Microbial communities in the tropical air ecosystem follow a precise diel cycle

Elena S. Gusareva[a,1], Enzo Acerbi[a,1], Kenny J. X. Lau[a], Irvan Luhung[a,1], Balakrishnan N. V. Premkrishnan[a], Sandra Kolundžija[a], Rikky W. Purbojati[a], Anthony Wong[a], James N. I. Houghton[a], Dana Miller[a], Nicolas E. Gaultier[a], Cassie E. Heinle[a], Megan E. Clare[a], Vineeth Kodengil Vettath[a], Carmon Kee[a], Serene B. Y. Lim[a], Caroline Chénard[a], Wen Jia Phung[a], Kavita K. Kushwaha[a], Ang Poh Nee[a], Alexander Putra[a], Deepa Panicker[a], Koh Yanqing[a], Yap Zhei Hwee[a], Sachin R. Lohar[a], Mikinori Kuwata[b], Hie Lim Kim[a,b], Liang Yang[a], Akira Uchida[a], Daniela I. Drautz-Moses[a], Ana Carolina M. Junqueira[c], and Stephan C. Schuster[a,2]

[a]Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, 637551 Singapore; [b]Asian School of the Environment, Nanyang Technological University, 637459 Singapore; and [c]Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 21941-590 Brazil

Edited by Edward F. DeLong, University of Hawaii at Manoa, Honolulu, HI, and approved October 3, 2019 (received for review June 14, 2019)

The atmosphere is vastly underexplored as a habitable ecosystem for microbial organisms. In this study, we investigated 795 time-resolved metagenomes from tropical air, generating 2.27 terabases of data. Despite only 9 to 17% of the generated sequence data currently being assignable to taxa, the air harbored a microbial diversity that rivals the complexity of other planetary ecosystems. The airborne microbial organisms followed a clear diel cycle, possibly driven by environmental factors. Interday taxonomic diversity exceeded day-to-day and month-to-month variation. Environmental time series revealed the existence of a large core of microbial taxa that remained invariable over 13 mo, thereby underlining the long-term robustness of the airborne community structure. Unlike terrestrial or aquatic environments, where prokaryotes are prevalent, the tropical airborne biomass was dominated by DNA from eukaryotic phyla. Specific fungal and bacterial species were strongly correlated with temperature, humidity, and CO_2 concentration, making them suitable biomarkers for studying the bioaerosol dynamics of the atmosphere.

microbial ecology | bioaerosols | air microbiome | temperature | tropics

Of the 3 planetary ecosystems, the lithosphere, hydrosphere, and atmosphere, only soil and aquatic environments have been extensively investigated for their microbial content (1). In contrast, the study of the atmosphere as a microorganism-harboring ecosystem has been technologically challenging, rendering it an understudied biosphere (2, 3). The atmosphere is assumed to contain microorganisms from local and remote sources, being dispersed for as long as they remain suspended in air (4–6). Previously reported numbers of cells found in a defined volume of air are based either on cultivation (colony-forming units; CFUs) or are estimated by amplified ribosomal molecular markers (2, 7). Furthermore, research into airborne microbial organisms is predominantly carried out in countries located within temperate climate zones or as global surveys with varying geographic and climatic parameters (8–10).

Here, we investigate the microbial composition of the outdoor air microbiome at a single site in the tropics (Singapore, N1.334731; E103.680996, sea-surface level) through a metagenomics time course study of unprecedented temporal and taxonomic resolution. Singapore is located in the intertropical convergence zone and experiences a biannual reversal of wind directions, which defines the 2 monsoon seasons (11, 12). The biannually alternating wind directions of the study site could result in vastly disparate sources of transported microorganisms. Hindcast estimations over 48 h using the HYSPLIT Trajectory Model package (13) indicate that during the northeast monsoon season, the prevailing airflow is transporting air masses from the general area of Taiwan and the Philippines. Wind trajectories during the southwest monsoon season identify the Timor and Arafura seas as possible sites of origin. Little variation in length of daylight across the seasons as well as absence of large day-to-day fluctuations in temperature and relative humidity are factors making the tropical region an ideal air microbiome research site. Using deep metagenomics sequencing and analysis, we describe the air microbiome variation within time intervals of hours, days, and up to months. The resulting dataset of 795 air samples provided a high level of reproducibility and enabled extensive statistical analysis. Models developed using Bayesian network analysis implicated specific environmental parameters as drivers for ecosystem diversity and composition. Environmental studies have reported a microbial cellular density of $\sim 10^{12}$ per L of soil and $\sim 10^{9}$ per L of water (14–17). In contrast, air, being a much less dense medium, is reported to harbor $\sim 10^{2}$ cells per L sampled, based on CFU counts and amplification of rRNA markers (5). The 7 to 10 log difference in cell density per volume represents a significant technological challenge for analyzing the biological content of air, as it mandates very large volumes to be collected over extended periods of time (2). We have devised an air-sampling approach that allows for 2-h or shorter intervals to be sampled, generating highly reproducible results for

Significance

This manuscript describes a precise diel cycle carried out by airborne microbiota in the tropics. 795 metagenomes from air samples taken from a single site show that fungi, bacteria, and plants all adhere to a specific timing for their presence in the near-surface atmosphere. The airborne community composition thereby shows an unexpected robustness, with the majority of the dynamics in taxa composition occurring within 24 h, but not across days, weeks, or months. Environmental parameters are the main drivers for the observed phenomenon, with temperature being the most important one.

Author contributions: S.C.S. designed research; K.J.X.L., I.L., S.K., A.W., J.N.I.H., D.M., N.E.G., C.E.H., M.E.C., V.K.V., C.K., S.B.Y.L., C.C., W.J.P., K.K.K., A.P.N., A.P., D.P., K.Y., Y.Z.H., S.R.L., M.K., H.L.K., L.Y., A.U., D.I.D.-M., A.C.M.J., and S.C.S. performed research; E.S.G., E.A., K.J.X.L., I.L., B.N.V.P., R.W.P., D.I.D.-M., A.C.M.J., and S.C.S. analyzed data; and E.S.G., E.A., I.L., D.I.D.-M., A.C.M.J., and S.C.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Data deposition: The raw metagenomic sequencing data reported in this paper have been deposited in the National Center for Biotechnology Information (NCBI) database (submission ID code SUB3630161; BioProject ID code PRJNA436039).

[1]E.S.G., E.A., and I.L. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: scschuster@ntu.edu.sg.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1908493116/-/DCSupplemental.

First published October 28, 2019.

downstream metagenomics analyses. In this regard, we can show that collected biomass contains sufficient quantity and complexity of DNA to enable ultradeep sequencing, generating more than 300 million reads from a single air sample. More importantly, the robustness and precision of our sampling technique, as illustrated by its replicability, allowed for the undertaking of experimental time-series surveys across days, weeks, and months (i.e., day variation experiment [DVE] 1 to 5). Air samples were collected 12 times per d, every 2 h, for 5 consecutive days, in 3-mo intervals (DVE2 to 5). This 13-mo schedule of sampling events covered Singapore's wet and dry seasons, including 2 monsoon seasons with wind directions from the northeast and southwest. Each time series (DVE2 to 5) resulted in 180 sequencing libraries. Thus, all 5 DVEs combined represent 795 air samples from 265 time points, which resulted in 9.11 billion reads of 250 bp, or 2.27 terabase pairs, the equivalent of ~760 human genomes (*SI Appendix*, Table S1).

Results

Taxonomic Diversity of a Single Tropical Air Sample. We first assessed the taxonomic diversity of a random air sample (DVE1011) in DVE1 relative to seawater, soil, and human gut microbiome samples from the same geographic region (Singapore, each normalized to 2 million reads per sample). The air sample contained the lowest percentage of identifiable microbial taxa (9 to 17%, night/day), while the human microbiome, being the most intensively studied (18), showed up to 64% identifiable taxa in our analysis at the superkingdom level (Fig. 1*A*) (19). The unidentifiable fraction of the air microbiome contained reads that were either not specific to a known taxon (35%, unassigned) or did not show any similarity to the nonredundant sequence database (48%, no hits). The "no hits" category likely consists of 1 of the following possibilities: 1) intergenic regions of eukaryotic organisms; 2) extracellular DNA that has been highly mutagenized by the exposure to harsh environmental conditions; or 3) as-yet undiscovered microorganisms that currently have no representation in sequence databases. Hence, compared with soil and seawater, air is currently the most underexplored ecosystem on a planetary scale.

Despite the low percentage of assigned reads, the air samples displayed species richness comparable to the other 3 ecosystems (Fig. 1*B*). Unlike these environments, which were dominated by bacterial phyla, the tropical air samples showed that a large proportion of the assigned reads belong to fungi (82%), while only 14.5% were assigned to bacteria, 2.6% to plant species, and 0.087% to archaea (Fig. 1*C*). As our sequencing data and metagenomics analysis allowed for species-level identification, we used known genome sizes of the top 40 fungal and bacterial genomes (*SI Appendix*, Fig. S1) to normalize the DNA read abundance, thereby approximating the number of bacterial and fungal cells (assuming a single chromosome per cell). The estimated proportions of bacterial and fungal cells were 62.1 and 37.9%, respectively (Fig. 1*D*). For nighttime samples, the

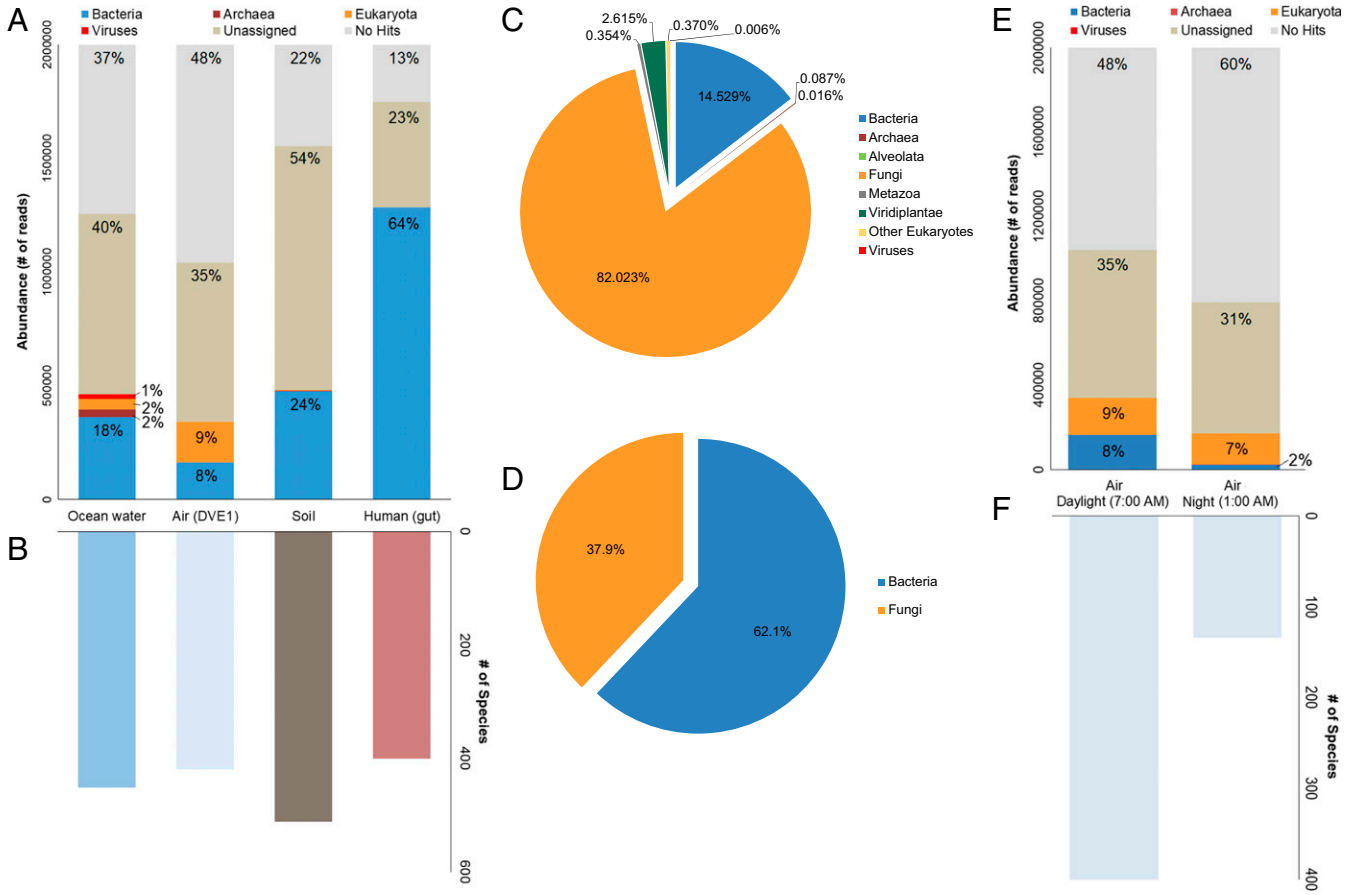


Fig. 1. Taxonomic structure of microbial communities. (*A*) Superkingdom-level taxonomical classification of relative abundances of microbes in samples representing the ecosystems ocean, air, soil, and human gut. (*B*) Number of species in samples representing the 4 corresponding ecosystems. (*C*) Relative abundances of the total of metagenomic reads from 5 DVEs, assigned to taxonomic groups. (*D*) Estimate of bacterial versus fungal cell ratio, based on read count normalized by genome size. (*E*) Relative abundances of microbes in daytime and nighttime air samples. These 2 samples were randomly selected from the complete DVE1 dataset. (*F*) Number of assigned species in the corresponding daylight and night air samples.

percentage of identifiable reads declined to only 9%, with most of the bacterial reads being absent (Fig. 1 *E* and *F*).

Environmental Time-Series Sampling. By comparing a 330-million-read, ultradeep air sample dataset with an independently sequenced 4-million-read dataset of the same 2-h sample (DVE1011), we determined that 2 million randomly subsampled reads were sufficient to identify and rank the 500 most abundant taxa from airborne communities (*SI Appendix*, Figs. S2–S5). This observation, together with the minimal biomass requirements, led to a sampling design for the environmental time series DVE2 to 5 with 2-h sampling intervals and 12 time points per d.

Statistical analysis was performed for DVE1 to 5 individually. Of the randomly selected 2 million reads of each sequenced air sample, on average 12.16 (DVE1, pilot), 10.30 (DVE2), 12.46 (DVE3), 11.66 (DVE4), and 8.25% (DVE5) were taxonomically identified, with the majority belonging to bacterial or fungal taxa (for details, refer to *SI Appendix*, Table S2). Our analysis of the metagenomics data allowed for taxonomic identification down to the species level (*SI Appendix*, Fig. S6, including blank samples in *SI Appendix*, Fig. S7). Of the microbial species identified in each time series (DVE1 to 5), 725 were detectable throughout the 13-mo sampling period and therefore are considered the core microbiome of the airborne community. In contrast, only between 74 and 206 species were unique to a specific time series (Fig. 2*A* and *SI Appendix*, Table S3).

The average richness of the tropical air microbial community (assessed by the number of identifiable species) throughout DVE2 to 5 was found to vary prominently over the course of a day (Fig. 2*B* and *SI Appendix*, Fig. S8; $R_{\mathrm{ANOSIM}} = 0.62$ to 0.81, $P_{\mathrm{ANOSIM}} < 0.001$; ANOSIM, analysis of similarities), while the day-to-day and month-to-month average variation displayed an unexpectedly small fluctuation (Fig. 2 *C* and *D* and *SI Appendix*, Fig. S8; $R_{\mathrm{ANOSIM}} = 0.01$ to 0.12, $P_{\mathrm{ANOSIM}} = 0.002$ to 0.28). The recurrent diel changes of the air microbial community were sufficiently reproducible within 24 h to allow for a single air sample to be traced back to the actual time of day it was collected. However, it is not possible to assign a sample to a specific month/season using our data as reference.

Despite achieving species-level taxonomic resolution, we summarized the temporal variability of airborne organisms in 7 taxonomic groups (Fig. 3*A*) in order to show the general applicability to each taxonomic group. Plant-associated reads were collapsed at the level of Viridiplantae, fungi at the 2 taxonomic levels of Ascomycota and Basidiomycota, and bacteria at the levels of Cyanobacteria, Firmicutes, Proteobacteria, and Actinobacteria. Within the allocated taxonomic groups, plant and bacterial phyla had similar diel dynamics to their relative abundances (Fig. 3*A*), for example, more frequent during daytime (midday), with only traces detected during nighttime. In contrast, Basidiomycota were the only taxonomic group whose abundance was high during the night and diminished during the daytime hours (Fig. 3*A*). The abundance of Ascomycota, on the other hand, predominantly increased during midday or in correlation with rain events. These patterns were observed in all 5 DVEs and were statistically significant (*SI Appendix*, Table S5; Wilcoxon signed-rank test and regression modeling analysis).

The community richness (observed number of species) of Ascomycota increased ~4-fold at noon when air temperature and solar radiation reached their maximum, while the Basidiomycota community structure was remarkably stable over a 24-h period (Fig. 3*B*). The richness of Viridiplantae changed 2- to 3-fold, being greatest at noontime. Of the aforementioned 4 groups of bacteria, the richness of Firmicutes and Proteobacteria communities increased up to 10-fold at noon, while Cyanobacteria and Actinobacteria communities were more stable over the course of a day, with less than a 5-fold increase (Fig. 3*B*). Corresponding alpha-diversity values of the microbial communities throughout DVE2 to 5 are presented in *SI Appendix*, Fig. S9.

Day and Night Biomass Fluctuations. In the time series DVE2 to 5, the total sampled biomass was consistently higher during night

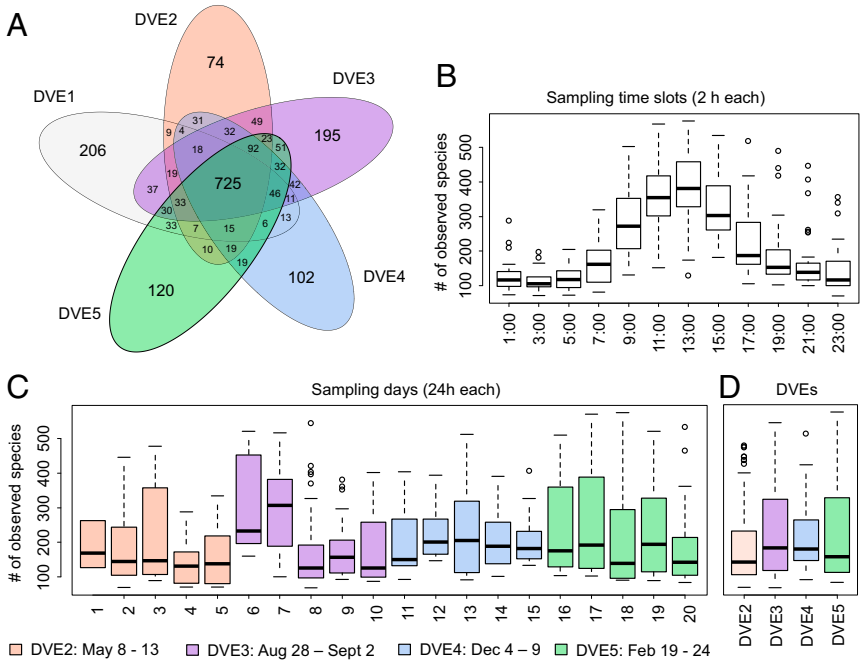


Fig. 2. Temporal variation of the airborne microbial community. (*A*) Venn diagram of species co-occurring in 5 environmental time series (DVE1 to 5). (*B–D*) Box plots of the number of observed species per sampling time slot. (*B*) Intraday distribution of species richness within 24 h (2-h sampling time interval). (*C*) Interday variation of species richness, considering the daily average within 5 d, covering DVE2 to 5. (*D*) Intermonth distribution of species richness, considering the average of 5-d sampling in intervals of 3 mo, covering a total of 13 mo (DVE2 to 5). Outliers are displayed as unfilled circles.

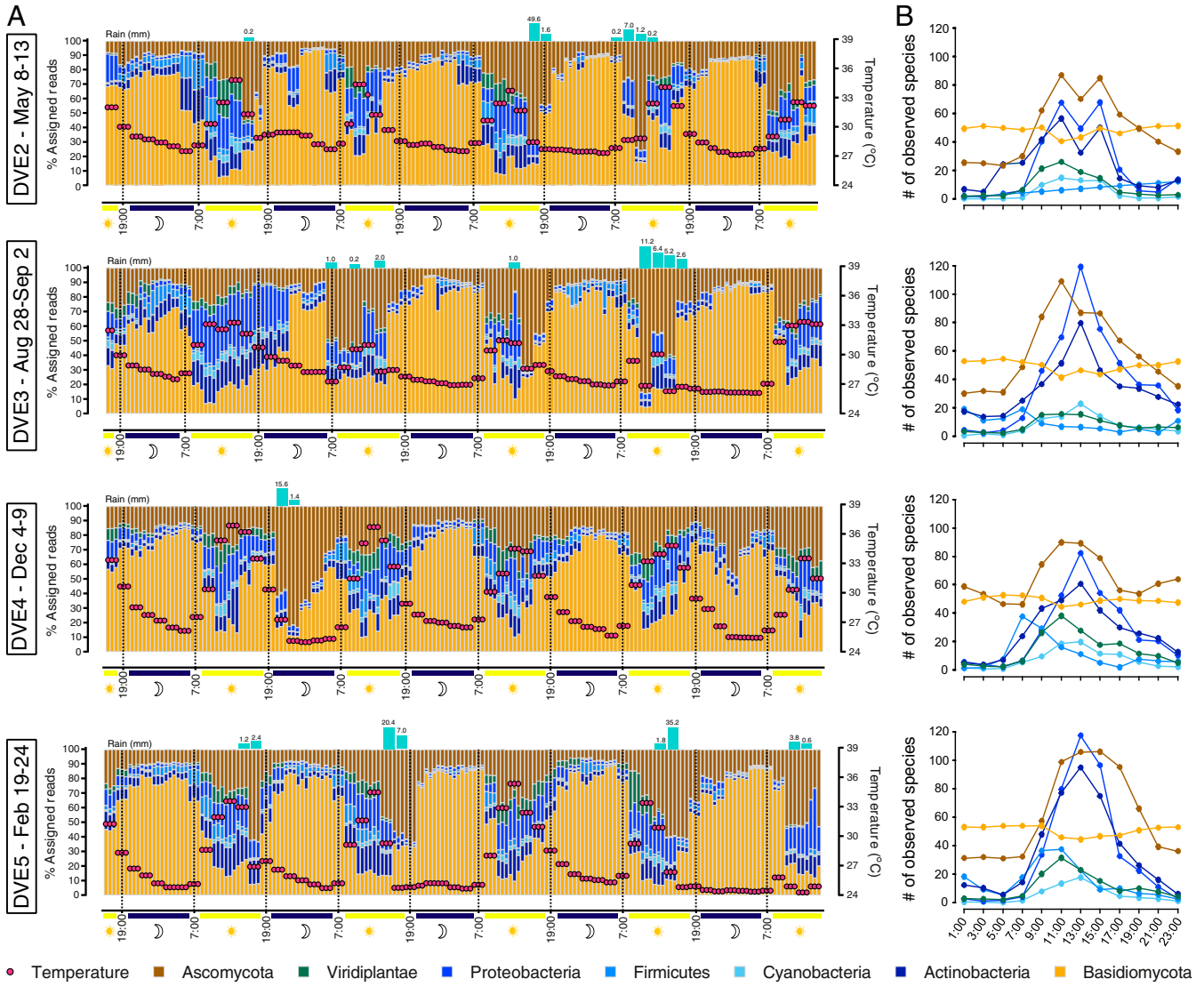


Fig. 3. Taxonomical breakdown and temporal resolution of the air microbial communities. (*A*) Relative abundance of 7 taxonomic groups of organisms in air. Rain events with recorded amounts of precipitation (blue bars) and temperature distribution (red buttons) are denoted for each time series. The timescale is indicated (*Bottom*), as well as day and night samples denoted by sun and moon symbols. (*B*) Observed number of species by taxonomical groups in each of the 4 DVEs. The graph depicts the change in species abundance per group, indicating the robustness of the Basidiomycota community versus strong daily variation of the community composition for the other 6 phyla.

hours (19:00 to 07:00), as approximated by DNA amount extracted per 2-h interval (Fig. 4*A*). The DNA amounts varied between 5 and 500 ng (per 72 m^3 sampled), resulting in up to a 100-fold difference between the early morning hours (05:00) and solar noon (13:00) (e.g., DVE2). This diel effect continued to exist during the dry seasons, but the DNA amount declined ~30-fold to 5 to 150 ng (e.g., DVE3 and 4). Such oscillations within 24 h are also apparent in the pairwise Bray–Curtis similarity analysis among consecutive samples and the periodic oscillation component extracted after applying single-spectrum analysis (Fig. 4*B*).

The significant differences in DNA yield per daytime–nighttime pair were confirmed for DVE2 to 5 using the Wilcox test (DVE2: $P = 7.24 \times 10^{-10}$; DVE3: $P = 3.98 \times 10^{-6}$; DVE4: $P = 4.98 \times 10^{-7}$; DVE5: $P = 4.63 \times 10^{-5}$). Multivariate association analysis confirmed that time of sampling within the day had a stronger effect on microbial community composition than the day on which the sampling was performed (*SI Appendix*, Table S4). The affiliation with a specific sampling time increased in statistical significance when a binary trait corresponding to "day" and "night" was assumed (*SI Appendix*, Table S4). Species that significantly oscillate in the course of a day are listed in *SI Appendix*, Table S6. Microbial community structure (Fig. 4 *C* and *D*) was also more dispersed during the light hours as opposed to a consistent dark-hour community composition ($P = 0.001$ for analysis of multivariate homogeneity of group dispersions in DVE2, DVE3, and DVE5, though less significant in DVE4: $P = 0.14$). There were at least 3 times more bacterial reads during daytime compared with nighttime (*SI Appendix*, Table S7).

To reconcile differences between whole-genome sequencing data used for qualitative analysis and quantitative comparative analysis in this study, we also undertook quantitative PCR analysis. Primers for the 16S and internal transcribed spacer (ITS) regions were used to quantify absolute abundances of bacterial and fungal OTUs, respectively. The comparative analysis showed congruence of relative and absolute species abundances for bacterial taxa but not for fungal organisms due to database and amplification biases (see below and *SI Appendix*, Fig. S10).

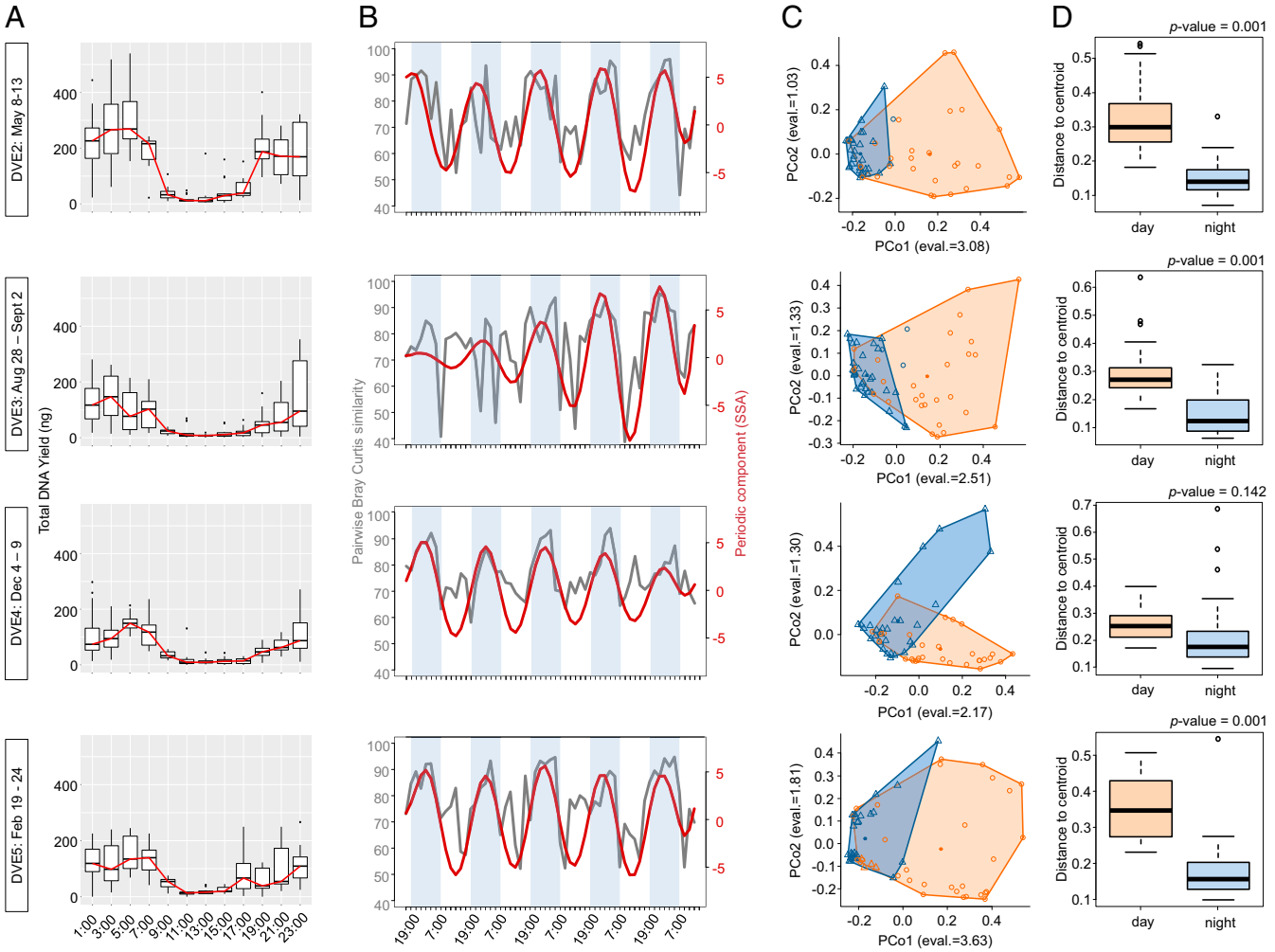


Fig. 4. Diel cycle of the air microbial community. (*A*) Box plots of total DNA yields (ng) collected during DVE2 to 5. (*B*) Pairwise Bray–Curtis similarity among consecutive samples (gray) and the periodic oscillation component (red) computed through single-spectrum analysis. (*C* and *D*) Structures of the airborne microbial communities in DVE2 to 5. (*C*) Areas of the 2 groups of samples representing day (7:00 to 19:00, orange circles) and night (19:00 to 7:00, blue triangles) on the first 2 principal coordinate (PCo1 and PCo2) axes. (*D*) Box plots of the distances to centroid for each group are indicated. *P* values for the permutation-based test of multivariate homogeneity of group dispersions (PERMDISP2) are shown. Unfilled circles represent outliers.

Responses to Changing Environmental Parameters. In addition to diel variation, airborne microorganisms responded to changing physical and chemical parameters of air such as temperature, CO_2, and rain (metadata for DVE2 to 5 are plotted in *SI Appendix*, Fig. S11*A*). In Singapore, the daily temperature cycle is usually consistent, with a minimum temperature of ∼26 °C at night and a maximum of ∼35 °C during the day (*SI Appendix*, Fig. S11*A*). Relative humidity varies from ∼50% during late afternoon to ∼90% at the end of the night. The general pattern of high humidity during nighttime and lower humidity during daytime hours is also observed when the absolute humidity is calculated for each time point (*SI Appendix*, Fig. S11*A*). *SI Appendix*, Table S8 shows the interdependency of temperature and relative humidity, also known as saturation water vapor pressure. The diel cycle of CO_2 in the near-ground–level atmosphere is largely the result of plant respiration and varies from 400 ppm during the night to 460 ppm during the day (*SI Appendix*, Fig. S11*A*). Contrasting windy conditions were recorded during the daytime of DVE3, resulting in an environmental parameter that clearly differentiated it from the other time-series surveys (*SI Appendix*, Fig. S11).

Of all meteorological factors, temperature appears to have the greatest impact on microbial communities in the air (Fig. 5*A*), followed by CO_2 levels (Fig. 5*B*). Increasing air temperature is positively correlated with a higher abundance of bacterial taxa and Ascomycota, while fungal taxa of Basidiomycota exhibit a negative correlation (Fig. 5*C*). This observation is supported by strong statistical correlations (Pearson's $R > 0.75$ or < -0.75). Particularly in DVE2, the abnormal temperature patterns during the second night and fourth day of the sampling event were remarkably mirrored by the pattern of abundance of several microbial taxa (Fig. 5*C*). Notably, positive correlations with temperature increases were seen for all bacterial phyla, while a negative correlation was typically observed only for Basidiomycota fungal species of the Agaricomycetes class (e.g., *Heterobasidion irregulare*, *Moniliophthora roreri*, *Auricularia delicata*, *Tulasnella calospora*). These observations were robust throughout the time series (DVE2 to 5) (see also *SI Appendix*, Table S9), and supported by highly significant associations obtained through regression modeling (*SI Appendix*, Table S10).

In addition, a consistently strong correlation and significant association were detected between CO_2 levels and abundance of Agaricomycetes, mainly represented by the species *A. delicata* (DVE2: $R = 0.59$; DVE3: $R = 0.59$; DVE4: $R = 0.82$; DVE5: $R = 0.66$) and *Rhizoctonia solani* (DVE2: $R = 0.67$; DVE3: $R = 0.59$;

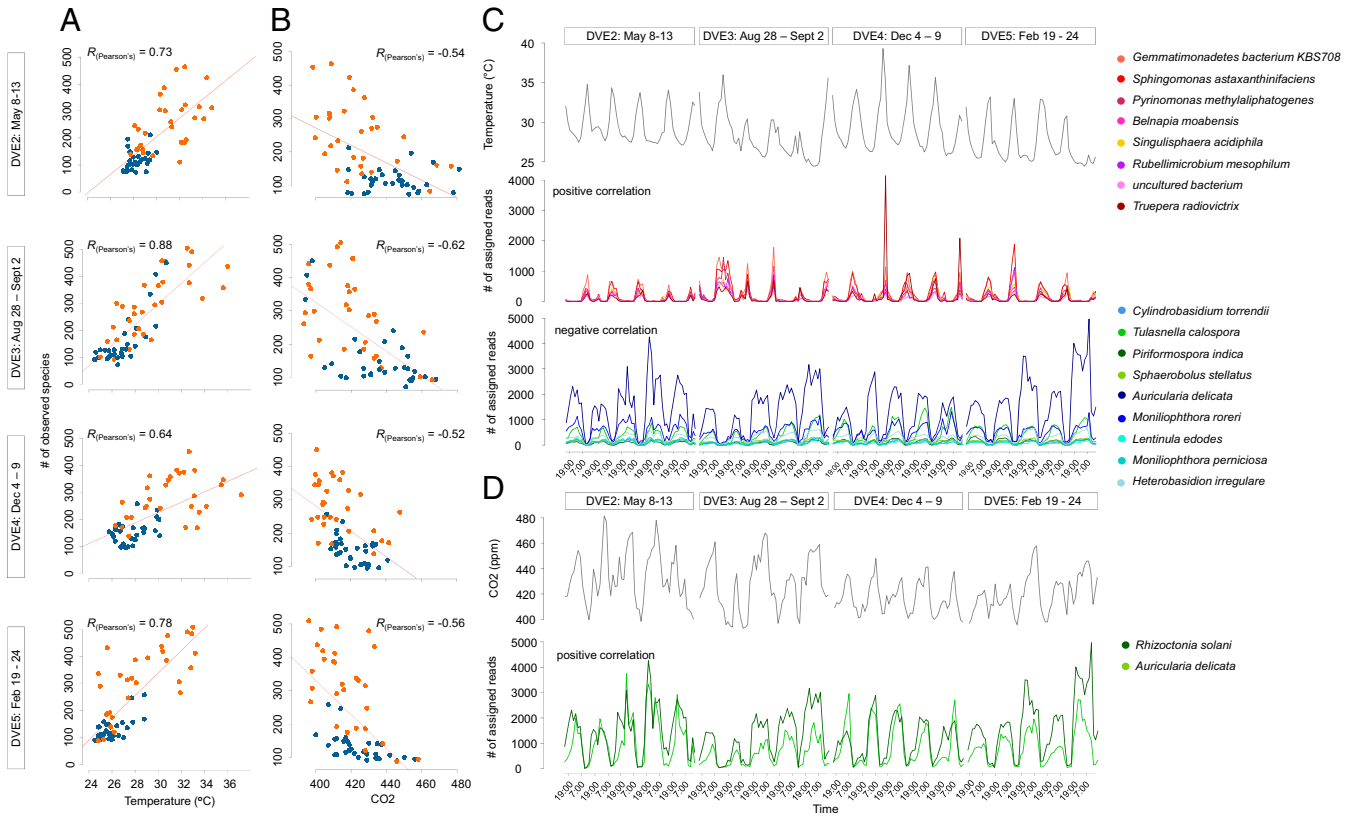
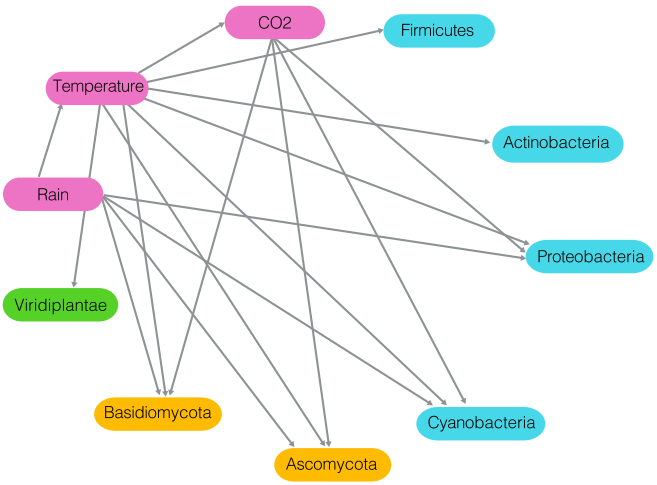


Fig. 5. Diel changes of airborne fungi and bacteria in response to environmental factors. (*A* and *B*) Plotting species richness (number of observed species or chao1 index) against temperature (*A*) and CO_2 (*B*) at a particular time point of sampling. The red line is a regression line (chao1 ~ temperature). Orange and blue dots correspond to day and night time points, respectively. Pearson's correlation indices between the chao1 index and temperature/CO_2 are indicated. (*C*) Temperature. Eight microbial species are observed to vary their abundances, as represented by read counts, in response to temperature changes. The majority of the species responding positively to temperature are bacteria. Nine fungal species of the airborne community respond negatively. (*D*) Carbon dioxide. Two fungal species respond to diel changes in CO_2. The observed fluctuations caused by both environmental factors are highly correlated (Pearson's $R > 0.75$ or <-0.75).

DVE4: $R = 0.84$; DVE5: $R = 0.68$) (Fig. 5*D* and *SI Appendix*, Table S10). Similarly, *Actinomycetospora chiangmaiensis* increased in abundance during rain in DVE2 ($R = 0.93$), while the Pezizomycotina fungi *Diaporthe ampelina* ($R = 0.80$) and *Pseudocercospora fijiensis* ($R = 0.88$) increased during rain events in DVE3. The other Pezizomycotina fungus, *Eutypa lata*, was consistently observed to increase in abundance either directly after a rain event (cross-correlation function [CCF] modeling analysis: DVE2: $R = 0.84$, t −1; DVE4: $R = 0.76$, t −1; DVE5: $R = 0.72$, t −1) or during rain events in DVE3 ($R = 0.81$) (*SI Appendix*, Fig. S12). Similar behavior was observed for another Pezizomycotina fungus, *D. ampelina* (CCF modeling analysis: DVE2: $R = 0.94$, t −10; DVE3: $R = 0.78$, t −1; DVE4: $R = 0.82$, t −1; DVE5: $R = 0.53$, t −1) (*SI Appendix*, Fig. S12).

Modeling. Bayesian network analysis (BNA) was applied to investigate the combined effect of multiple observed environmental variables. The above-identified environmental parameters impacting the abundance of airborne microorganisms were modeled in a BNA (20) (Fig. 6). This was conducted in addition to the statistical analysis and underlines the congruence of environmental parameter oscillations and the dynamics of the microbial communities. Furthermore, the BNA model assesses several scenarios that capture a "typical day" in a tropical climate. These include a "high-temperature scenario" (scenario A), a "low-temperature scenario" (scenario B), as well as an "intense rain scenario" (scenario C), which describe the effects of rapid temperature drops and high humidity caused by local thunderstorms. The BNA model therefore constructs and visualizes a network of probabilistic interactions between various environmental factors and relative abundances of taxonomic groups (for a more detailed description, see *SI Appendix*).

Airborne Particle Distribution and Environmental Pollutants. Different morphologies and sizes of biotic and abiotic particles were visualized using light and scanning electron microscopy (*SI Appendix*, Fig. S13). The particles and microbial cells appear dispersed and not physically attached to a substratum (*SI Appendix*, Fig. S13*A*). Furthermore, no strong correlation between various particle sizes was evident, as observed by optical particle counts (OPCs) (*SI Appendix*, Fig. S14) and the total amount of collected biomass (e.g., total extracted DNA yield) (*SI Appendix*, Table S11). There was also no correlation between particle size and the abundance of the 7 taxonomical groups described above (*SI Appendix*, Table S12). A combination of fluorescence and light microscopy indicated a significant fraction of particles represented abiotic matter (*SI Appendix*, Figs. S13 *B–D*). In addition, multicellular organisms, as well as single-celled organisms with 1 or more nuclei, were identified (*SI Appendix*, Fig. S13*B*). Some of these appeared to have been imaged during the dikaryotic stage in the fungal life cycle of Ascomycota and Basidiomycota (*SI Appendix*, Fig. S13*B*). We also monitored the hourly concentration of environmental pollutants, such as NO_x (nitric oxides) and SO_x (sulfur oxides), in relationship to the reported fluctuations in biomass and taxonomic distribution. However, both chemicals could be excluded as drivers for the diel cycle of airborne microorganisms (*SI Appendix*, Figs. S15 and S16).



	Scenario 1	Scenario 2	Scenario 3
CO2	low 88% high 12%	low 44% high 56%	low 44% high 56%
Temperature	low 0% high 100%	low 100% high 0%	low 100% high 0%
Rain	below 5mm 96% above 5mm 4%	below 5mm 96% above 5mm 4%	below 5mm 0% above 5mm 100%
Viridiplantae	low 9% medium 25% high 66%	low 52% medium 39% high 9%	low 52% medium 39% high 9%
Basidiomycota	low 58% medium 39% high 3%	low 15% medium 29% high 56%	low 88% medium 6% high 6%
Ascomycota	low 11% medium 41% high 48%	low 50% medium 27% high 23%	low 6% medium 6% high 88%
Cyanobacteria	low 9% medium 24% high 67%	low 58% medium 32% high 10%	low 19% medium 33% high 48%
Proteobacteria	low 4% medium 30% high 66%	low 55% medium 36% high 9%	low 6% medium 43% high 51%
Actinobacteria	low 11% medium 34% high 55%	low 50% medium 33% high 17%	low 50% medium 33% high 17%
Firmicutes	low 22% medium 44% high 34%	low 42% medium 26% high 33%	low 42% medium 26% high 33%

Fig. 6. Network of probabilistic interactions between taxonomical groups and environmental parameters inferred from the data. Each node represents a variable of the system, and each edge denotes an influence of one over the other. In addition, each node is associated with a probability table indicating the likelihood of a variable being in a particular state in the absence of information about the system. The probability table of each node is updated (recalculated) once information regarding the state of other nodes becomes available. Scenario 1 models a typical midday time point when the temperature reaches its maximum. In response to high temperature, other variables are recalculated making a new prediction (e.g., CO_2 is likely to be low, while bacteria are likely to be in high relative abundance). Low-temperature scenario 2, typical for a nighttime setting, and scenario 3, a rainfall scenario, modeling a local meteorological event typical for the tropics.

Discussion

Metagenomics analysis of air samples is technically challenging due to the low amount of obtainable biomass and has therefore previously been restricted to long sampling time intervals, large volumes of air (21–23), and amplification-based approaches using gene markers (24–27). The composition of airborne microorganisms was previously assumed to vary continuously due to changes in meteorological, spatial, and temporal factors (9, 25, 28–31). However, sampling across 5 environmental time series in the tropics, spanning up to 13 mo, resulted in a highly robust set of 725 core taxa. Considering the rate of airflow and the opposing direction of monsoon winds across the tropical seasons, the observed robustness of the tropical airborne microbial community was unexpected, particularly when compared with reports from temperate climates (9, 28, 32, 33). A large variation in microbial composition, however, was observed between day and night samples, while the day-to-day and month-to-month variation was less significant (Fig. 2 and *SI Appendix*, Table S7). The robust and reliable assessment of airborne community composition with high temporal resolution enables the interrogation of even ultralow biomass environments to determine the scales and boundaries of understudied ecosystems, such as the atmosphere. Future studies investigating the sources and sinks of airborne microbial communities will have to consider the large taxonomic differences encountered between day and night. Therefore, it is important to note that consistent solar elevation angles (true solar time) need to be considered in future horizontal and vertical surveys, as these will govern the most relevant environmental parameters, such as temperature and relative humidity. The temporal distribution of >1,200 taxa reported in our metagenomics survey is evident from the data from the 5 environmental time series presented here. The mechanisms behind the observed diel biomass fluctuations are at present unknown; however, the active release of biological particulates, such as spores, has previously been reported (34). In contrast, the removal of bioaerosols likely involves the cessation of the active release of biological matter in combination with dry and wet deposition. In addition, the dynamics of bioaerosol concentrations may also be governed by meteorological processes of the near-surface atmosphere, such as microdroplet formation of atmospheric water.

Our in-depth taxonomical analysis of ambient tropical air revealed a microbial diversity with similar complexity to other well-studied ecosystems, such as seawater, soil, and the human gut. The proportion of eukaryotic microorganisms in tropical air was high compared with other planetary ecosystems. Of the eukaryotes, Basidiomycota and Ascomycota fungi were the most abundant, while only small amounts of plant DNA were detected. Unlike marine ecosystems, the air contained only traces of phage and archaea. Although Ascomycota is the richest fungal phylum in nature (35) and usually prevails in the air of temperate climate zones (8), Basidiomycota have been reported as more abundant in the tropics (32), despite their temporal distribution and community dynamics being previously unknown.

The observed high taxonomic diversity of the tropical air microbiome is particularly noteworthy, as it derives from only a fraction of the total sequence data generated. With 9% of a typical nighttime sample and 17% of a daytime sample matching existing sequence information, the air ecosystem appears similar in species diversity and richness to that reported for ocean (36) and soil (37) microbiomes. The richness of air samples exceeded 400 species when considering a 2-h time interval and ~1,200 species across a single time series of 5 d.

The dynamics and long-term stability of tropical air microbial communities are likely driven by environmental factors such as temperature, relative humidity, rain events, solar irradiance, and CO_2 levels. These factors vary in daily repeating patterns and are typical for global-scale tropical atmospheric air masses (e.g., Hadley cells) (38). Importantly, temperature was determined to be the most important environmental parameter driving taxonomic diversity of airborne communities, and it also impacts the water vapor pressure (e.g., relative humidity) (39). Temperature as a driver of microbial diversity was also reported in a survey of the global marine systems (40).

The highly consistent meteorological conditions of the tropics are ideal for studying an atmospheric near-surface ecosystem, as evidenced by the data collection from a single location with high temporal and taxonomic resolution, over extended periods of time. The precise responses of airborne microbiota to fluctuations in temperature and CO_2 levels elucidate how changing physical/environmental factors drive bioaerosol dynamics (41). However, the

underlying mechanisms of the ecosystem responses are presently understudied. Temperature, as the main determinant, affected the prevalence of bacterial and fungal species in an opposing manner, namely more bacteria and less fungi at higher temperatures. Another such example is the response to rain events by organisms such as *E. lata* and *D. ampelina*. Changes in the abundance of these species may be driven by responses to local sources, such as spores being aerosolized by an active mechanism that is triggered by changes in humidity or rainfall.

Some species responded in a consistent fashion to meteorological changes. *Gemmatimonadetes bacterium KBS708* and *Sphingomonas astaxanthinifaciens* responses indicate temperature changes, while abundances of *A. delicata* and *R. solani* indicate increased CO_2 levels, and the abundance of *E. lata* was observed to be associated with rain. These changes in response to meteorological conditions cause a several thousand-fold increase/decrease of some microorganisms (e.g., Fig. 5) and were revealed by the high temporal resolution and sensitivity of the metagenomics-based approach employed. Increases in species abundance on this scale are unlikely to be achieved via growth over a short period of time. This suggests that, at least for the airborne fungal microbiota described in this study, some form of active dispersal mechanism is being deployed, many of which have been unraveled by mycologists over the past decades (34). In contrast, removal of biomass from air need not rely on an active mechanism. Thus, in addition to known means of aerosol removal from the atmosphere by dry (42) and wet (43) deposition, for example, the washout occurring during strong rainfall, we propose a mechanism based on the daily temperature fluctuation in the tropics. In this regard, gaseous water forms nanodroplets as the temperature approaches the dew point during night hours. The formation of these nanodroplets, by aggregating into larger droplets, could contribute to the wet deposition that removes aerosols from the near-surface atmosphere (43). In addition, it is relevant to note that the atmospheric boundary layer contracts during the night, which also positively contributes to an increase in airborne biomass during the night.

Interestingly, the recorded OPC data do not suggest any specific particle size associated with increased amounts of DNA. Hence, the majority of sampled biomass may consist of nonaggregated cells that do not adhere to a specific substratum. This, in turn, would indicate that aggregation of bioaerosols is not a major factor in the deposition process.

Summary. Our study of tropical airborne biomass revealed that also ultralow biomass environments can now be analyzed with high taxonomic resolution. In this regard, tropical air was demonstrated to contain an unexpected richness and diversity of bacterial and fungal taxa, despite only 9 to 17% of the metagenomics reads being identifiable. The high temporal resolution of the sampling approach enabled the observation that airborne microbial communities follow a precise diel cycle, with the daily variation of microbial diversity being much larger within a single day than day-to-day or month-to-month. The robustness of the community structure is underlined by 725 microbial taxa in every environmental time series taken over 13 mo. The experimental approach and analysis from this study will inform future environmental surveys that will investigate the spatial dynamics and natural variability of airborne microbial communities. Lastly, by studying the interrelationship of microbiomes from the 3 planetary ecosystems (atmosphere, terrestrial, and aquatic), it can now be shown that temperature is a global driver of microbial community dynamics.

Materials and Methods

Sample Collection. All air samples were collected at an open rooftop balcony of the School of Chemical and Biomedical Engineering at Nanyang Technological University in Singapore (GPS coordinates N1.346247, E103.679467). The rooftop is located among buildings of similar height, ~20 m aboveground, with tropical forest surrounding the area from west to north of the university campus. The sampling site is predicted to remain within the urban boundary layer throughout the day and night. None of the sampling activities in this study was affected by recurring polluting events such as the Southeast Asian haze. Filter-based air samplers (SASS 3100; Research International) were used with high flow rate, mounted with SASS bioaerosol electret filters (Research International) with 6-cm diameter and particle retention efficiency of 50% for 0.5-μm particles. All air samplers were fixed on a tripod at 1.5-m height above the concrete floor of the balcony. Sampling was performed at a 300 L/min airflow rate for a duration of 2 h. Upon the completion of each sampling event, the SASS filters were transferred to filter pouches and transported back to the laboratory for immediate processing or storage at −20 °C for later use.

A pilot experiment was first conducted to assess the sampling and downstream pipeline. Triplicate air samples were collected during 5 nonconsecutive time slots over the course of 1 d (5:00 to 7:00, 9:00 to 11:00, 12:00 to 14:00, 15:00 to 17:00, and 17:00 to 19:00). The sampling was repeated for 5 consecutive days from February 1 to 5, 2016 (DVE1). A total of 75 samples were collected and processed.

In light of the pilot results, 4 sets of 24-h sampling events were conducted from May 8 to 13 (DVE2), August 28 to September 2 (DVE3), and December 4 to 9 (DVE4) in 2016 and February 19 to 24 (DVE5) in 2017. This sampling regime enabled coverage of Singapore's 2 monsoon seasons: northeast monsoon (December through March) and southwest monsoon (June through September).

DVE2 to 5 were identical-by-design to DVE1, but were extended to cover 120 consecutive hours of sampling for each DVE: Bioaerosol filters were collected and replaced every 2 h, 12 times a day, for 5 consecutive days. Three technical replicates were collected per time point in DVE2 to 5, resulting in 180 samples for each experiment, or a combined total of 720 samples. Successful sample processing, including DNA extraction and library preparation, was achieved for 98.2% of all samples, leading to the following final numbers of sequenced samples per sampling event: 173 (DVE2), 177 (DVE3), 178 (DVE4), and 179 (DVE5).

For each sampling event, filters and reagent blanks were collected, sequenced, and analyzed with the same protocol as the air samples in order to assess potential contamination.

Samples for microscopy were collected using Coriolis μ samplers (Bertin Technologies) over a period of 3 h at a flow rate of 300 L/min during DVE4 on December 9, 5:00 to 8:00. Each Coriolis was set up at a height of 160 cm. Samples were collected in 15 mL ultrapure water, and evaporation was mitigated with a top off-flow rate of 0.5 mL of ultrapure water per min. A blank, 15 mL ultrapure water, was also taken and checked visually under the microscope to ensure the absence of contamination.

Collection of Sensor Data. Sensor data were captured locally using Met ONE HPPC 6+ to measure particle counts (size range 0.3, 0.5, 1.0, 2.0, and 5.0 μm), Rotronic CP11 for CO_2 (ppm), TSI VelociCalc Air Velocity Meter 9545 for wind speed (m/s), temperature (°C), and relative humidity (%), and APOGEE model MU-100 for measuring combined UVA and UVB radiation (W/m^2). Rain data were obtained from the nearby Earth Observatory of Singapore (EOS) from a tipping bucket-type rain gauge (mm). All sensors except those at the EOS were placed in close proximity to the SASS air samplers (0.3 to 3 m away) under a roof to protect them from direct rain and sunlight. The rain sensors from the EOS were located on a rooftop, ~1 km away from the sampling site. One instrument of each type was used for DVE1, with a redundant particle counter added for DVE2 to 5, redundant VelociCalc and CO_2 monitors added for DVE4 and 5, and a UV monitor added for DVE2 to 5. A UV monitor was acquired after DVE1 and was therefore only used for DVE2 to 5. Sampling rates for CO_2, particle count, and VelociCalc sensors were set to 1 min for all DVEs with the exception of DVE4 and 5, for which the particle counter was set to 2- and 5-min intervals, respectively. In addition, the CO_2 sensor was set to record data in 3-min sampling intervals for DVE5. UV measurements were recorded every 30 min. Data from redundant sensors were averaged over the 2-h sampling interval with the exception of rain data, which are reported as total (mm). Global Data Assimilation Systems datasets were acquired from the National Oceanic and Atmospheric Administration.

DNA Extraction. Each SASS filter was first transferred to a sterile tube. PBS/Triton X-100 (PBS-T) was used as washing buffer and added to the tube as needed. Using sterile forceps, the SASS filter was moved up and down in the tube a few times to ensure that the entire filter was soaked in PBS-T. After rinsing, the filter was squeezed with forceps and the buffered solution with washed-out particles was transferred to a sterile conical tube to complete the first washing step. Soaking and squeezing were repeated 3 times for each filter, using fresh PBS-T for each wash step. The combined volume of the

3 washes was subsequently filtered through a 0.02-μm Anodisc filter (Whatman) using a vacuum manifold (DHI). Lastly, DNA was extracted from the Anodisc with the DNeasy PowerWater Kit (Qiagen), according to the manufacturer's protocol with slight modifications (44).

Metagenomics Sequencing. Extracted DNA samples were quantitated on a Qubit 2.0 fluorometer, using the Qubit dsDNA HS (High Sensitivity) Assay Kit (Invitrogen). Immediately prior to library preparation, sample quantitation was repeated on a Promega QuantiFluor fluorometer, using Invitrogen's Picogreen assay. If the concentration of a sample determined by Qubit and Picogreen varied by more than 10%, quantitation was repeated for a third time using the Picogreen assay. In general, a concentration of 0.25 ng/μL and above could be quantitated accurately with both the Qubit dsDNA HS and the Picogreen assay. For concentrations below 0.25 ng/μL, quantitation became less accurate and the Qubit tended to overestimate the concentration. Hence, with the exception of blanks, samples with a concentration of <0.25 ng/μL were not processed further unless absolutely necessary.

High-throughput sequencing libraries were prepared using the Swift Biosciences Accel-NGS 2S Plus DNA Kit, following the instructions provided by the manufacturer. DNA was sheared to ∼450 bp on either a Covaris S220 or E220 focused ultrasonicator. All libraries were dual-barcoded with indices from the Swift Biosciences 2S Dual Indexing Kit.

Library quantitation was performed using Invitrogen's Picogreen assay and the average library size was determined on a Bioanalyzer DNA 7500 chip (Agilent). Library concentrations were normalized to 4 nM and the concentration was validated by qPCR on a ViiA-7 real-time thermocycler (Applied Biosystems) using Kapa Biosystem's Library Quantification Kit for Illumina Platforms. Libraries were then pooled at equal volumes and sequenced on Illumina HiSeq 2500 rapid runs at a final concentration of 10 to 11 pM and a read length of 251-bp paired end (Illumina HiSeq 2500 V2 rapid sequencing chemistry).

Next-Generation Sequencing (NGS) Data Processing and Analysis. The metagenomics data generated for the air samples were processed for adaptor removal and quality trimming with a Phred quality score threshold of Q20 using Cutadapt v. 1.8.1 (45). Two million reads (250 bp) were randomly selected from each sample as a representative set and aligned against the NCBI nonredundant protein database downloaded on February 25, 2016 using the alignment tool RAPSearch v. 2.15 (46, 47).

Resulting alignments were imported into MEGAN v. 5.11.3, which assigns taxon IDs based on NCBI taxonomy (19). To achieve the desired taxonomic specificity, we used the following filtering parameters: min score 100 (bit score), max expected 0.01 (e value), top percent 10 (top 10% of the highest bit score), min support 25 (minimum number of reads required for taxonomic assignment), lowest common ancestry (LCA) percent 100 (naive), and min complexity 0.33 (sequence complexity). LCA for each read on the NCBI taxonomy is assigned using MEGAN's LCA algorithm. In instances where all of the above filtering criteria were fulfilled, reads were assigned to various levels of taxonomic classification ranging from domain to species. In our study, species-level classification was only reached if at least 25 reads uniquely aligned to a single species in the database with a bit score equivalent to a 100% match on the protein level over at least 50% of the 250-bp read. Due to limits of existing public sequence databases, some sequencing reads did not result in meaningful alignments and were classified in the no-hits category. Unassigned reads were categorized in instances when alignments were made but did not fulfill one of the filtering parameters, namely the bit score.

Statistical Analysis of DNA Yield. Technical replicates of total DNA yields were averaged per time point. Wilcox test was used to compare total extracted DNA yields collected during day (7:00 to 19:00) and night (19:00 to 7:00) time slots.

Statistical Analysis of Metagenomics Data. Thresholds for taxon identification used for the presence/absence analysis in the Venn diagram were set at 25 reads per sample analyzed (Fig. 2*A*). For statistical analysis (Figs. 4 *C* and *D* and 5) a more stringent set of criteria was applied. For each time point of DVE2 to 5, the 3 technical replicates were averaged and only taxa which exceeded 25 reads on average across 60 time points were included in the subsequent statistical analysis. As a result, a lower overall number of species was included in the downstream statistical analysis (195 for DVE1, 139 for DVE2, 36 for DVE3, 42 for DVE4, and 176 for DVE5).

Regression modeling in R v. 3.2.3 was used to assess differences in relative abundances of the 7 most common airborne taxa (i.e., Ascomycota, Basidiomycota, Cyanobacteria, Firmicutes, Proteobacteria, Actinobacteria, and Viridiplantae) at different time slots of a sampling day (2-h time resolution) and between day and night (12-h time resolution). Analysis of similarities was applied to assess the significance of differences in microbial communities between day and night, between experimental days in a single time-series experiment, and between time-series DVE2 to 5, as implemented in the *vegan* package in R. Generalized linear regression framework was used to assess the association of sampling time and date factors with the corresponding multivariate species abundance, as implemented in the *mvabund* package in R (correlations between species response variables were accounted for, while other settings were used as default). Pearson's correlations between multivariate species abundance and a number of environmental factors were also estimated (e.g., temperature, millimeters of rain, CO_2, wind velocity, relative humidity). Alpha-diversity indices (chao1, *Simpson D* and *E*) were calculated in QIIME v. 1.8.0 (48). Bray–Curtis dissimilarity distances among centroids for each DVE experiment were calculated with the *vegan* package in R. Principal coordinate analysis was used as ordination method. Single-spectrum analysis (SSA) was applied on the pairwise Bray–Curtis similarities among air samples belonging to consecutive time points (*Rssa* package in R v. 1.0, using default parameters and neigh 50). To compare concordance of the technical replicates, similarity percentages were calculated in Primer v. 7.0.13 (PRIMER-e).

To model existing relations of probabilistic dependence among environmental variables and taxonomical groups, structural and parameter learning of Bayesian networks was performed using the software GeNIe 2.2.2011.0. Variables associated with taxonomical groups were discretized in 3 bins of equal numbers of counts, while variables associated with environmental parameters were discretized in 2 bins following the same principle. The variable associated with the amount of rain was discretized in 2 bins corresponding to rain events that produced either more or less than 5 mm water. Interactions among taxonomical groups were excluded through the introduction of forbidden arcs. In addition, variables associated with taxonomical groups were assigned to a higher temporal tier; this prevented the taxonomical groups from having an effect on variables assigned to a lower (earlier) temporal tier. Structural learning was performed through the Bayesian search algorithm (49), specifying 3 as the maximum number of parents allowed per node. The number of iterations was set to 20, sample size to 50, link probability to 0.1, and prior link probability to 0.0001.

Quantitative PCR. After DNA extraction, a set of selected samples was subjected to qPCR analysis. TaqMan qPCR assays (StepOnePlus; Applied Biosystems, Life Technologies) were carried out to quantitate the number of bacterial, fungal, and *Schizophyllum commune* rRNA copies in the samples. For this purpose, 2 sets of universal primers and probes targeting the 341-to-806 region of the 16S rRNA gene (465 bp) for bacteria and the 18S rRNA gene (350 bp) for fungi were chosen (50, 51). In addition, a set of primers and probes was designed for specific quantification of the fungal species *S. commune* in our samples, based on its conserved 18S rRNA gene sequence.

Primers and probes chosen for bacteria were forward primer 341F 5′-CCTACGGGDGGCWGCA-3′, reverse primer 806R 5′-GGACTACHVGGGTMTCTAATC-3′, and TaqMan probe 6FAM-5′-CAGCAGCCGCGGTA-3′-BBQ. To detect fungi, forward primer FungiQuant-F 5′-GGRAAACTCACCAGGTCCAG-3′, reverse primer FungiQuantR 5′-GSWCTATCCCCAKCACGA-3′, and TaqMan probe FungiQuant-PrbLNA 6FAM-5′-TGGTGCATGGCCGTT-3′-BBQ were used. The custom-designed primers and probe for *S. commune* detection were forward primer 5′-CGCGTCTCCGATGTGATAAT-3′, reverse primer 5′-CTCAGTCAAGAGACGGTTAGAAG-3′, and TaqMan probe 6FAM-5′-TTCTACG-TCGTTGACCATCTCGGG-3′-BBQ. TaqMan Fast Advanced Master Mix was used for the qPCR run with the following conditions: denaturing step: 95 °C for 20 s; cycling step: 35 cycles of 95 °C for 1 s and 60 °C for 20 s.

A qPCR standard of reference sequence was developed to estimate gene copy number (GCN) in the extracted DNA solution based on the Ct values. Bacterial 16S and fungal 18S rRNA genes were amplified using the Hifi HotStart ReadyMix PCR Kit (Kapa Biosystems) according to the manufacturer's protocol. The primers used for amplification were the same primers as the ones used for qPCR. The purified PCR products were cloned into One Shot TOP10 Chemically Competent *E. coli* using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen). Serial dilution of the plasmid DNA solution was used to generate standard curves for qPCR, correlating the known copy number with the corresponding Ct value. qPCR analysis results are presented in terms of GCN per μL.

For *S. commune* qPCR standards, fruiting bodies of *S. commune* were directly collected from the university campus using sterile utensils. The collected *S. commune* samples were then cut and cultivated on malt extract agar with streptomycin. After colonies had formed, a single colony was then

chosen and isolated on separate malt extract agar. DNA was extracted from the isolated colony using the DNeasy PowerWater DNA Kit. Species identification as *S. commune* was then confirmed by PCR with primers ITS4F and ITS5R, followed by Sanger sequencing. Both our universal fungal primers and custom-designed primers for *S. commune* were tested on the extracted DNA sample (original DNA sample, prior to PCR). Both sets of primers and probes showed proper amplification from the DNA sample (qPCR effort with the universal bacteria primers and probes set yielded no amplification). A dilution series of this extracted DNA was then used as the standard for the *S. commune* qPCR effort.

ACKNOWLEDGMENTS. This study was supported by Academic Research Fund Tier 3, Singapore Ministry of Education (Grant MOE2013-T3-1-013). We acknowledge the Institute of Medical Biology–Institute of Molecular and Cell Biology Joint Electron Microscopy Suite for assistance in sample processing, image acquisition, imaging, and data analysis in this study.

1. A. M. Womack, B. J. Bohannan, J. L. Green, Biodiversity and biogeography of the atmosphere. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 3645–3653 (2010).
2. H. Behzad, T. Gojobori, K. Mineta, Challenges and opportunities of airborne metagenomics. *Genome Biol. Evol.* **7**, 1216–1226 (2015).
3. A. M. Delort *et al*., A short overview of the microbial population in clouds: Potential roles in atmospheric chemistry and nucleation processes. *Atmos. Res.* **98**, 249–260 (2010).
4. D. J. Smith *et al*., Free tropospheric transport of microorganisms from Asia to North America. *Microb. Ecol.* **64**, 973–985 (2012).
5. E. Mayol *et al*., Long-range transport of airborne microbes over the global tropical and subtropical ocean. *Nat. Commun.* **8**, 201 (2017).
6. S. M. Burrows *et al*., Bacteria in the global atmosphere—Part 2: Modeling of emissions and transport between different ecosystems. *Atmos. Chem. Phys.* **9**, 9281–9297 (2009).
7. Z. Fang, Z. Ouyang, H. Zheng, X. Wang, L. Hu, Culturable airborne bacteria in outdoor environments in Beijing,China. *Microb. Ecol.* **54**, 487–496 (2007).
8. A. S. Amend, K. A. Seifert, R. Samson, T. D. Bruns, Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 13748–13753 (2010).
9. A. Franzetti, I. Gandolfi, E. Gaspari, R. Ambrosini, G. Bestetti, Seasonal variability of bacteria in fine and coarse urban air particulate matter. *Appl. Microbiol. Biotechnol.* **90**, 745–753 (2011).
10. A.-l. Xu, Z.-w. Song, X.-l. Lang, X. Chen, Y. Xia, Seasonal variability in bacterial and fungal diversity and community composition of the near-surface atmosphere in coastal megacity. *Aerobiologia* **33**, 555–575 (2017).
11. Meteorological Service Singapore, Climate of Singapore, http://www.weather.gov.sg/climate-climate-of-singapore. Accessed 28 June 2019.
12. L. S. Chia, A. Rahman, D. Tay, eds., *The Biophysical Environment of Singapore* (Singapore University Press, 1991), pp. 13–27.
13. A. F. Stein *et al*., NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bull. Am. Meteorol. Soc.* **96**, 2059–2077 (2015).
14. R. L. Kepner, Jr, J. R. Pratt, Use of fluorochromes for direct enumeration of total bacteria in environmental samples: Past and present. *Microbiol. Rev.* **58**, 603–615 (1994).
15. X. Raynaud, N. Nunan, Spatial ecology of bacteria at the microscale in soil. *PLoS One* **9**, e87217 (2014).
16. B. Lighthart, Mini-review of the concentration variations found in the alfresco atmospheric bacterial populations. *Aerobiologia* **16**, 7–16 (2000).
17. J. T. Wilson, J. F. McNabb, D. L. Balkwill, W. C. Ghiorse, Enumeration and characterization of bacteria indigenous to a shallow water-table aquifer. *Ground Water* **21**, 134–142 (1983).
18. J. Qin *et al*.; MetaHIT Consortium, A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
19. D. H. Huson, A. F. Auch, J. Qi, S. C. Schuster, MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
20. N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers. *Mach. Learn.* **29**, 131–163 (1997).
21. S. Yooseph *et al*., A metagenomic framework for the study of airborne microbial communities. *PLoS One* **8**, e81862 (2013).
22. C. Fahlgren, G. Bratbak, R. A. Sandaa, R. Thyrhaug, U. L. Zweifel, Diversity of airborne bacteria in samples collected using different devices for aerosol collection. *Aerobiologia* **27**, 107–120 (2011).
23. J. L. Radosevich, W. J. Wilson, J. H. Shinn, T. Z. DeSantis, G. L. Andersen, Development of a high-volume aerosol collection system for the identification of air-borne microorganisms. *Lett. Appl. Microbiol.* **34**, 162–167 (2002).
24. R. M. Bowers *et al*., Sources of bacteria in outdoor air across cities in the midwestern United States. *Appl. Environ. Microbiol.* **77**, 6350–6356 (2011).
25. E. L. Brodie *et al*., Urban aerosols harbor diverse and dynamic bacterial populations. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 299–304 (2007).
26. S. W. Kembel *et al*., Architectural design influences the diversity and structure of the built environment microbiome. *ISME J.* **6**, 1469–1479 (2012).
27. Z. Wu, Y. Tsumura, G. Blomquist, X. R. Wang, 18S rRNA gene variation among common airborne fungi, and development of specific oligonucleotide probes for the detection of fungal isolates. *Appl. Environ. Microbiol.* **69**, 5389–5397 (2003).
28. J. Fröhlich-Nowoisky, D. A. Pickersgill, V. R. Després, U. Pöschl, High diversity of fungi in air particulate matter. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12814–12819 (2009).
29. R. M. Bowers *et al*., Characterization of airborne microbial communities at a high-elevation site and their potential to act as atmospheric ice nuclei. *Appl. Environ. Microbiol.* **75**, 5121–5130 (2009).
30. I. Gandolfi, V. Bertolini, R. Ambrosini, G. Bestetti, A. Franzetti, Unravelling the bacterial diversity in the atmosphere. *Appl. Microbiol. Biotechnol.* **97**, 4727–4736 (2013).
31. R. M. Bowers *et al*., Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. *Environ. Sci. Technol.* **47**, 12097–12106 (2013).
32. A. M. Womack *et al*., Characterization of active and total fungal communities in the atmosphere over the Amazon rainforest. *Biogeosciences* **12**, 6337–6349 (2015).
33. N. Fierer *et al*., Short-term temporal variability in airborne bacterial and fungal populations. *Appl. Environ. Microbiol.* **74**, 200–207 (2008).
34. D. Savage, M. J. Barbetti, W. J. MacLeod, M. U. Salam, M. Renton, Seasonal and diurnal patterns of spore release can significantly affect the proportion of spores expected to undergo long-distance dispersal. *Microb. Ecol.* **63**, 578–585 (2012).
35. M. Blackwell, The fungi: 1, 2, 3 ... 5.1 million species? *Am. J. Bot.* **98**, 426–438 (2011).
36. M. L. Sogin *et al*., Microbial diversity in the deep sea and the underexplored "rare biosphere." *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12115–12120 (2006).
37. C. Wagg, S. F. Bender, F. Widmer, M. G. A. van der Heijden, Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5266–5270 (2014).
38. J. Lu, G. A. Vecchi, T. Reichler, Expansion of the Hadley cell under global warming. *Geophys. Res. Lett.* **34**, L06805 (2007).
39. M. G. Lawrence, The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications. *Bull. Am. Meteorol. Soc.* **86**, 225–234 (2005).
40. S. Sunagawa *et al*.; Tara Oceans Coordinators, Ocean plankton. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
41. M. H. Y. Leung, D. Wilkins, E. K. T. Li, F. K. F. Kong, P. K. H. Lee, Indoor-air microbiome in an urban subway network: Diversity and dynamics. *Appl. Environ. Microbiol.* **80**, 6760–6770 (2014).
42. L. Zhu *et al*., Spatiotemporal characteristics of particulate matter and dry deposition flux in the Cuihu wetland of Beijing. *PLoS One* **11**, e0158616 (2016).
43. L.-C. Guo, L.-J. Bao, J.-W. She, E. Y. Zeng, Significance of wet deposition to removal of atmospheric particulate matter and polycyclic aromatic hydrocarbons: A case study in Guangzhou, China. *Atmos. Environ.* **83**, 136–144 (2014).
44. I. Luhung *et al*., Protocol improvements for low concentration DNA-based bioaerosol sampling and analysis. *PLoS One* **10**, e0141158 (2015).
45. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
46. Y. Ye, J. H. Choi, H. Tang, RAPSearch: A fast protein similarity search tool for short reads. *BMC Bioinformatics* **12**, 159 (2011).
47. Y. Zhao, H. Tang, Y. Ye, RAPSearch2: A fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* **28**, 125–126 (2012).
48. J. G. Caporaso *et al*., QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
49. G. F. Cooper, E. Herskovits, A Bayesian method for the induction of probabilistic networks from data. *Mach. Learn.* **9**, 309–347 (1992).
50. C. M. Liu *et al*., BactQuant: An enhanced broad-coverage bacterial quantitative real-time PCR assay. *BMC Microbiol.* **12**, 56 (2012).
51. C. M. Liu *et al*., FungiQuant: A broad-coverage fungal quantitative real-time PCR assay. *BMC Microbiol.* **12**, 255 (2012).