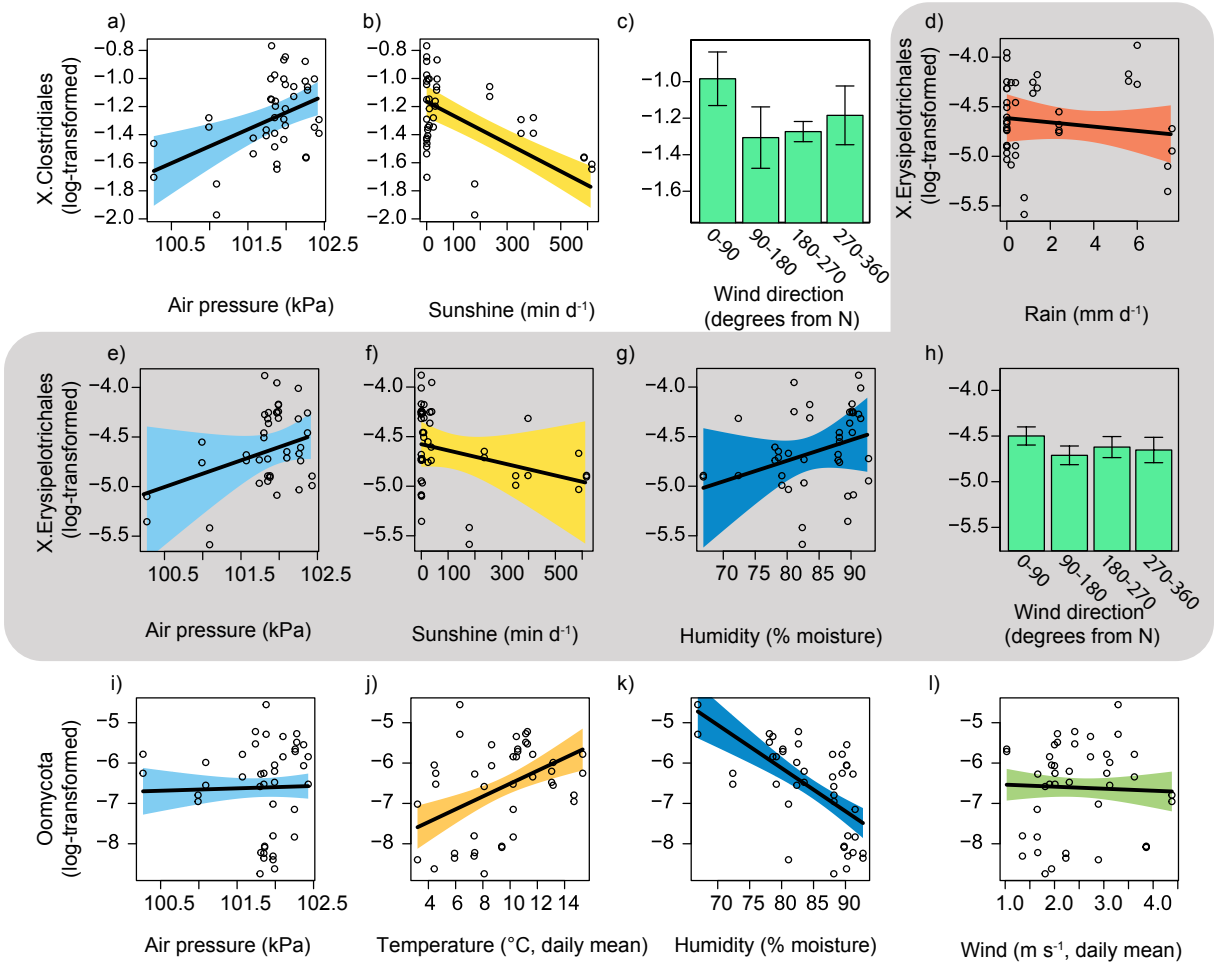

Fig. 2. Estimated relationships of three key aerobiome taxa (see main text) to selected meteorological predictors. Log-transformed relative abundances are shown for Clostridiales (white section on top row), Erysipelotrichales (grey section; middle row and top right) and Oomycota (bottom row) in relation to the most important predictors based on model selection. To account for model uncertainty model averaging was used and the mean and 95% confidence intervals of the model averaged coefficients are shown. Panels with the same predictors share the same color for confidence intervals.

Two subdominant groups, Ichthyosporea (cluster A) and Acanthamoebidae (associated with clusters A and C) correlated with daily sunshine duration and less strongly with air pressure. None of the major plant families were affected by weather conditions, while the less abundant Boraginaceae (cluster B), Cupressaceae and Sapindaceae (clusters B and C) were indicative of relatively low temperatures and high humidity.

3.3. Clusters of co-occurring taxa reveal dispersal of pathogens and parasites with their most likely hosts

Network analysis, combining all OTUs from the three marker datasets, identified eight 'co-response clusters' of OTUs with tightly correlated abundance patterns along the 21 sampling days (Fig. 3A; Additional File E). Clusters 2, 3 and 6 are dominated by bacteria, with especially cluster 3 and 6 containing a variety of obligate gut-related taxa (Fig. 3C; dominant OTUs belonging to Clostridiales, Bacillales, Lactobacillales and Erysipelotrichales). Among these gut bacteria, a few highly abundant OTUs identified as *Clostridia* sp. grouped particularly closely together (upper right corner in Fig. 3C). The majority of ascomycete and basidiomycete fungi (Fig. 3B) grouped together in cluster 4, 5, 7 and to some extent 8 (Fig. 3A), with cluster 7 also containing a suite of bacteria belonging to the Actinomycetales. The drain fly OTU clustered in a subsection of group 8 that stands apart from the fungal clusters in Fig. 3A, and that also contains the majority of (endo)parasites (Fig. 3D), such as known insect-related parasitic protists (Apicomplexa) and pathogenic fungi (Entomophthoromycota), as well as nine of the less abundant Oomycete OTUs. Five dominant Oomycete OTUs (representing 47% of the Oomycete sequences) grouped together without connection to the main network as did most of the plant OTUs (group 1 in Fig. 3B). Plotting correlations between OTUs and average daily temperature (Fig. 3E), highlights the existence of co-occurrence groups with similar responses to weather conditions, while confirming the differences in response between e.g. *Clostridia* sp. and other (gut-related) bacteria.

4. Discussion

Here, we provide the first full inventory of the aerobiome, and show that next to bacteria and fungi, a taxonomically and functionally diverse community of protists is dispersed by means of air currents. Furthermore, we provide new insights in how weather variables, as well as co-dispersal via insects, pollen and plant fragments may affect the potential of microbes to be air-dispersed.

Our results of the community composition of bacteria (Bowers et al. 2013; Bowers et al. 2011a) and fungi (Bowers et al. 2013; Fröhlich-Nowoisky et al. 2009; Yamamoto et al. 2012) are in line with previous studies confirming that airborne microbial communities do not entirely resemble those in soils, aquatic and other studied environments, but rather are composed of taxa from a variety of source environments with composition depending on local land use (Bowers et al. 2011a). The bacterial groups in our samples, were dominated by the phylum Firmicutes, composed of members of the two dominant orders Clostridiales and Erysipelotrichales, as well as the abundant Bacteriodales, which all frequently occur in mammalian gut and feces (Bowers et al. 2011b). This dominance of fecal-associated organisms is indicative of organic manure application and a general enrichment of these taxa early in fertilizer application periods (Bowers et al. 2013), but also with a relatively high abundance of these groups in suburban areas (Bowers et al. 2011a), as present here. Likewise, the observed list of plant taxa included both

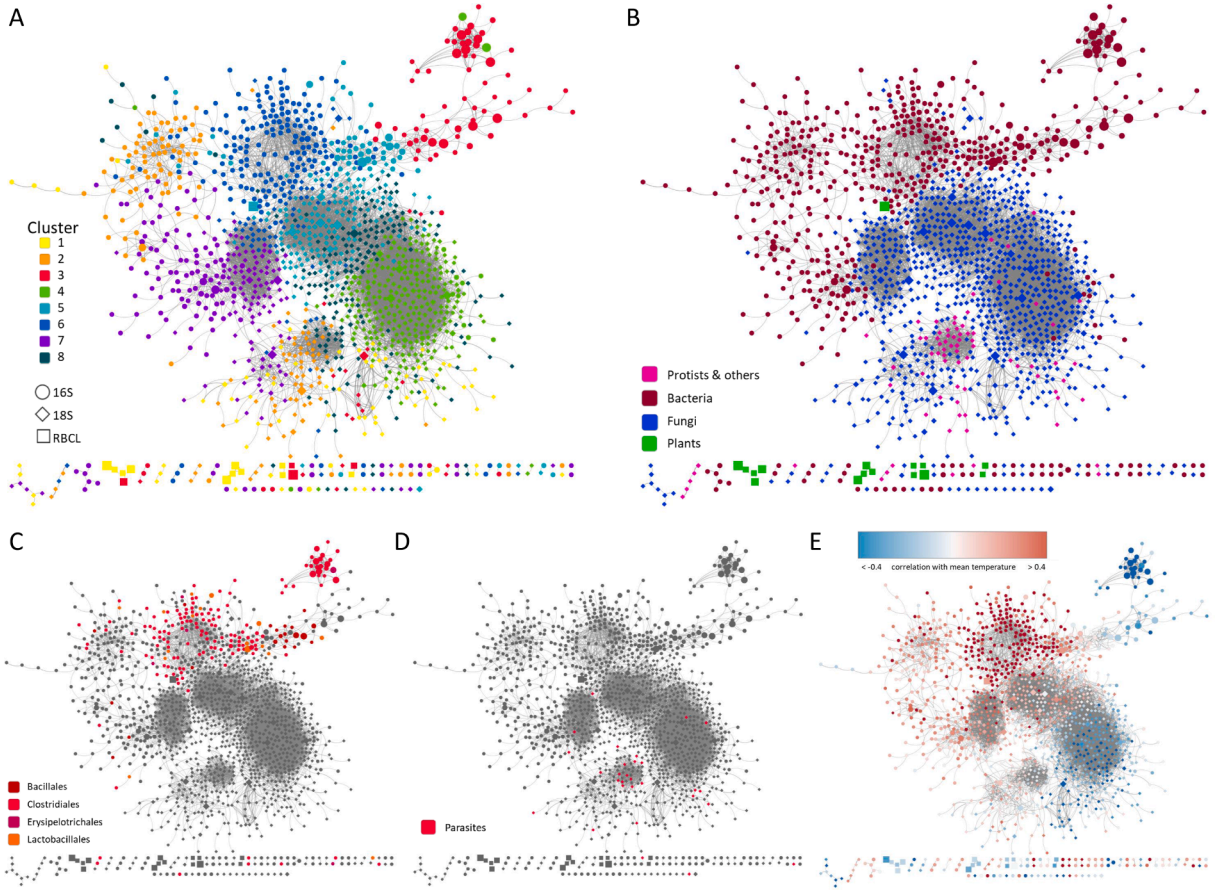

Fig. 3. Co-response network visualizing groups of OTUs with correlated abundance patterns. Correlations were based on the total set of OTUs obtained from three combined marker datasets, indicated with different symbols (see panel A for symbol legend; symbol size scales with increasing average relative abundance in the dataset). Lines represent positive correlations (Spearman $r > 0.8$) between pairs of OTUs; with shorter lines corresponding to higher correlations. OTUs without any such correlation are not shown. Panel A highlights eight main co-response groups based on k-means clustering (see methods); Panel B highlights the main taxonomic kingdoms; Panel C highlights OTUs of gut-related bacterial orders; Panel D highlights OTUs of taxa identified as (endo)parasites; Panel E shows correlations of all OTUs with mean temperatures per sampling day.

locally grown crop species (e.g. *Solanum tuberosum*) and species from local gardens (e.g. *Cedrus atlantica*) and nearby green roofs (*Sedum album*).

The dominance of the gram-positive Firmicutes among bacteria is attributed to their thick cell wall making them tolerant to adverse conditions such as high UV radiation levels and low moisture conditions. As expected, we found that particularly spore forming organisms are dominating the aerobiome, which is illustrated by the dominance of Clostridiales and Bacillales bacteria, basidiomycete and ascomycete fungi and spore-forming protists. Particularly for plant pathogenic oomycetes there could be positive selection for airborne dispersal as this will facilitate them to reach new hosts. These fungi-like protists are major pathogens of agricultural crops responsible for profound yield-losses, such as in potato (e.g. *Phytophthora*) (Brown and Hovmøller 2002; Judelson and Blanco 2005).

Our results suggest that weather parameters are important determinant of the aerobiome community composition. While given the limited sample size of our current study observed relations should be interpreted with caution, we did identify some putative drivers that deserve further exploration. For instance, the relative abundance of oomycetes in air increased with temperature and decreased with humidity and wind speed, suggesting that the relatively large spores of oomycetes, in comparison to fungal spores, are more dependent on suitable weather conditions to spread. Interestingly, only microbial taxa appeared to be directly affected by weather conditions, indicating that the metazoan and plant taxa recorded are airborne independent of weather conditions and that only few groups determine the overall differences in composition between distinct weather conditions.

We uncovered several unexpected members of the aerobiome, including animal parasites (Apicomplexans, Entomophthoromycota and Opalinata). The strong link between these microbial clades with animals provides additional support that these taxa, that are often found in environmental sequencing efforts in both aquatic systems (de Vargas et al. 2015) and soils (Geisen et al. 2015; Mahé et al. 2017), actually parasitize animals. Our correlation networks support the link between parasites and a host, suggesting that many correlations in HTS-based analyses actually depict species interactions. This potential makes approaches such as ours particularly useful to decipher the entire aerobiome, as it allows determining co-dispersal of species.

Differences in airborne bacterial and fungal communities are known to be present at different heights (Robinson et al. 2020) and between geographical regions (Barberán et al. 2015; Flies et al. 2020), and likely such differences are also present for other groups of eukaryotes. It must therefore be noted that, given the fact that our data were obtained at a local scale, within a period of 21 days and at a single height, our findings are not easily transferable to other aerobiomes. We see our study as a proof of principle for integrative microbiome inventories showing weather-dependent daily variations in aerobiome composition. Clearly, additional studies of this kind, ideally based on larger sample sizes, will be needed to determine how aerobiome communities respond to weather conditions across time and space and eventually derive more general conclusions on dominant factors shaping their composition.

5. Conclusions

Using a three-marker high-throughput sequencing approach, our study uncovered a high diversity of temporally fluctuating, co-occurring microorganisms that collectively form the aerobiome. Our data highlight that spore-formation serves as a successful dispersal strategy also in little-studied groups such as protists, as all non-parasitic protists in our dataset were spore-forming. Furthermore, network analyses identified tight associations among aerobiome members, including host-symbiont interactions. Finally, the automated sampling protocol combined with metabarcoding might be transformed to validate and improve epidemiological models, and to establish early warning systems in biomonitoring approaches when focusing on plant pathogens, but also animal and human pathogens.

CRediT authorship contribution statement

G. Arjen de Groot: Conceptualization, Investigation, Methodology, Writing - original draft. **Stefan Geisen:** Conceptualization, Investigation, Methodology, Writing - original draft. **E.R. Jasper Wubs:** Conceptualization, Formal analysis, Writing - original draft. **Liz Meulenbroek:** Investigation, Methodology, Writing - review & editing. **Ivo Laros:** Writing - review & editing. **L. Basten Snoek:** Formal analysis, Writing - review & editing. **Dennis R. Lammertsma:** Investigation, Methodology, Writing - review & editing. **Lars H. Hansen:** Writing - review & editing. **Pieter A. Slim:** Conceptualization, Funding acquisition, Writing - review & editing.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Acknowledgements

We thank Dr. Loek Kuiters for valuable comments on previous versions of this manuscript.

This research was part of the KennisBasis IV strategic research program 'Sustainable Spatial Development of Ecosystems, Landscapes, Seas and Regions' (KB-24-002-005), funded by the Dutch Ministry of Economic Affairs and carried out by Wageningen University & Research. SG was supported by an NWO-VENI grant from the Netherlands Organization for Scientific Research (016.Veni.181.078). ERJW was funded by an NWO-RUBICON grant (Netherlands Organization for Scientific Research, contract no. 019.181EN.01) and by ERC-ADV 323020 SPECIALS to W.H. van der Putten. The funding bodies had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2021.106551.

References

Adams, R.I., Miletto, M., Taylor, J.W., Bruns, T.D., 2013. Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. ISME J. 7, 1262.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., Huse, S.M., 2009. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PLoS ONE 4, e6372.

Astorga, B., Lorenzo-Morales, J., Martín-Navarro, C.M., Alarcón, V., Moreno, J., González, A.C., Navarrete, E., Piñero, J.E., Valladares, B., 2011. *Acanthamoeba* belonging to T3, T4, and T11: genotypes isolated from air-conditioning units in Santiago, Chile. J. Eukaryotic Microbiol. 58, 542–544.

Baas-Becking, L.G.M., 1934. Geobiologie, of inleiding tot de milieukunde ed˙eds. Van Stockum.

Bar-On, Y.M., Phillips, R., Milo, R., 2018. The biomass distribution on Earth. Proc. Natl. Acad. Sci. 115, 6506–6511.

Barberán, A., Ladau, J., Leff, J.W., Pollard, K.S., Menninger, H.L., Dunn, R.R., Fierer, N., 2015. Continental-scale distributions of dust-associated bacteria and fungi. Proc. Natl. Acad. Sci. 112, 5756.

Barton, K., Barton, M.K., 2015. Package 'MuMIn'. R package version 1156.

Bass, D., Stentiford, G.D., Wang, H.-C., Koskella, B., Tyler, C.R., 2019. The pathobiome in animal and plant diseases. Trends Ecol. Evol. 34, 996–1008.

Bowers, R.M., Clements, N., Emerson, J.B., Wiedinmyer, C., Hannigan, M.P., Fierer, N., 2013. Seasonal Variability in Bacterial and Fungal Diversity of the Near-Surface Atmosphere. Environ. Sci. Technol. 47, 12097–12106.

Bowers, R.M., McLetchie, S., Knight, R., Fierer, N., 2011a. Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. ISME J. 5, 601–612.

Bowers, R.M., Sullivan, A.P., Costello, E.K., Collett, J.L., Knight, R., Fierer, N., 2011b. Sources of Bacteria in Outdoor Air across Cities in the Midwestern United States. Appl. Environ. Microbiol. 77, 6350–6356.

Bring, J., 1994. How to standardize regression coefficients. Am. Statist. 48, 209–213.
Brock, D.A., Douglas, T.E., Queller, D.C., Strassmann, J.E., 2011. Primitive agriculture in a social amoeba. Nature 469, 393.
Brown, J.K.M., Hovmøller, M.S., 2002. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. Science 297, 537–541.
Burnham, K.P., Anderson, D.R., 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociol. Methods Res. 33, 261–304.
Clobert, J., Danchin, E., Dhondt, A., Nichols, J., 2001. Dispersal. Oxford University Press.
De Cáceres, M., Legendre, P., 2009. Associations between species and groups of sites: indices and statistical inference. Ecology 90, 3566–3574.
De Cáceres, M., Legendre, P., Moretti, M., 2010. Improving indicator species analysis by combining groups of sites. Oikos 119, 1674–1684.
de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O., Lima-Mendez, G., Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Coordinators, T.O., Acinas, S.G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M.E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P., Karsenti, E., 2015. Eukaryotic plankton diversity in the sunlit ocean. Science 348.
Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods 10, 996–998.
Fay, M.F., Swensen, S.M., Chase, M.W., 1997. Taxonomic affinities of *Medusagyne oppositifolia* (Medusagynaceae). Kew Bull. 52, 111–120.
Fisher, M.C., Henk, D.A., Briggs, C.J., Brownstein, J.S., Madoff, L.C., McCraw, S.L., Gurr, S.J., 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature 484, 186–194.
Flies, E.J., Clarke, L.J., Brook, B.W., Jones, P., 2020. Urbanisation reduces the abundance and diversity of airborne microbes - but what does that mean for our health? A systematic review. Sci. Total Environ. 738, 140337.
Foissner, W., 2006. Biogeography and dispersal of micro-organisms: a review emphasizing protists. Acta Protozool. 45, 111–136.
Foissner, W., 2008. Protist diversity and distribution: some basic considerations. Biodivers. Conserv. 17, 235–242.
Frankel, R., Galun, E., 2012. Pollination mechanisms, reproduction and plant breeding ed^eds. Springer Science & Business Media.
Fröhlich-Nowoisky, J., Pickersgill, D.A., Després, V.R., Pöschl, U., 2009. High diversity of fungi in air particulate matter. Proc. Natl. Acad. Sci. 106, 12814–12819.
Geisen, S., Tveit, A.T., Clark, I.M., Richter, A., Svenning, M.M., Bonkowski, M., Urich, T., 2015. Metatranscriptomic census of active protists in soils. ISME J. 9, 2178–2190.
Genitsaris, S., Kormas, K.A., Christaki, U., Monchy, S., Moustaka-Gouni, M., 2014. Molecular diversity reveals previously undetected air-dispersed protist colonists in a Mediterranean area. Sci. Total Environ. 478, 70–79.
Greub, G., Raoult, D., 2004. Microorganisms resistant to free-living amoebae. Clin. Microbiol. Rev. 17, 413–433.
Grueber, C.E., Nakagawa, S., Laws, R.J., Jamieson, I.G., 2011. Multimodel inference in ecology and evolution: challenges and solutions. J. Evol. Biol. 24, 699–711.
Hawksworth, D.L., Lucking, R., 2017. Fungal diversity revisited: 2.2 to 3.8 million species. Microbiol. Spectrum 5.
Hellweger, F.L., van Sebille, E., Fredrick, N.D., 2014. Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. Science 345, 1346.
Herlemann, D.P., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., Andersson, A.F., 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J. 5, 1571–1579.
Judelson, H.S., Blanco, F.A., 2005. The spores of *Phytophthora*: weapons of the plant destroyer. Nat. Rev. Microbiol. 3, 47.
Lacey, M.E., West, J.S., 2007. The air spora: a manual for catching and identifying airborne biological particles ed^eds. Springer Science & Business Media.
Little, D.P., Barrington, D.S., 2003. Major evolutionary events in the origin and diversification of the fern genus *Polystichum* (Dryopteridaceae). Am. J. Bot. 90, 508–514.
Lukacs, P.M., Burnham, K.P., Anderson, D.R., 2009. Model selection bias and Freedman's paradox. Ann. Inst. Stat. Math. 62, 117.
Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., Singer, D., Mayor, J., Bunge, J., Sernaker, S., Siemensmeyer, T., Trautmann, I., Romac, S., Berney, C., Kozlov, A., Mitchell, E.A.D., Seppey, C.V.W., Egge, E., Lentendu, G., Wirth, R., Trueba, G., Dunthorn, M., 2017. Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. Nat. Ecol. Evol. 1, 0091.
Möhler, O., DeMott, P., Vali, G., Levin, Z., 2007. Microbiology and atmospheric processes: the role of biological particles in cloud physics. Biogeosciences 4, 1059–1071.
Morris, W.F., Hufbauer, R.A., Agrawal, A.A., Bever, J.D., Borowicz, V.A., Gilbert, G.S., Maron, J.L., Mitchell, C.E., Parker, I.M., Power, A.G., Torchin, M.E., Vázquez, D.P., 2007. Direct and interactive effects of enemies and mutualists on plant performance: a meta-analysis. Ecology 88, 1021–1029.
Muñoz, J., Felicísimo, Á.M., Cabezas, F., Burgaz, A.R., Martínez, I., 2004. Wind as a long-distance dispersal vehicle in the southern hemisphere. Science 304, 1144–1147.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Team, R.C., 2013. NLME: Linear and nonlinear mixed effects models. R package version 31-128.
R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. Vienna, Austria: R Foundation for Statistical Computing.
Ramirez, K.S., Snoek, L.B., Koorem, K., Geisen, S., Bloem, L.J., ten Hooven, F., Kostenko, O., Krigas, N., Manrubia, M., Caković, D., van Raaij, D., Tsiafouli, M.A., Vreš, B., Čelik, T., Weser, C., Wilschut, R.A., van der Putten, W.H., 2019. Range-expansion effects on the belowground plant microbiome. Nat. Ecol. Evol. 3, 604–611.
Rivera, F., Lugo, A., Ramirez, E., Bonilla, P., Calderon, A., Rodriguez, S., Ortiz, R., Gallegos, E., Labastida, A., Chavez, M.P., 1992. Seasonal distribution of air-borne protozoa in Mexico City and its suburbs. Water Air Soil Pollut. 61, 17–36.
Robinson, J.M., Cando-Dumancela, C., Liddicoat, C., Weinstein, P., Cameron, R., Breed, M.F., 2020. Vertical stratification in urban green space aerobiomes. Environ. Health Perspect. 128, 117008.
Rout, M.E., Callaway, R.M., 2012. Interactions between exotic invasive plants and soil microbes in the rhizosphere suggest that 'everything is not everywhere'. Ann. Bot. 110, 213–222.
Schardl, C.L., Leuchtmann, A., Spiering, M.J., 2004. Symbioses of grasses with seedborne fungal endophytes. Annu. Rev. Plant Biol. 55, 315–340.
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504.
Simpson, G.L., 2016. Permutate: functions for generating restricted permutations of data. R package version 0.9-0.
Xiong, W., Song, Y., Yang, K., Gu, Y., Wei, Z., Kowalchuk, G.A., Xu, Y., Jousset, A., Shen, Q., Geisen, S., 2020. Rhizosphere protists are key determinants of plant health. Microbiome 8, 27.
Yamamoto, N., Bibby, K., Qian, J., Hospodsky, D., Rismani-Yazdi, H., Nazaroff, W.W., Peccia, J., 2012. Particle-size distributions and seasonal diversity of allergenic and pathogenic fungi in outdoor air. ISME J. 6, 1801–1811.