PNAS
www.pnas.org

Supplementary Information for

A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations

Joan Cáliz[$] (orcid.org/0000-0001-5714-7431), Xavier Triadó-Margarit[$] (orcid.org/0000-0001-5339-3221), Lluís Camarero (orcid.org/0000-0003-4271-8988) and Emilio O. Casamayor (orcid.org/0000-0001-7074-3318)
[$]equal contribution

Emilio O. Casamayor
Email: casamayor@ceab.csic.es

This PDF file includes:

Supplementary text
Figs. S1 to S8
Table S1
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1812826115

Supplementary text

Bioaerosol collection

Atmospheric wet precipitation (both rain and snow) was collected with an automatic wet and dry passive sampler MTX ARS 1010 (MTX, Bologna, Italy) equipped with two 667 cm^2 area polyethylene collection buckets and a hygroscopic sensor cell (1). The wet collector remained covered preventing atmospheric inputs until the hygroscopic sensor was activated. Particles deposited in the wet container (i.e. those washed from the atmosphere) were retained onto precombusted (450 °C, 4 hours) Whatman GF/F filters and then dried in a laboratory oven for 4 hours and kept in a dark and dry place (1).
For the molecular analysis, the variable V4 and V5 regions of the 16S rRNA gene (~250 nt) were amplified with the primers F515 (5'–GTGCCAGCMGCCGCGGTAA–3') and R806 (5'– GGACTACHVGGGTWTCTAAT–3') (2). For the eukaryotic V9 region of the 18S rRNA gene (~150 nt) we used primers 1391F (5'-GTACACACCGCCCGTC-3') and EukBr (5'-TGATCCTTCTGCAGGTTCACCTAC-3') (3), following the Earth Microbiome Project (EMP) protocols (www.earthmicrobiome.org).
For determination of the chemical composition of the water samples (rain/snow), major cations (Na^+, K^+, Ca^{2+}, Mg^{2+} and NH_4^+) and anions (Cl^-, NO_3^- and SO_4^{2-}) were analyzed by capillary electrophoresis (Quanta 4000, Waters). Water pH was determined using a low ionic strength electrode (Crison). Acid neutralizing capacity (ANC) was measured by potentiometric Gran titration. Dissolved inorganic (DIC) and organic (DOC) carbon were determined by conversion to CO_2 by acidification and catalytic combustion (DOC) and IR spectrometry to measure the CO_2 produced, with a Shimadzu TOC-5000 analyzer.

Sequence processing
Raw rRNA genes sequences were processed using the UPARSE pipeline (4). After merging of read pairs, filtering by read length and with an expected error of 0.25, ~42% of the original reads were retained (i.e., 5,622,407 reads for 16S rRNA gene and 6,417,502 for eukaryotes). The reads were dereplicated and clustered into operational taxonomic units (OTUs) at cut-off 0.03% identity after chimera removal (UCHIME) and excluding the singletons. More than 92% of the globally trimmed, quality filtered sequence pool was mapped back into OTUs. A total of 4,164 prokaryote and 8,248 eukaryote OTUs were obtained and taxonomically assigned with SILVA_119 (5). Chloroplast, mitochondria, Metazoa, Embryophyta (mostly pollen), and unclassified reads were excluded for further analyses. In order to minimize biased effects for differences in sampling effort, the original OTU table was average rarefied (100 random subsamplings, (6)) and set to a depth of 12,500 prokaryotes and 10,000 eukaryotes sequences per sample.
Environmental descriptive terms were extracted from closest matches (99% identity) using the SEQenv pipeline (7) for the most abundant OTUs using ENVO terms (8). Sixty-five and fifty-five per cent of the average abundance of prokaryotes and eukaryotes, respectively, were annotated to ENVO terms. It is important here to remark that predicted sources have to be interpreted with caution as annotations rely on homology searches, as well as available source entries, against sequences of from public databases.

Statistical Analyses

Factor Analysis (FA) is a dimension reduction technique that substantially reduces the number of original variables, and only a few factors are extracted from the data set. The factors are linear combinations of the original variables, and are computed in order to both be uncorrelated among each other, and be strongly correlated to different groups of input variables. Factors can therefore be interpreted as underlying processes driving the original variables distribution.

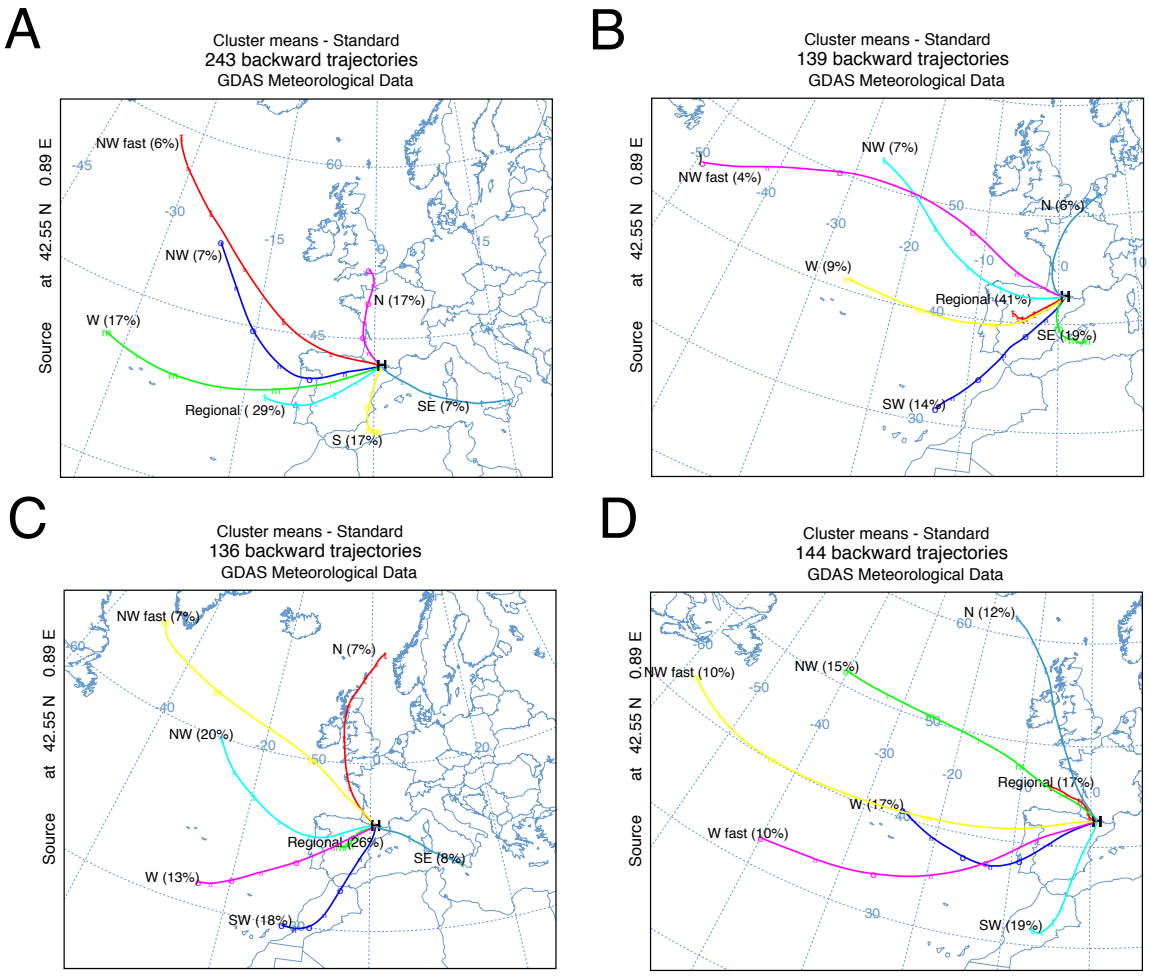

Fig. S1. Cluster centroids and relative frequency of back-trajectories associated to each cluster during period from 2007 to 2013, analyzed by season: spring (A), summer (B), autumn (C) and winter (D).

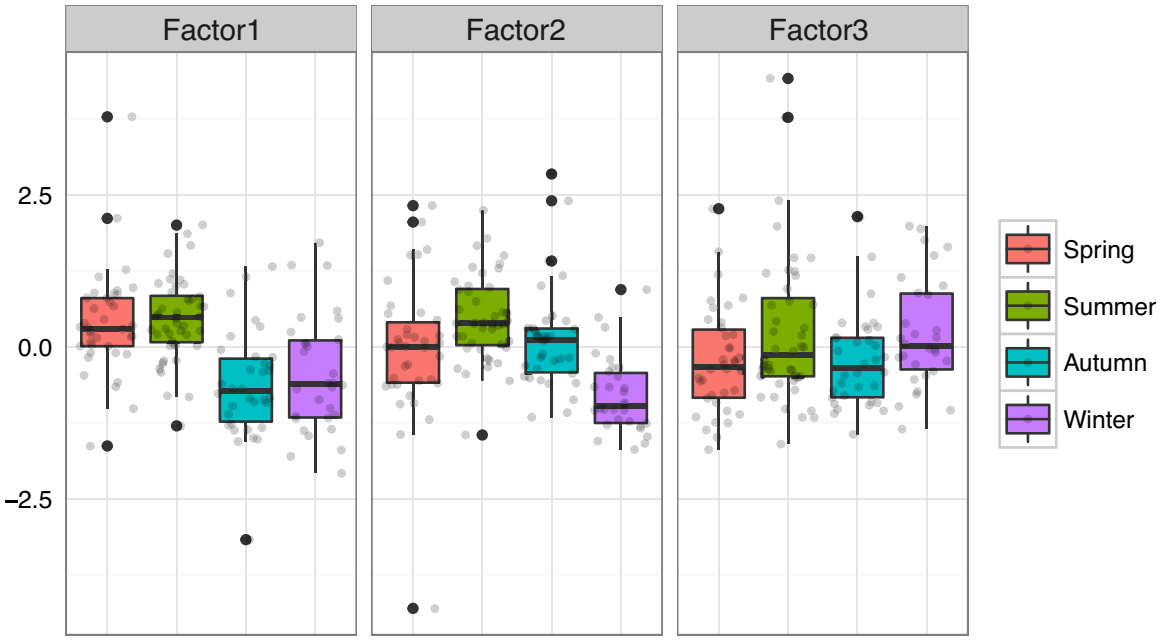

Fig. S2. Seasonal variation of factors described by the chemical composition of wet depositions: factor one (NO_3^-, SO_4^{2-}, NH_4^+, DOC and Mg^{2+}), factor two (pH, ANC, DIC, Ca^{2+} and Mg^{2+}) and factor three (Cl^- and Na^+).

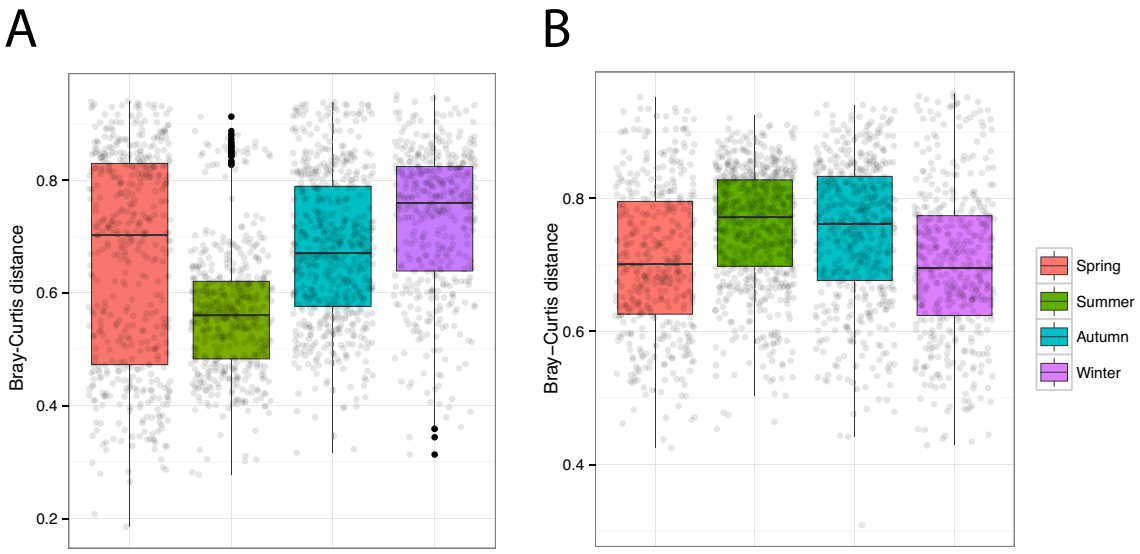

Fig. S3. Intra-seasonal community dissimilarity (Bray Curtis index) for bacteria (A) and eukaryotes (B).

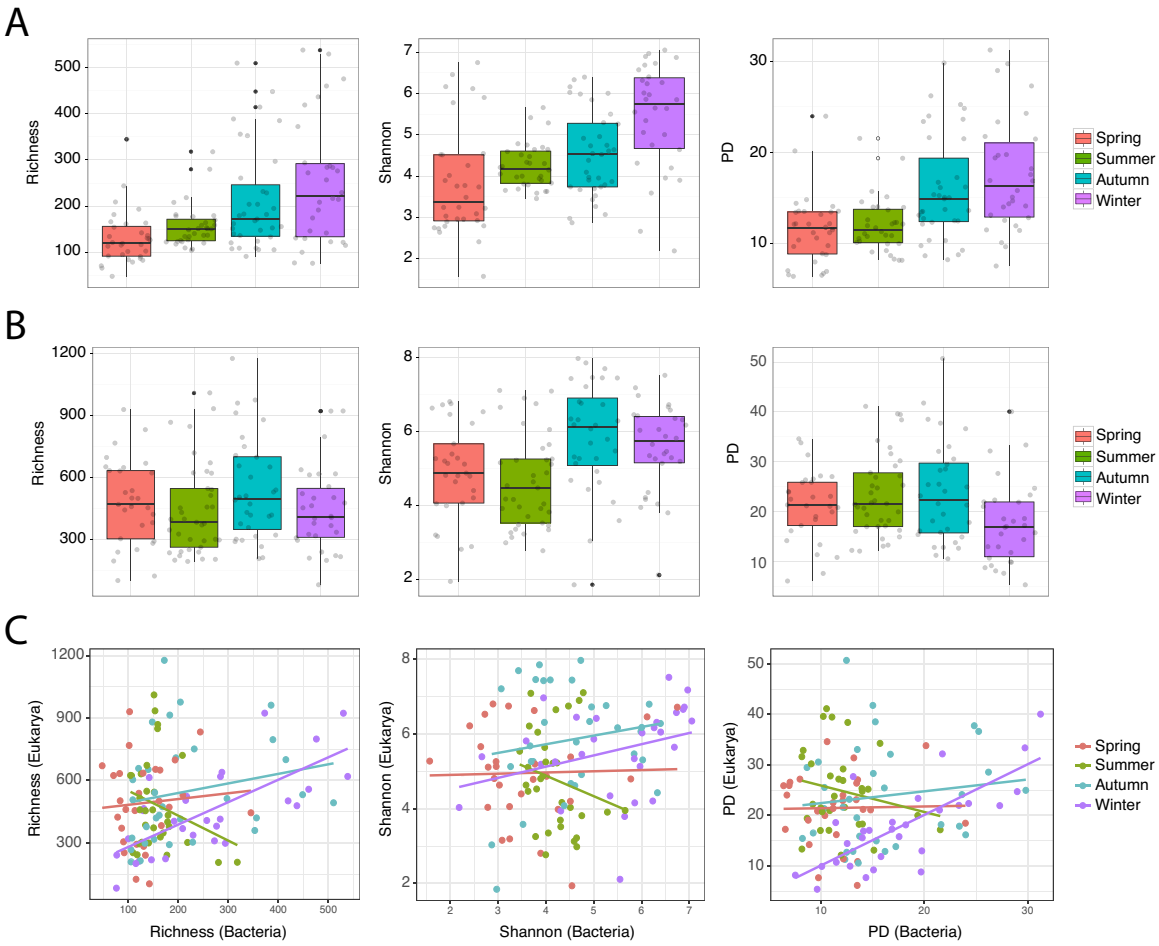

Fig. S4. Seasonal comparison of alpha diversity indices (richness, Shannon and Faith's phylogenetic diversity – PD) for airborne bacteria (A) and eukaryotes (B). Seasonal pairwise comparisons of alpha diversity indices between bacteria and eukaryotes (fitted linear model regression in color lines) (C).

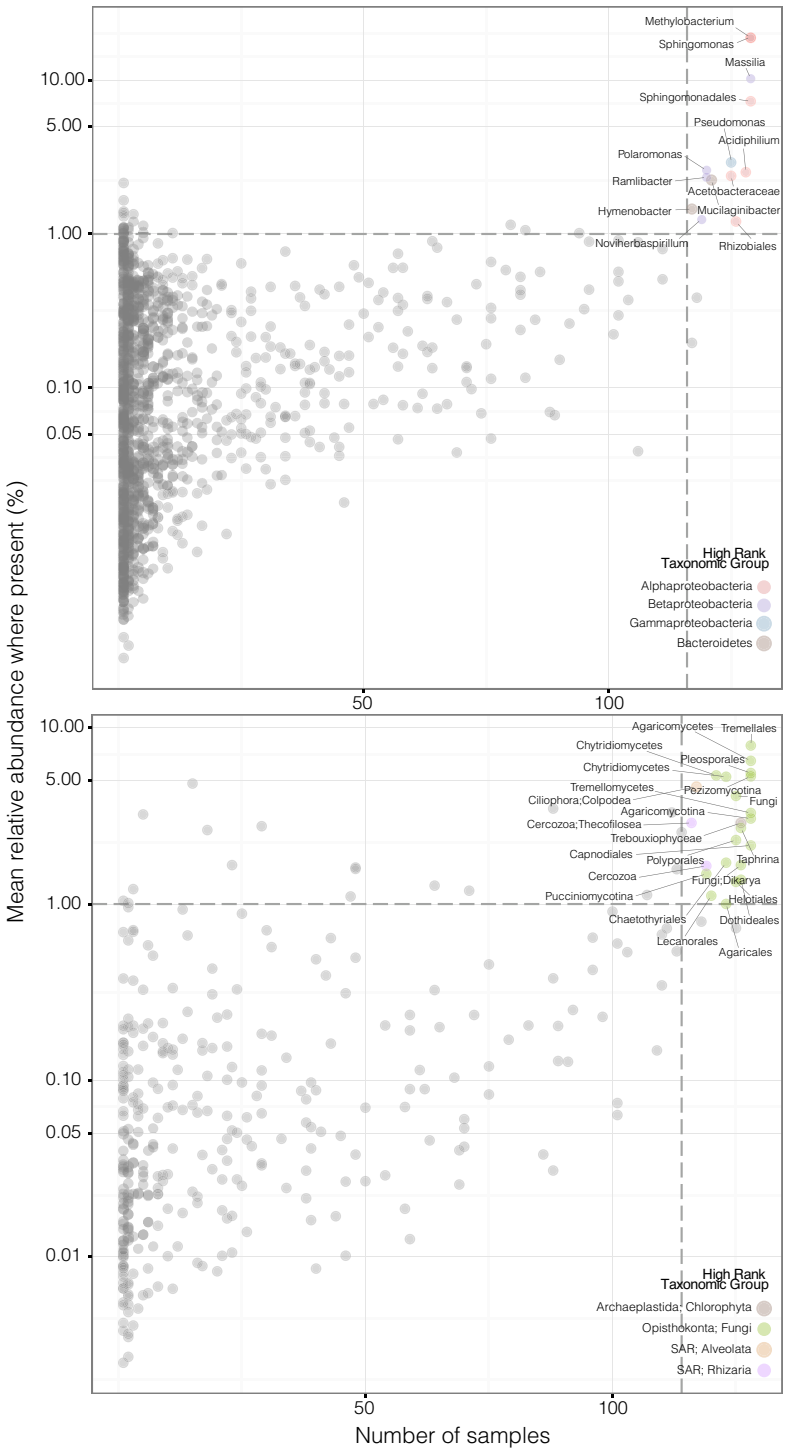

Fig. S5. Mean relative abundances (where present) and occupancy relationships for bacterial and eukaryal taxa (upper and lower panel, respectively).

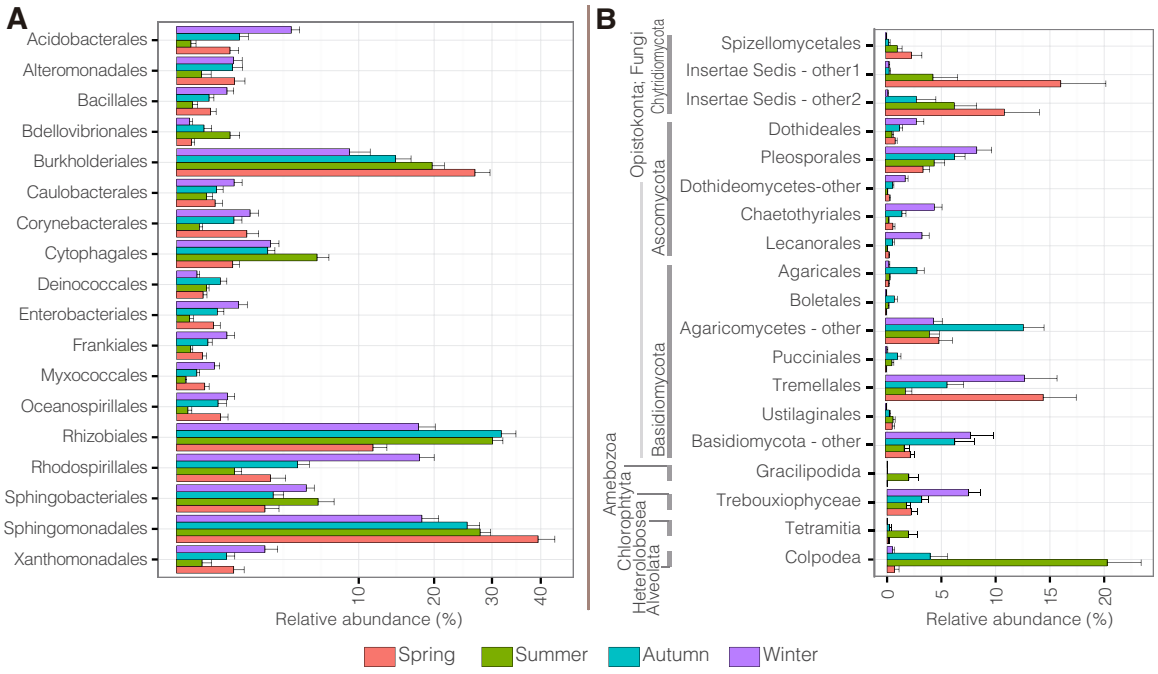

Fig. S6. (A) Bacterial orders and (B) Eukaryal taxa (when possible, at the order level) with significantly different relative abundances among seasons. Minor groups, i.e., relative abundance < 1%, are not shown.

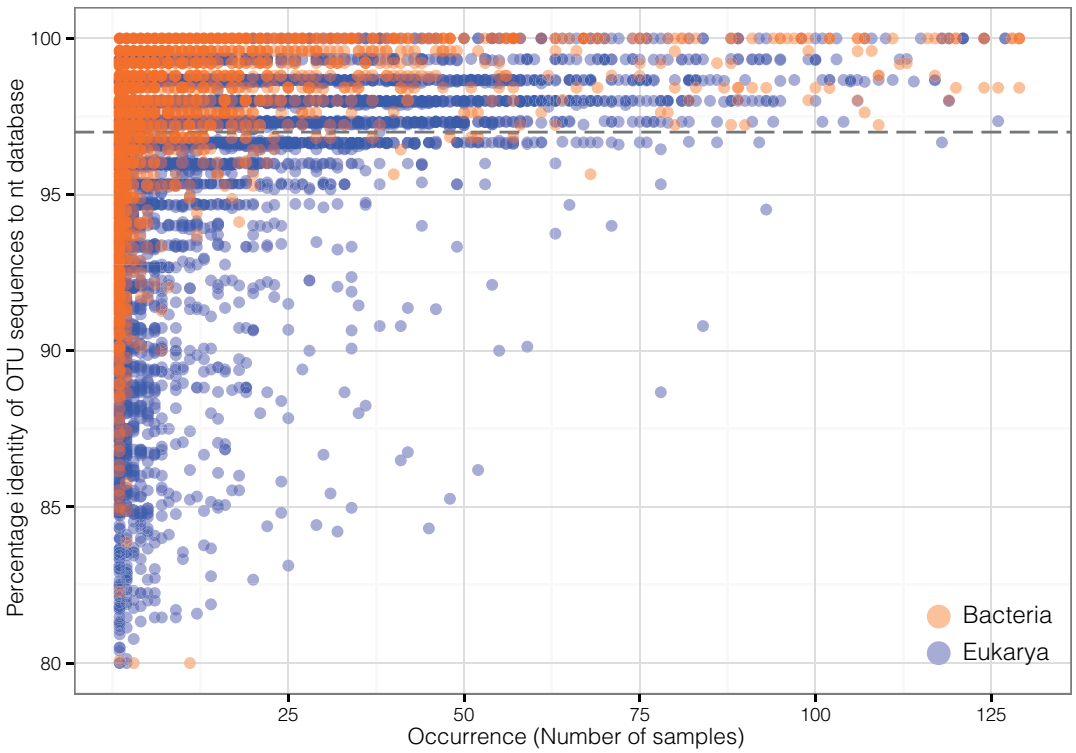

Fig. S7. OTU sequence resemblance to databases, based on its percent identity values, and occupancy relationships. Data from both bacterial and eukaryal OTUs are shown.

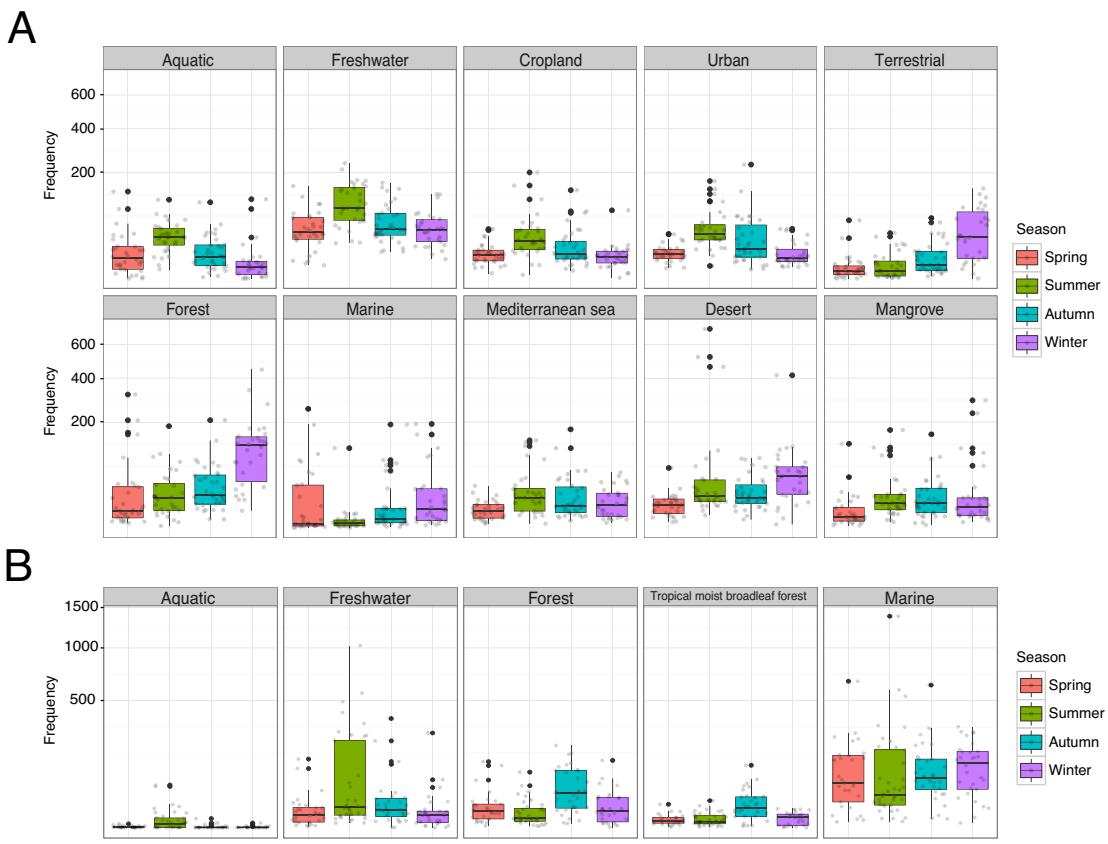

Fig. S8. Seasonal distribution of depositions according to the predicted source of biomes related to airborne bacteria (A) and eukaryotes (B). Biomes are ordered from left to right according to seasonal prevalence. Note that because the skewed distribution, the y-axis is squared.

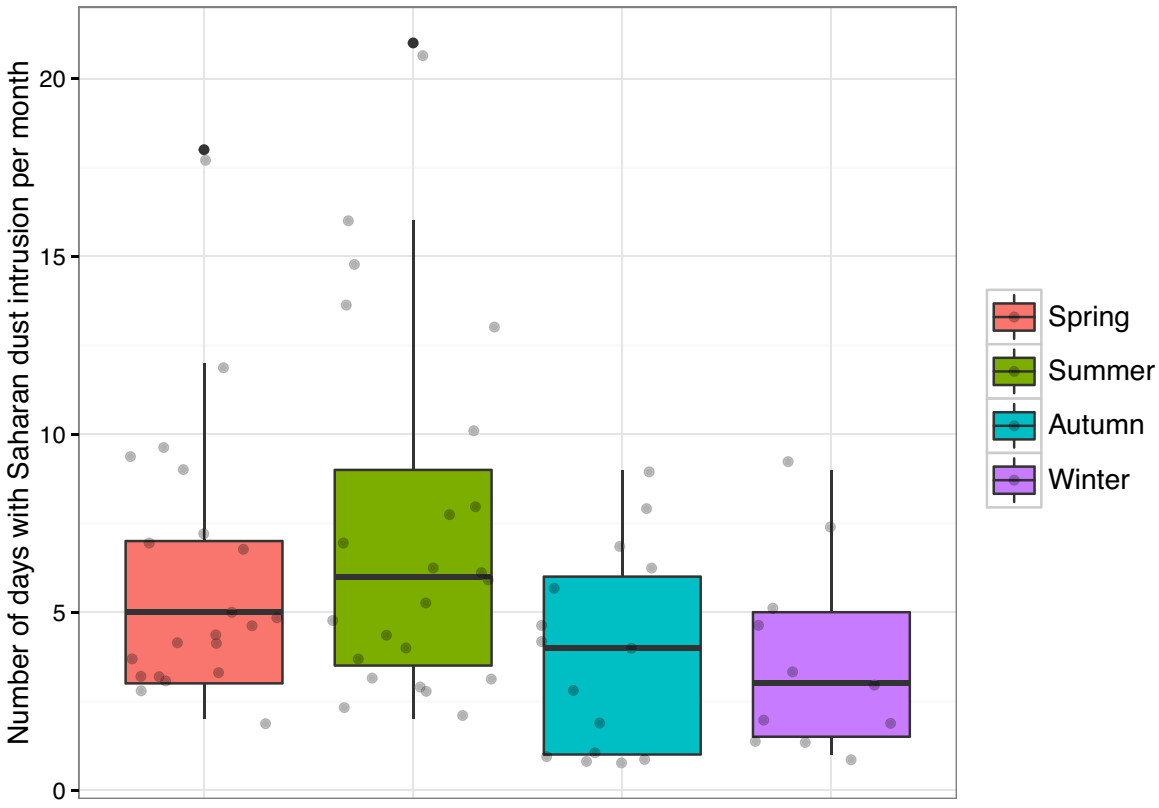

Fig. S9. Comparison of the occurrence of Saharan dust intrusions per month among seasons. Data retrieved from www.calima.ws

Table S1. Correlation coefficients of the three factors extracted by Factor Analysis

	Factor1	Factor2	Factor3
NO_3^-	0.924		
SO_4^{2-}	0.772		
NH_4^+	0.687		
DOC	0.606		
Mg^{2+}	0.600	0.604	
pH		0.831	
ANC		0.802	
DIC		0.770	
Ca^{2+}		0.733	
Cl^-			0.925
Na^+			0.851

References for SI reference citations

1. Hervas A, Camarero L, Reche I, & Casamayor EO (2009) Viability and potential for immigration of airborne bacteria from Africa that reach high mountain lakes in Europe. *Environ Microbiol* 11(6):1612-1623.
2. Caporaso JG*, et al.* (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* 108 Suppl 1:4516-4522.
3. Stoeck T*, et al.* (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* 19(s1):21-31.
4. Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 10(10):996-998.
5. Quast C*, et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41(Database issue):D590-596.
6. Caliz J*, et al.* (2015) Influence of edaphic, climatic, and agronomic factors on the composition and abundance of nitrifying microorganisms in the rhizosphere of commercial olive crops. *PLoS One* 10(5):e0125787.
7. Sinclair L*, et al.* (2016) Seqenv: linking sequences to environments through text mining. *PeerJ* 4:e2690.
8. Buttigieg PL, Morrison N, Smith B, Mungall CJ, & Lewis SE (2013) The environment ontology: contextualising biological and biomedical entities. *J Biomed Semant* 4(1):1-9.