

ORIGINAL ARTICLE

Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments

Robert M Bowers¹, Shawna McLetchie², Rob Knight^{3,4} and Noah Fierer^{1,5}

¹Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA; ²Department of Biology, Hanover College, Hanover, IN, USA; ³Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO, USA; ⁴Howard Hughes Medical Institute, University of Colorado, Boulder, CO, USA and ⁵Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA

Although bacteria are ubiquitous in the near-surface atmosphere and they can have important effects on human health, airborne bacteria have received relatively little attention and their spatial dynamics remain poorly understood. Owing to differences in meteorological conditions and the potential sources of airborne bacteria, we would expect the atmosphere over different land-use types to harbor distinct bacterial communities. To test this hypothesis, we sampled the near-surface atmosphere above three distinct land-use types (agricultural fields, suburban areas and forests) across northern Colorado, USA, sampling five sites per land-use type. Microbial abundances were stable across land-use types, with $\sim 10^5$ – 10^6 bacterial cells per m³ of air, but the concentrations of biological ice nuclei, determined using a droplet freezing assay, were on average two and eight times higher in samples from agricultural areas than in the other two land-use types. Likewise, the composition of the airborne bacterial communities, assessed via bar-coded pyrosequencing, was significantly related to land-use type and these differences were likely driven by shifts in the sources of bacteria to the atmosphere across the land-uses, not local meteorological conditions. A meta-analysis of previously published data shows that atmospheric bacterial communities differ from those in potential source environments (leaf surfaces and soils), and we demonstrate that we may be able to use this information to determine the relative inputs of bacteria from these source environments to the atmosphere. This work furthers our understanding of bacterial diversity in the atmosphere, the terrestrial controls on this diversity and potential approaches for source tracking of airborne bacteria.

The ISME Journal advance online publication, 4 November 2010; doi:10.1038/ismej.2010.167

Subject Category: microbial population and community ecology

Keywords: bioaerosol; 16S rRNA; land-use; ice nucleation; pyrosequencing; atmospheric microorganisms

Introduction

Studies of aerosols in the near-surface atmosphere typically focus on non-biological particles despite increasing recognition that biological particles may represent a significant portion of the particulates suspended in the atmosphere (Jaenicke, 2005). Of those studies examining biological aerosols, most have focused on the quantification and identification of fungal spores, pollens and dust mite allergens (Beggs and Kerr, 2000). Although there is a history of research examining bacterial distributions

in the near-surface atmosphere (Lighthart and Shaffer, 1995a, b; Lighthart, 1997; Tong and Lighthart, 1998), the majority of this work has largely been restricted to surveys of culturable bacteria (those bacteria which can readily be grown and isolated under laboratory conditions). As most bacteria cannot be readily cultured (Pace, 1997), the vast majority of airborne bacteria have effectively been excluded from these culture-based studies (Amato *et al.*, 2005) and comprehensive surveys of airborne bacterial diversity are uncommon.

Our limited understanding of bacterial diversity in the atmosphere represents a critical knowledge gap in atmospheric research given that bacteria are ubiquitous in the atmosphere with concentrations ranging between 10^4 and 10^8 cells per m³ of air (Lange *et al.*, 1997; Bauer *et al.*, 2003; Fabian *et al.*, 2005; Albrecht *et al.*, 2007; Bowers *et al.*, 2009).

Correspondence: RM Bowers, Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, UCB 216, Boulder, CO 80309-0216, USA.

E-mail: robert.bowers@colorado.edu

Received 5 July 2010; revised 24 September 2010; accepted 24 September 2010

These airborne bacteria may have important impacts on human health, serving directly as human pathogens or triggering allergic reactions (Schwartz *et al.*, 1995; Hussman, 1996). In addition, bacterial pathogens of agricultural crops and livestock can be transported through the atmosphere (Romantschuk and Bamford, 1986; Gonzalez *et al.*, 2000). More generally, owing to their small size and relatively high concentrations, airborne bacteria are likely to represent a significant portion of the PM_{2.5} aerosol fraction (those aerosols less than 2.5 μm in diameter), an aerosol fraction that is commonly the focus of air quality monitoring and remediation efforts (Boreson *et al.*, 2004; Menetrez *et al.*, 2007). Although it is often assumed that airborne bacteria are merely passive inhabitants of the atmosphere, there is increasing evidence that bacteria have the potential to alter atmospheric processes by serving as ice nucleating particles in the atmosphere (Constantinidou *et al.*, 1990; Mohler *et al.*, 2007, 2008).

Culture-independent molecular methods, such as rRNA gene sequencing and/or hybridization (that is, PhyloChip), are becoming the method of choice for describing the total diversity and composition of microbial communities in environmental samples. Such methods have been used to survey bacterial diversity in the atmosphere (Tong and Lighthart, 2000; Radosevich *et al.*, 2002; Wilson *et al.*, 2002; Angenent *et al.*, 2005; Maron *et al.*, 2005; Brodie *et al.*, 2007; Despres *et al.*, 2007a,b; Fierer *et al.*, 2008a,b; Bowers *et al.*, 2009). However, there are still critical gaps in our understanding of airborne bacterial communities, including our limited understanding of the spatial variation in airborne bacterial community structure and the factors driving this variation. In particular, we know very little about the potential sources of atmospheric bacteria and how changes in land-use or land cover may influence the diversity and composition of airborne bacterial communities.

There are likely two main terrestrial sources of bacteria to the near-surface atmosphere: leaf surfaces and soil (dust) (Lindemann and Upper, 1985; Lighthart and Shaffer, 1994; Lighthart and Shaffer, 1995a,b; Lighthart, 1997). However, the relative importance of these primary sources is not known, nor do we know the extent to which airborne bacterial communities overlap with those bacterial communities found in these source environments. It has previously been suggested that airborne bacterial communities are similar to those in soil (Brodie *et al.*, 2007), but it is not clear whether this finding applies to other near-surface atmospheric environments. We might expect land-use type to have a significant influence on the structure of airborne bacterial communities (Despres *et al.*, 2007a,b) due to shifts in the relative importance of these habitats as sources of bacteria to the atmosphere, or due to the specific nature of the communities found on leaves and soil. For example, because different plant types harbor distinct bacterial communities on their

leaf surfaces, (Redford *et al.*, 2010) we would expect the types of bacteria transported from leaf surfaces to the atmosphere to depend on local vegetation characteristics. Likewise, land-use type may have indirect effects on airborne bacterial community composition due to shifts in meteorological conditions, as previous work has demonstrated that the abundance and composition of airborne bacteria can be affected by local changes in near-surface atmospheric conditions (Jones and Harrison, 2004).

In this study, we describe airborne bacterial diversity in the near-surface atmosphere of the Colorado Front Range and determine the effects of local weather and land-use type on the abundance, composition and ice nucleating potential of these airborne bacterial communities. We hypothesized that the atmosphere harbors unique bacterial communities compared with those found in possible source environments (that is, soil and leaf surfaces), and that land-use type would influence bacterial community structure owing to changes in the types and relative contributions of these bacterial source environments. To test these hypotheses, we analyzed airborne bacterial communities collected from multiple locations within each of the three dominant land-use types in the Colorado Front Range using a bar-coded pyrosequencing approach described previously (Hamady *et al.*, 2008; Lauber *et al.*, 2009; Fierer *et al.*, 2010). We then directly compared the composition of these airborne bacterial communities with those found in other previously collected outdoor air samples and to the potential source environments of soil and leaf surfaces.

Materials and methods

Bioaerosol collection

Aerosol samples were collected from three dominant land-use types in the Colorado Front Range: agricultural fields, suburban areas and forests. Samples were collected from five separate locations per land-use type distributed across the northern half of the Colorado Front Range (Figure 1) for a total of 15 samples. The agricultural sites were located in areas dominated by croplands, and the sampling sites were located immediately adjacent to either corn (*Zea mays*) or barley (*Hordeum vulgare*) fields. The suburban sites varied with respect to the extent of urbanization: sites were located in each of five different population centers from the Denver suburbs north to Fort Collins. The forest sites were in Subalpine fir (*Abies lasiocarpa*), Engelmann spruce (*Picea engelmannii*) and Lodgepole pine (*Pinus contorta*) forests. Table 1 contains details on the sampling locations, sampling times and meteorological conditions at the time of sampling.

Air sampling was conducted over a 2-week period during the summer of 2009. We typically sampled one site per day from each of the three different

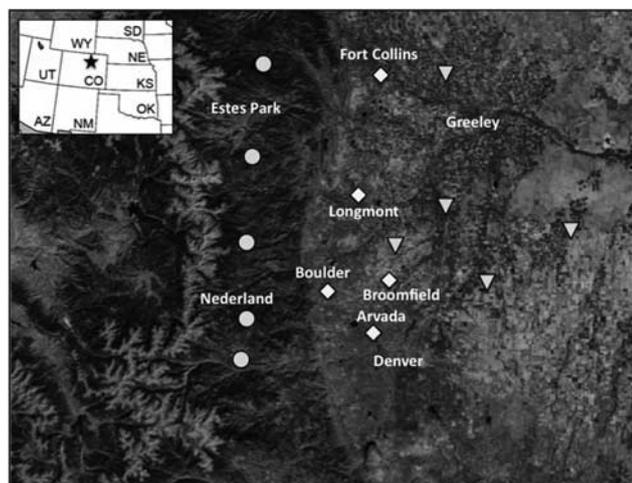


Figure 1 Map of the locations from which aerosol samples were collected over a 2-week time span during the end of June to early July, 2009. Symbols indicate land-use type (circles = forest, diamonds = suburban and the upside down triangles = mixed agricultural sites). Inset graph shows the state of Colorado and the surrounding states, with the star representing the sampled region for this study.

land-use types and rotated the order of collection between land-use types to minimize the effect of time of day on our estimates of bacterial abundance and community composition (Table 1). Air samples were collected at 2.5 m above ground level using vacuum filtration onto 0.22 μm cellulose nitrate filters (Fisher Scientific, Pittsburgh, PA, USA) at a flow rate of ~30 l min⁻¹ per filter for 1.5 h. The sampling height was consistent across all samples even though the canopy height was different for each of the three land-use types. For those sampling sites with significant canopy heights (all forest sites and some suburban sites), the sampling equipment was still set up at the 2.5 m height, but in a central location that would not be touched if all trees, buildings or other standing objects were to fall in the direction of the sampling equipment. The sample filters used in this study were each individually wrapped and contained within a pre-sterilized filter cup. The entire filter cup containing the 0.22 μm filter was used as the sampling device and transported in whole (filter cup with filter inside and sealed in parafilm) back to the lab. This apparatus is simple, but owing to the limited handling of the filter itself, this particular setup drastically reduces any possible contamination during the handling and transport processes. The total volume of air that passed through each filter during each sampling period ranged from 0.8 to 1.1 m³ of air. Triplicate filters were collected for each sample, and one filter of each triplicate set was used for each of the following assays: total bacterial abundance measurements via epifluorescence microscopy, ice nucleation assays and DNA extraction with the subsequent bacterial community analyses. Blank filters were analyzed alongside sample filters to test for

Table 1 Description of samples collected and the corresponding meteorological conditions

Date	Land-use type	Latitude, longitude	Elevation (m)	Approx. sampling location	Start time (hours)	End time (hours)	Temp (C)	Relative humidity (%)	Dew point temp (C)	Wind speed (km h ⁻¹)	Max wind speed (km h ⁻¹)	Wind direction	Barometric pressure (Mpa)	Solar rad. (W m ⁻²)
06/24/09	Agricultural	40°00' N 105°03' W	1552	E. of Boulder, CO	1357	1456	27.7	34.9	54.7	13.3	27.9	SE	10.17	997.43
06/25/09	Agricultural	40°09' N 105°02' W	1504	E. of Longmont, CO	0815	0914	22.5	58.7	57.1	2.4	6.3	N	10.26	461.92
06/25/09	Agricultural	40°21' N 105°01' W	1540	S. of Loveland, CO	1020	1120	26.1	43.3	55.7	6.3	13.6	NE	10.21	534.00
06/29/09	Agricultural	40°31' N 104°45' W	1485	E. of Ft. Collins, CO	1131	1231	27.9	40.2	55.4	16.0	25.0	SSE	10.27	990.38
07/02/09	Agricultural	39°56' N 104°47' W	1542	E. of Brighton, CO	0942	1042	21.9	60.4	57.1	6.8	16.0	??	10.22	717.43
06/16/09	Suburban	40°00' N 105°15' W	1612	Boulder	0740	0900	16.4	51.4	43.4	1.0	5.0	SSE - ENE	10.11	258.59
06/24/09	Suburban	40°10' N 105°06' W	1518	Longmont	1548	1629	29.3	35.1	54.1	8.7	18.9	S	10.22	126.00
06/29/09	Suburban	40°25' N 105°04' W	1528	Ft. Collins	0905	1005	23.7	40.1	48.8	6.4	15.8	S	10.25	638.30
07/01/09	Suburban	40°34' N 105°04' W	1521	Arvada	1353	1453	31.1	28.6	50.2	17.4	35.7	NW	10.05	903.00
07/01/09	Suburban	39°55' N 105°04' W	1645	Broomfield	1536	1636	31.1	22.2	44.5	14.7	27.1	NNE-NNW	10.03	514.54
06/22/09	Forest	40°03' N 105°31' W	2973	Roosevelt Nat'l Forest, CO	0833	0925	15.7	26.2	25.4	7.7	22.3	W	8.74	554.69
06/26/09	Forest	40°18' N 105°31' W	2880	Roosevelt Nat'l Forest, CO	0927	1027	22.6	49.7	52.5	9.6	20.2	NE	9.59	427.50
06/26/09	Forest	40°38' N 105°25' W	2531	Roosevelt Nat'l Forest, CO	1226	1326	24.4	55.3	58.8	3.7	13.8	NE	9.61	445.00
06/30/09	Forest	39°55' N 105°21' W	2574	Roosevelt Nat'l Forest, CO	1151	1252	23.5	12.1	18.5	10.5	26.5	SW	9.15	1051.43
06/30/09	Forest	39°48' N 105°19' W	2549	Roosevelt Nat'l Forest, Golden Gate Canyon State Park, CO	1343	1443	22.4	15.5	22.6	12.0	27.9	EEN	8.85	1103.29

contamination and, following extraction and amplification, blank filters were consistently found to be free of bacterial contaminants. All filter samples were stored on dry ice during transport from the field to the lab, followed by storage at -20°C until the various analyses were performed. Meteorological data (see Table 1) were collected during the sampling periods using a portable Vantage Pro 2 Weather Station (Davis Instruments, Hayward, CA, USA).

Total bacterial abundance measurements and ice nucleation assays

Aerosol particles were shaken free from the filters in a small petri dish in 8 ml of high performance liquid chromatography-grade water for 2 h at room temperature. Filters were routinely examined under a microscope after this shaking process to assure that most visible particles were removed from the filter. The bacterial cells were stained with 4'-6'-diamidino-2-phenylindole, a DNA binding dye and counted at $\times 1000$ magnification using a protocol described previously (Bowers *et al.*, 2009). To the best of our abilities, we counted 4'-6'-diamidino-2-phenylindole stained particles between 0.5 and $10\ \mu\text{m}$ in diameter, which is a rough estimate of the size range for bacterial cells in the environment. Given that there are likely to be bacterial cells that fall outside this size range, as well as small fungal cells that fall within this range, we fully acknowledge that our bacterial abundance measurements are only an approximation of the true abundance levels in the near-surface atmosphere. All filter samples contained at least 20 cells per field at $\times 1000$ magnification in the stock wash solution and at least 30 fields per sample were counted. If there were more than 100 cells per field, the appropriate dilutions were made. Bacterial abundances are reported as the number of cells per m^3 of air, taking into account the dilution and the volume of air that passed through each filter. For ice nucleation assays, particles were removed from the filters using the same protocol described above, except that the filters were shaken at 4°C instead of room temperature to minimize warming, which may alter the ice nucleating capabilities of bacterial cells (Nemecek-Marshall *et al.*, 1993). Drop-freeze assays were performed using a protocol described previously (Bowers *et al.*, 2009), which was modified from a protocol described in Vali (1971). All sample drops freezing above -10°C were recorded as containing high-temperature ice nuclei and were used to calculate the abundance of ice nuclei in each air sample. Throughout this manuscript we use the term high-temperature ice nuclei instead of biological ice nuclei, as we could not be absolutely certain that a particle initiating an ice nucleation event was biological. However, Christner *et al.* (2008) did demonstrate that the majority of high-temperature ice nuclei (those particles which nucleate ice above -10°C) in snow samples were of biological origin.

Likewise, previous work has demonstrated that few, if any, abiotic particles can nucleate ice at temperatures as high as those at which specific bacterial species can nucleate ice (Maki *et al.*, 1974; Vali *et al.*, 1976; Mohler *et al.*, 2007). The relative concentration of high-temperature ice nuclei in each sample was calculated using the cumulative nucleus concentration calculation described in Vali (1971).

DNA extraction and PCR amplification

Bacterial DNA was extracted directly from the filters using the PowerSoil DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA). Individual filters were cut and loaded into the bead tube of the DNA extraction kit and heated to 65°C for 10 min followed by 2 min of vortexing. The remaining steps of the DNA extraction were carried out according to the manufacturer's instructions. Bacterial community composition was determined using a recently developed bar-coded pyrosequencing procedure, which facilitates multiplexed sequencing of partial 16S rRNA genes. The protocol used in this study is identical to that described previously (Bowers *et al.*, 2009; Fierer *et al.*, 2008a,b), including both the PCR conditions and primer sequences. Negative controls (both no template and template from unused filters) were included in all steps of the process, from DNA extraction to PCR amplification, in order to check for contamination. PCR amplicons from each sample were pooled at approximately equal amounts into a single tube for pyrosequencing at the University of South Carolina Environmental Genomics Core Facility on a 454 Life Sciences Genome Sequence FLX (Roche, Branford, CT, USA) machine. With this protocol, we obtained sequence read lengths that averaged 260 bp in length. Previous work has shown that reads of this length over the V2 region of the 16S rRNA gene region provide sufficient information for accurate phylogenetic clustering of bacterial communities (Liu *et al.*, 2008) and accurate taxonomy assignment (to at least the family level of taxonomic resolution (Liu *et al.*, 2008).

Sequence analysis

Sequences were analyzed and processed using the recently developed QIIME package (Caporaso *et al.*, 2010b). Briefly, the QIIME pipeline takes all sequences from a single pyrosequencing run and assigns sample IDs using a mapping file and the barcode assigned to each sample. Sequences were removed from the analysis if they were < 200 bp in length, had a quality score of < 25 , contained ambiguous characters, an uncorrectable barcode, or did not contain the primer sequence. The remaining sequences were clustered into phylotypes using CD-HIT (Li *et al.*, 2001, 2002; Li and Godzik, 2006) with a minimum coverage of 99% and a minimum identity of 97%. A representative sequence was

chosen for each phylotype by selecting the longest sequence that had the highest number of hits to other sequences of that particular phylotype. Representative sequences were aligned using PyNAST (Caporaso *et al.*, 2010a) against the Greengenes coresets (DeSantis *et al.*, 2006). Taxonomy assignments were made using the ribosomal database project classifier (Wang *et al.*, 2007). A phylogenetic tree containing the aligned sequences was then produced using FastTree (Price *et al.*, 2009). To determine the phylogenetic similarity between sample types, the weighted UniFrac metric was used (Lozupone *et al.*, 2007). UniFrac distances are based on the fraction of branch length shared between two communities within a phylogenetic tree made from the 16S rRNA gene sequences of all bacterial communities being compared. The UniFrac algorithm provides an estimate of the overall phylogenetic similarity between each pair of communities and therefore, avoids some of the pitfalls associated with taxon-based community analyses (Lozupone *et al.*, 2007).

We compared the airborne communities found in the Colorado Front Range samples (across land-use types) with a previously-described dataset of airborne bacterial communities, which was composed of samples collected at a high-elevation research site, Storm Peak laboratory (3220 m) located in northern Colorado (Bowers *et al.*, 2009) to determine whether these sample sets share similar bacterial taxa. We also conducted a meta-analysis combining the 16S rRNA gene sequence datasets of the Front Range and high-elevation (Storm Peak laboratory) air samples to two separate datasets representing a diversity of soil types across North and South America ($N=88$) (Lauber *et al.*, 2009) and to leaf-surface samples consisting of more than 60 unique tree species ($N=112$) (Redford *et al.*, 2010) in order to directly compare the air communities with those communities found in potential source environments. For this analysis, QIIME was used in a manner similar to that described above with the following modifications. As different fragments of the 16S rRNA gene were sequenced in different studies, the sequences could not be clustered directly, so instead of using CD-HIT for phylotype assignment, sequences were assigned to phylotypes using BLAST against the Greengenes database (DeSantis *et al.*, 2006) to identify their closest matching sequences. This BLAST-based sequence assignment is identical to the method described in Hamady *et al.* (Hamady *et al.*, 2009). Briefly, this method uses the Greengenes coresets (DeSantis *et al.*, 2006), which was formatted into a BLAST database using formatdb, and subsequently used for BLAST searches of all air sequences, soil and leaf-surface sequences from the studies mentioned above using megablast. The resulting hit tables were parsed to make sample mapping files, wherein each hit was mapped to its closest hit in the formatted Greengenes database. Query sequences that had no hit below an e-value threshold of $1e-10$

were excluded from the analyses (20 203 of 307 341 or 6% of total sequences were removed). The remaining sequences were clustered into phylotypes using a minimum coverage of 99% and a minimum identity of 97%. Owing to unequal sampling efforts (differing number of sequences per sample) the dataset was rarefied to 500 sequences per sample to remove sample heterogeneity. Alpha diversity was assessed by examining rarefaction curves. A tree containing the same set of sequences from the Greengenes core set, available on the Fast UniFrac website (http://128.138.212.43/fastunifrac/tutorial.psp#sample_download) was used to assess the phylogenetic relationships between air, soil and leaf-surface datasets.

Statistical analyses

The variation in total microbial abundance and ice nuclei concentrations across the three land-use types were assessed using an ANOVA, followed by *post-hoc* tests using a Bonferroni adjusted *P*-value for each pair-wise comparison. Relationships between microbial community similarity (determined using UniFrac), bacterial abundance, ice nuclei concentrations and meteorological parameters were investigated using Mantel tests as implemented in the PRIMER V6 software package (Clarke and Warwick, 2001). We also used this software package to conduct permutational MANOVA tests to determine whether any categories of samples harbored significantly distinct bacterial communities. To create the heatmap displayed in Figure 5, the heatmap function in *R* (V2.11.0) was used on the relative abundance values of the most dominant bacterial phyla and subphyla across the grouped environment types: air, soil and leaf surface. Replicate sample types (air from each of the three land-use types, air from Storm Peak laboratory, leaf-surface samples from Gymnosperms and Angiosperms, and soils binned by pH), were clustered using hierarchical agglomerative clustering generating Euclidean distances between sample types to construct the dendrogram. The SIMPER (similarity percentage) function in PRIMER V6 was used to identify bacterial taxa indicative and unique to the soil and leaf-surface environments, with the relative abundances of these 'indicator taxa' subsequently used to determine the relative contribution of each source environment to the bacterial communities in the air samples. Briefly, SIMPER determines the contribution of each taxonomic category to the average Bray–Curtis dissimilarity between pairs of grouped samples (for example, the different environment types).

Results

Meteorological conditions

Atmospheric conditions during the 2-week sampling period were typical of summer-time

conditions across the Colorado Front Range (Table 1). Meteorological conditions were quite variable across the 15 sampling periods and there was more variability in meteorological conditions between the five sampling events within a given land-use type than between different land-use types. With the exception of barometric pressure (which we would expect to be lower at the higher-elevation forest sites), none of the measured meteorological variables (see Table 1) were significantly different between the land-use types (ANOVA, overall $P \gg 0.05$, $F \ll 1.5$).

Concentrations of bacteria and high-temperature ice nuclei

Mean bacterial abundances ranged from 1.0×10^5 to 2.6×10^6 cells per m^3 air across the agricultural, suburban and forest locations (Figure 2). Total bacterial abundances were relatively stable with no significant differences between land-use types (ANOVA, overall $P = 0.94$, $F = 0.07$). Similarly, there were no statistically significant relationships between bacterial abundances and any of the measured meteorological variables listed in Table 1 ($r < 0.3$ and $P > 0.05$ in all cases).

The concentrations of high-temperature ice nuclei were considerably more variable than the total bacterial abundances and were significantly different between the three land-use types ($P = 0.05$, $F = 3.64$) (Figure 2). The agricultural land-use type had the highest abundance of high-temperature ice nuclei, with an average of 36 nuclei per m^3 air. The suburban and forested land-use types had far fewer high-temperature ice nuclei with 18 and 4.5 ice nuclei per m^3 air, respectively (Figure 2). This shift in ice nuclei abundances across land-use types did not correlate to any of the measured meteorological variables ($r < 0.3$, $P > 0.05$ in all cases), suggesting

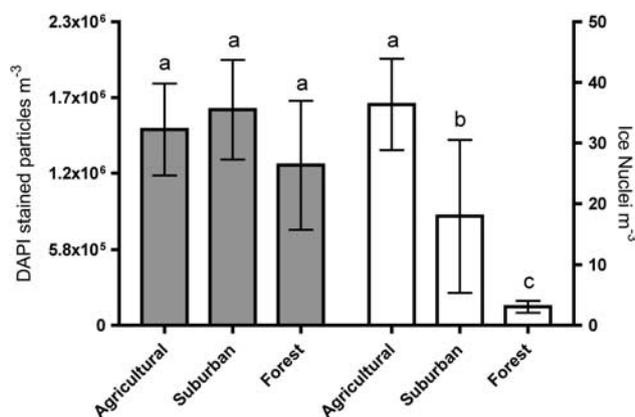


Figure 2 Left y axis, (■) Total bacterial abundance in the collected samples from the Colorado Front Range as determined via direct microscopy, and right y axis, (□) total number of high-temperature ice nuclei as determined via the drop-freeze assay. Letters above the bars indicate a significant difference at $P < 0.05$. Error bars indicate ± 1 s.e.m.

that land-use, not the prevailing meteorological conditions at the time of sampling, had the strongest influence on near-surface concentrations of high-temperature ice nuclei.

Composition of airborne bacterial communities across land-use types

A total of 10,997 partial 16S rRNA gene sequences were generated from the 15 aerosol samples for an average of 730 sequences per sample (range of 540–930 sequences per sample). From the 10,997 sequences, 671 unique phylotypes were identified (with a phylotype defined as those sequences sharing $\geq 97\%$ identity). To compare diversity levels between samples and control for differences in sequencing depth between samples, we conducted rarefaction analyses with 500 randomly selected sequences per sample. At this depth of coverage, we identified an average of 93 phylotypes per sample (range 63–152) and found no effect of land-use type (ANOVA, $P = 0.27$, $F = 1.67$) on phylotype richness (Supplementary Figure S1A), or on the overall phylogenetic structure (Supplementary Figure S1B) across the collected air samples. Actual richness was not estimated as rarefaction analysis revealed that observed richness continually increased as more sequences were included. The airborne bacterial communities were all dominated by sequences assigned to the *Proteobacteria*, *Actinobacteria* and *Firmicutes* phyla. Within the *Proteobacteria*, the betaproteobacterial subphylum was most abundant, with this subphylum being almost entirely composed of representatives from the *Burkholderiales* family of bacteria (92% of the betaproteobacterial sequences).

The shifts in bacterial community composition observed across land-use types are evident if we examine the relative abundances of specific bacterial taxa (Figure 3). For example, the *Actinobacteria*

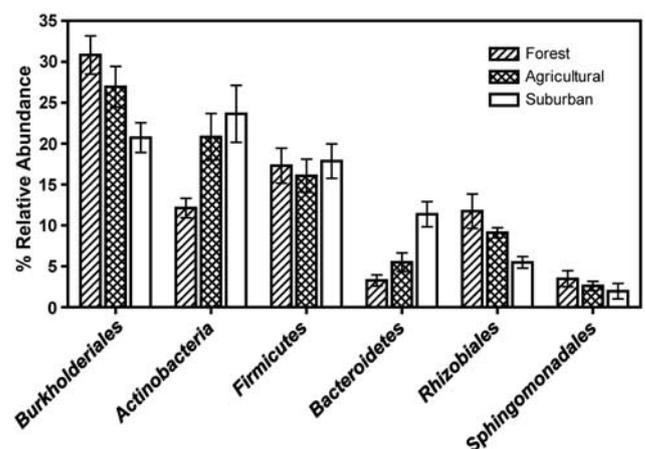


Figure 3 The dominant bacterial phyla and subphyla found in air samples collected from across the three land-use types of the Colorado Front Range: forest ($n = 5$), agricultural ($n = 5$) and suburban ($n = 5$). Proteobacterial groups are designated by the Greek symbols α and β . Error bars indicate ± 1 s.e.m.

were in greatest abundance from the agricultural and suburban locations (averaging 21 and 24%, respectively) than in the forest samples (12%). The relative abundance of *Bacteroidetes* was higher in the air from the suburban land-use (11.4%) than in the agricultural and forest land-use types (6 and 3% respectively). Likewise, forest air had the highest number of *Rhizobiales* sequences (12%), and the suburban land-use had the fewest *Rhizobiales* sequences (6%).

None of the measured meteorological characteristics (singly or in combination) were significantly related to the observed shifts in bacterial community composition, with pair-wise distances between communities calculated using the UniFrac algorithm ($r < 0.3$, $P > 0.05$ in all cases). Instead, land-use type was significantly related to bacterial community structure (Figure 4a) as confirmed by the permutational MANOVA analyses (Main Effect, $P = 0.001$). Using a Bonferroni corrected P -value of 0.02 to determine significance ($\alpha = 0.05/3$ pair-wise comparisons), each pair of land-use types harbored significantly distinct airborne bacterial communities ($P < 0.02$). This effect of land-use type on the overall structure of bacterial communities is evident in Figure 4a, wherein, the samples from the different land-use types separate with minimal overlap across the first principal coordinate axis.

Comparing airborne bacterial communities with those in potential source environments

The bacterial communities characterized in the land-use samples described above were significantly different from those airborne bacterial communities collected at Storm Peak laboratory (using identical sampling and analytical procedures) in winter-like conditions during the early spring of 2008 (Bowers *et al.*, 2009) (Figure 4b, $P < 0.001$). The main taxa driving these differences were the *Actinobacteria* (20% in land-use air samples and 3% of the Storm Peak laboratory samples) and *Firmicutes* (17% in land-use air samples and 1.5% of the Storm Peak laboratory samples) and the *Pseudomonadales* family (1.7% across land-use air samples and 26% of the Storm Peak laboratory air samples). The majority of the pseudomonads in the Storm Peak laboratory samples were related to cold-adapted bacteria such as the *Psychrobacter* genus and these same cold-adapted organisms were entirely absent from the land-use samples collected in the summer.

If we compare all of the airborne bacterial communities (both those described in this study and those from the Storm Peak laboratory study (Bowers *et al.*, 2009) with those communities found in potential source environments, we find that the air communities are relatively similar to one another and quite distinct from those bacterial communities found on leaves and in soil (Figure 4c). The observed phylotype richness and phylogenetic diversity was far lower in the air samples and

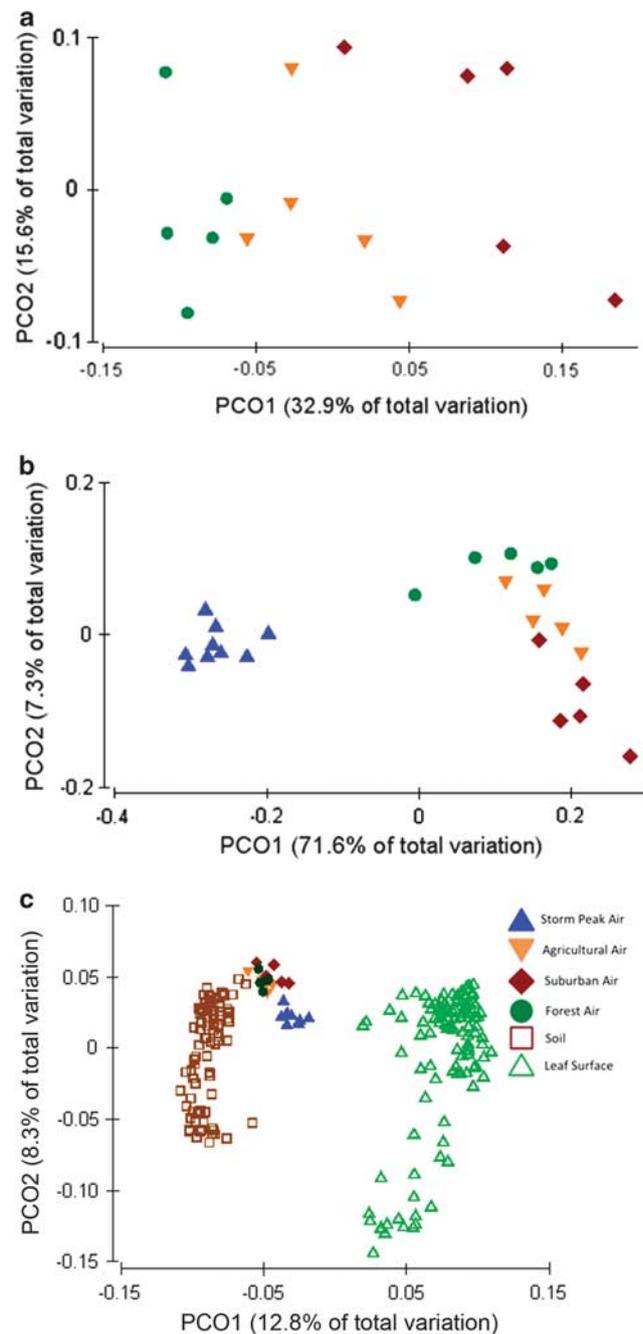


Figure 4 Principal coordinates analysis (PCoA) of the pair-wise distances between bacterial communities as calculated using the UniFrac algorithm (a–c). (a) represents the relationship between the bacterial communities of the three dominant land-use types of the Colorado Front Range and (b) shows the relationship between the bacterial communities from the Colorado Front Range and those collected at 3200 m elevation at Storm Peak laboratory. Panel (c) show the relationships between the bacterial communities found in the three distinct habitats: air, leaf-surfaces and soil. Filled symbols refer to the air samples across land-use type and from Storm Peak laboratory and the open symbols correspond to the two source environments, leaf-surface and soils.

leaf-surface samples compared with the soil samples, as evidenced by the corresponding rarefaction curves (Supplementary Figures S1A and B).

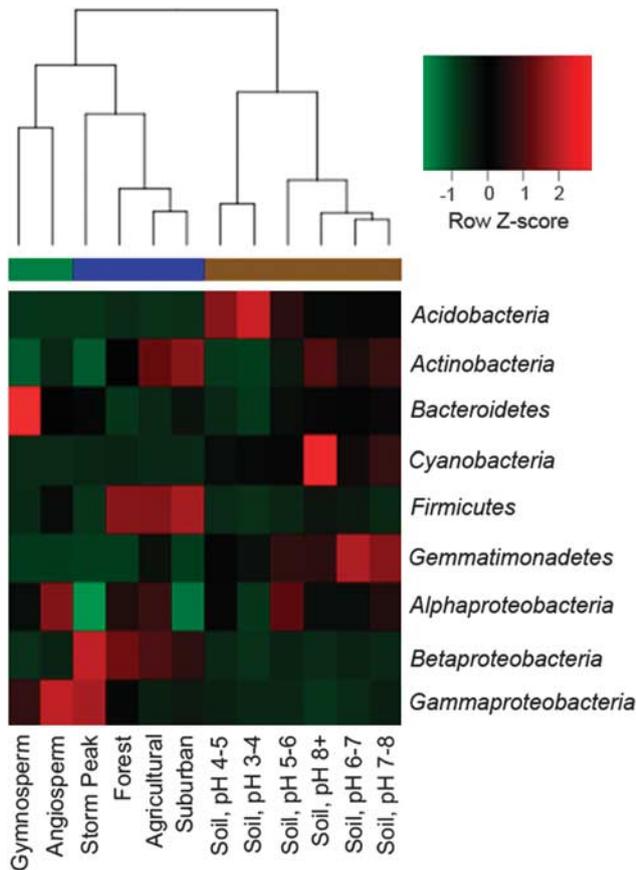


Figure 5 Heatmap displaying the relative abundances across a wide range of air, leaf-surface and soil samples. Soil samples were binned into six different pH categories. Leaf-surface samples were split into gymnosperms and angiosperms and air samples are represented by the Storm Peak laboratory samples collected at the Storm Peak research facility (3200 m above sea level), and the air samples of the three dominant land-use types of the Colorado Front Range: forest, agricultural and suburban sample types. These three ecosystem level groups: air, leaf-surface and soils are designated as blue, green and brown on the color bar, respectively. Environments are clustered based on the percent relative abundance of the nine phylum to subphylum level classifications shown as rows in this figure. Each row was scaled so that the mean of each taxonomic group across sample types was calculated and colored by corresponding z-score of each cell. The heatmap was made with R version 2.11.0 using the heatmap function.

The specific taxa driving the differences in community composition from the three environments (air, leaves and soil) are shown in Figure 5. Although each of these three environments harbor significantly distinct bacterial communities (permutational MANOVA $P < 0.001$), the hierarchical clustering at the top of Figure 5 implies that the leaf-surface is more similar to the air samples than the air samples are to the soil samples, a pattern likely driven by the very low abundance of *Acidobacteria* in air and on leaves compared with the high relative abundance of *Acidobacteria* in most soil environments (Lauber *et al.*, 2009). Also of note is the high relative abundance of *Gammaproteobacteria* sequences on

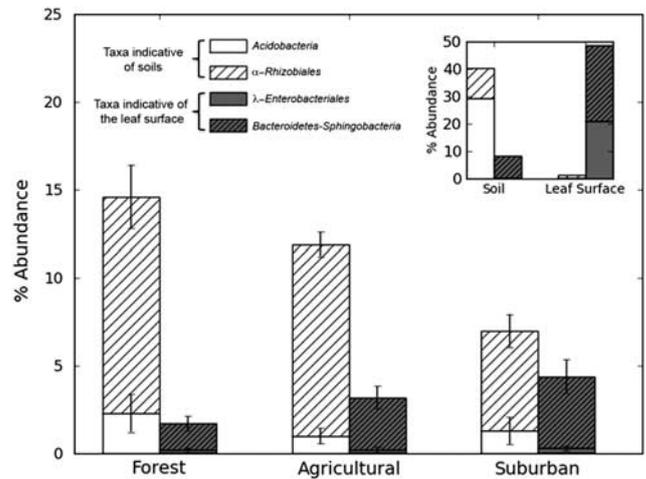


Figure 6 Percent relative abundance of airborne bacterial taxa across the three land-use types that were designated as indicator taxa from soil and leaf-surface environments. Inset. The indicator taxa derived from soil and leaf-surface environments and their percent relative abundances.

the leaf surfaces and in the air samples from both the forest land-use type and from the Storm Peak laboratory samples (Figure 5).

We compared the relative abundances of those taxa indicative of either the soil or leaf-surface environments across the 15 collected aerosol samples in order to estimate the relative contribution of these source environments with the bacteria found in the air samples collected from across the land-use types. The bacterial taxa identified as being indicative of the soil or leaf-surface environments and their relative abundances in the land-use air samples (forest, agricultural and suburban) are shown in Figure 6, and the same indicator taxa from the leaf and soil environments are shown as the inset graph in Figure 6. The airborne communities from the forest sites had the highest relative abundance of taxa indicative of soil whereas those airborne communities collected from the suburban land-use type had the highest relative abundance of taxa indicative of leaf-surface bacteria.

Discussion

The concentrations of bacteria in the collected samples were relatively constant and did not change in a predictable manner with changes in meteorological conditions or land-use type. However, the concentrations measured in this study are similar to the ranges reported from other near-surface air samples wherein, similar culture-independent methods were employed (Matthias-Maser and Jaenicke, 2000; Fabian *et al.*, 2005; Bowers *et al.*, 2009). One might expect there to be differences in airborne bacterial concentrations across diverse land-use types as some studies have shown very high bacterial concentrations in and around

agricultural areas (Albrecht *et al.*, 2007; Lange *et al.*, 1997). However, these studies were conducted near likely point sources, such as a compost facility (Albrecht *et al.*, 2007) and a pig farm (Lange *et al.*, 1997), whereas our goal was to estimate the ambient bacterial abundances from sites that were not necessarily adjacent to biological point sources. We also hypothesized that changing meteorological conditions would lead to predictable shifts in microbial abundances (Jones and Harrison, 2004); however, we observed no significant relationships between meteorological variables and bacterial abundances. Supporting our observations, Matthias-Maser *et al.* (Matthias-Maser and Jaenicke, 2000) demonstrated that only larger bioparticles (that is, fungal spores, pollen and plant debris) exhibit significant shifts in abundance with changes in atmospheric conditions and that concentrations of smaller particles (that is, bacteria. $2\mu\text{m} > r > 0.2\mu\text{m}$) exhibit minimal variation. Although we might observe more variability in bacterial abundances, if sampling over a longer time period, our results suggest that there is minimal short-term variability in bacterial concentrations in the near-surface atmosphere related to differences in meteorological conditions or land-use characteristics.

In contrast to the near-constant concentrations of total bacterial cells, land-use type did have a significant effect on the total number of high-temperature ice nuclei (those nuclei which freeze $> -10^\circ\text{C}$). The concentration of ice nuclei in the agricultural air samples was, on average, two and eight times higher than in the air samples collected from the suburban and forest land-use types, respectively (Figure 2). Based on previous work (Christner *et al.*, 2008), we tentatively assume that the majority of these high temperature ice nuclei represent intact microbial cells. If this is true, it is interesting to note that the majority of known ice nucleating bacterial species have been isolated from annual plant species, including many of our most common agricultural crop species (Lindow *et al.*, 1978; O'Brien and Lindow, 1988). Furthermore, Lindow *et al.* (1978) demonstrated that coniferous tree species possess relatively few ice nuclei active bacteria. This difference in plant-associated ice nucleating bacteria may explain why our forested sites (dominated by conifers) had lower concentrations of ice nuclei than the other two land-use types.

The airborne bacterial communities sampled from the Colorado Front Range are qualitatively similar to those identified from other surface-level air surveys (Radosevich *et al.*, 2002; Bowers *et al.*, 2009; Despres *et al.*, 2007a,b; Fierer *et al.*, 2008a,b), as much of the bacterial community was dominated by *Actinobacteria*, *Firmicutes*, *Betaproteobacteria* and *Gammaproteobacteria*. We also note the overall similarity between the bacterial communities in the air samples collected from the Colorado Front Range with those collected from a high elevation site during early spring when the surrounding

land-surface was largely snow covered (Figure 4b). The airborne communities are not identical, but they do share similar dominant taxa and the air communities are far more similar to one another than they are to bacterial communities of the other terrestrial habitats (and possible source environments) (Figure 4c).

The air bacterial communities surveyed in this study were not panmictic as the atmosphere over the different land-use types harbored bacterial communities that were significantly different from one another (Figure 4a). This subtle, yet significant, effect of land-use type on the airborne bacterial community composition suggests that local terrestrial environments are a likely source for at least a portion of the near-surface atmospheric community. However, the similarity between the spatially and seasonally separated air samples used in this interhabitat analysis (land-use samples and high-elevation samples) also indicates that a fraction of the airborne community may be ubiquitously distributed throughout the near-surface atmosphere (Figure 4c).

Using indicator taxa analysis, we identified taxa largely responsible for the separation of the soil and leaf-surface communities (Figure 4c and Figure 6, inset), and then used these indicator taxa to determine the relative importance of these potential environments as sources of bacteria to the atmosphere across the three Colorado Front Range land-use types. Of the indicator taxa representing either soil or leaf-surface environments (Figure 6, inset), the *Rhizobiales* soil indicator and *Sphingobacteria* leaf-surface indicator were the two most prevalent bacterial groups found in the air samples across land-use types (Figure 6). Forest and agricultural sites had more of a soil signature, whereas the suburban sites were slightly more enriched with taxa indicative of the leaf-surface habitat (Figure 6). The high soil influence and relatively low leaf-surface influence on the airborne forest communities were a somewhat surprising outcome. However, this pattern may be a result of the relatively low leaf biomass (and leaf-associated microbial biomass) in these forests, and/or a result of sampling below the forest canopy. However, regardless of location, the overall structure of the communities found in air are distinct from those found in soil or on leaves (as evidenced from Figure 4c) and this is likely the result of air communities being a mixture of bacteria from soils and leaves and/or certain taxa from these source environments not being able to survive or remain intact in the atmosphere. Additional research will be needed to determine the generality and accuracy of this approach as we have no way to independently confirm these findings. Future work will need to integrate molecular surveys of airborne communities with atmospheric chemical transport models across more extensive sample sets in order to more confidently determine bacterial sources and map their atmospheric

trajectories over extended time periods. However, just as we can use bacterial taxa indicative of animal feces to track fecal contamination of bodies of water (Brownell *et al.*, 2007; Abdelzaher *et al.*, 2010; Wery *et al.*, 2010) or use bacterial communities on the hands of different individuals to identify touched objects (Fierer *et al.*, 2010), we may be able to use bacterial taxa indicative of possible source environments to identify and compare inputs of bacteria with the atmosphere.

Conclusion

The bacterial communities of the near-surface atmosphere across the Colorado Front Range differed across the three distinct land-use types. These differences could not be predicted from the prevailing meteorological conditions, suggesting that the characteristics of the local terrestrial surfaces (for example, vegetation cover, land management and amount of bare soil) have a greater influence on airborne community composition than local atmospheric conditions. These patterns were further investigated using a bacterial 'source tracking' analysis, to identify possible differences in the relative importance of soils and leaf surfaces as sources of atmospheric bacteria. Although the atmosphere is a unique microbial habitat that harbors bacterial communities that are fundamentally distinct from those found in soil and on leaf surfaces, we demonstrate that the differences in airborne bacterial communities above the different land-use types were likely owing, in part, to differences in the relative contributions of bacteria from specific source habitats.

Acknowledgements

We would like to thank Donna Berg-Lyons for her assistance with the pyrosequencing preparation, Jesse Stombaugh for his assistance with the bioinformaticians, Dr Ray Fall for providing access to the equipment used for the ice nucleation assays, and Dr John Basey for providing us with access to a fluorescent microscope. We would also like to thank members of the Fierer lab for their helpful input on this paper. This study was supported by grants to N.F. from the Cooperative Institute for Research in Environmental Sciences, the National Science Foundation and the U.S. Department of Agriculture.

References

Abdelzaher AM, Wright ME, Ortega C, Solo-Gabriele HM, Miller G, Elmir S *et al.* (2010). Presence of pathogens and indicator microbes at a non-point source subtropical recreational marine beach. *Appl Environ Microbiol* **76**: 724–732.

Albrecht A, Witzemberger R, Bernzen U, Jackel U. (2007). Detection of airborne microbes in a composting facility

by cultivation based and cultivation-independent methods. *Ann Agric Environ Med* **14**: 81–85.

Amato P, Ménager M, Sancelme M, Laj P, Mailhot G, Delort A-M. (2005). Microbial population in cloud water at the Puy de Dôme: Implications for the chemistry of clouds. *Atmos Environ* **39**: 4143–4153.

Angenent LT, Kelley ST, Amand A, Pace NR, Hernandez MT. (2005). Molecular identification of potential pathogens in water and air of a hospital therapy pool. *Proc Natl Acad Sci USA* **102**: 4860–4865.

Bauer H, Giebl H, Hitzemberger R, Kasper-Giebl A, Reischl G, Zibuschka F *et al.* (2003). Airborne bacteria as cloud condensation nuclei. *J Geophys Res* **108**: 4658–4663.

Beggs CB, Kerr KG. (2000). The threat posed by airborne micro-organisms. *Indoor Built Environ* **9**: 241–245.

Boreson J, Dillner AM, Peccia J. (2004). Correlating bioaerosol load with PM_{2.5} and PM₁₀ concentrations: a comparison between natural desert and urban-fringe aerosols. *Atmos Environ* **38**: 6029–6041.

Bowers RM, Lauber CL, Wiedinmyer C, Hamady M, Hallar AG, Fall R *et al.* (2009). 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.

Brodie EL, DeSantis TZ, Moberg Parker JP, Zubietta IX, Piceno YM, Andersen GL. (2007). Urban aerosols harbor diverse and dynamic bacterial populations. *Proc Natl Acad Sci USA* **104**: 299–304.

Brownell MJ, Harwood VJ, Kurz RC, McQuaig SM, Lukasik J, Scott TM. (2007). Confirmation of putative storm-water impact on water quality at a Florida beach by microbial source tracking methods and structure of indicator organism populations. *Water Res: Identifying Sources of Fecal Pollution* **41**: 3747–3757.

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**: 266–267.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Christner BC, Morris CE, Foreman CM, Cai R, Sands DC. (2008). Ubiquity of biological ice nucleators in snowfall. *Science* **319**: 1214.

Clarke KR, Warwick RM. (2001). A further biodiversity index applicable to species lists: variation in taxonomic distinctness. *Mar Ecol-Progress Series* **216**: 265–278.

Constantinidou HA, Hirano SS, Baker LS, Upper CD. (1990). Atmospheric dispersal of ice nucleation-active bacteria: the role of rain. *Phytopathology* **80**: 934–937.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.

Despres V, Nowoisky J, Klose M, Conrad R, Andreae MO, Poschl U. (2007a). Molecular genetics and diversity of primary biogenic aerosol particles in urban, rural, and high-alpine air. *Biogeosci Discuss* **4**: 349–384.

Despres VR, Nowoisky JF, Klose M, Conrad R, Andreae MO, Poschl U. (2007b). Characterization of primary biogenic aerosol particles in urban, rural, and high-alpine air by DNA sequence and restriction fragment analysis of ribosomal RNA genes. *Biogeosciences* **4**: 1127–1141.

- Fabian MP, Miller SL, Reponen T, Hernandez MT. (2005). Ambient bioaerosol indices for indoor air quality assessments of flood reclamation. *J Aerosol Sci* **36**: 763–783.
- Fierer N, Hamady M, Lauber CL, Knight R. (2008a). The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc Natl Acad Sci USA* **105**: 17994–17999.
- Fierer N, Liu Z, Rodriguez-Hernandez M, Knight R, Henn M, Hernandez MT. (2008b). Short-term temporal variability in airborne bacterial and fungal populations. *Appl Environ Microbiol* **74**: 200–207.
- Fierer N, CL Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. (2010). Forensic identification using skin bacterial communities. *Proc Natl Acad Sci USA* **107**: 6477–6481.
- Gonzalez AJ, Landeras E, Mendoza MC. (2000). Pathovars of *Pseudomonas syringae* causing bacterial brown spot and halo blight in *Phaseolus vulgaris* L. are distinguishable by ribotyping. *Appl Environ Microbiol* **66**: 850–854.
- Hamady M, Lozupone C, Knight R. (2009). Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J* **4**: 17–27.
- Hamady M, Walker JJ, Harris KJ, Gold NJ, Knight R. (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* **5**: 235–237.
- Hussman T. (1996). Health effects of indoor-air microorganisms. *Scand J Work Environ Health* **22**: 5–13.
- Jaenicke R. (2005). Abundance of cellular material and proteins in the atmosphere. *Science* **308**: 73.
- Jones M, Harrison RM. (2004). The effects of meteorological factors on atmospheric bioaerosol concentrations - a review. *Sci Total Environ* **326**: 151–180.
- Lange J, Thorne P, Lynch N. (1997). Application of flow cytometry and fluorescent *in situ* hybridization for assessment of exposures to airborne bacteria. *Appl Environ Microbiol* **63**: 1557–1563.
- Lauber CL, Hamady M, Knight R, Fierer N. (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol* **75**: 5111–5120.
- Li W, Godzik A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.
- Li W, Jaroszewski L, Godzik A. (2001). Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17**: 282–283.
- Li W, Jaroszewski L, Godzik A. (2002). Tolerating some redundancy significantly speeds up clustering of large protein databases. *Bioinformatics* **18**: 77–82.
- Lighthart B. (1997). The ecology of bacteria in the alfresco atmosphere. *FEMS Microbiol Ecol* **23**: 263–274.
- Lighthart B, Shaffer BT. (1994). Bacterial flux from chaparral into the atmosphere in mid-summer at a high desert location. *Atmospheric Environ* **28**: 1267–1274.
- Lighthart B, Shaffer BT. (1995a). Viable bacterial aerosol particle size distributions in the midsummer atmosphere at an isolated location in the high desert chaparral. *Aerobiologia* **11**: 19–25.
- Lighthart B, Shaffer BT. (1995b). Airborne bacteria in the atmosphere surface layer: temporal distribution above a grass seed field. *Appl Environ Microbiol* **61**: 1492–1496.
- Lindemann J, Upper CD. (1985). Aerial dispersal of epiphytic bacteria over bean plants. *Appl Environ Microbiol* **50**: 1229–1232.
- Lindow SE, Arny DC, Upper CD. (1978). Distribution of ice nucleation-active bacteria on plants in nature. *Appl Environ Microbiol* **36**: 831–838.
- Liu Z, DeSantis TZ, Andersen GL, Knight R. (2008). Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucl Acids Res* **36**: e120–e130.
- Lozupone CA, Hamady M, Kelley ST, Knight R. (2007). Quantitative and qualitative (beta) diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* **73**: 1576–1585.
- Maki LR, Galyan EL, Chang-Chien M-M, Caldwell DR. (1974). Ice nucleation induced by *Pseudomonas syringae*. *Appl Environ Microbiol* **28**: 456–459.
- Maron P-A, Lejon DPH, Carvalho E, Bizet K, Lemanceau P, Ranjard L *et al.* (2005). Assessing genetic structure and diversity of airborne bacterial communities by DNA fingerprinting and 16S rDNA clone library. *Atmospheric Environ* **39**: 3687–3695.
- Matthias-Maser S, Jaenicke R. (2000). The size distribution of primary biological aerosol particles in the multiphase atmosphere. *Aerobiologia* **16**: 207–210.
- Menetrez MY, Foarde KK, Esch RK, Dean TR, Betancourt DA, Moore SA *et al.* (2007). The measurement of ambient bioaerosol exposure. *Aerosol Sci Technol* **41**: 884–893.
- Mohler O, DeMott PJ, Vali G, Levin Z. (2007). Microbiology and atmospheric processes: the role of biological particles in cloud physics. *Biogeosciences Discuss* **4**: 2559–2591.
- Mohler O, Georgakopoulos DG, Morris CE, Benz S, Ebert V, Hunsmann S *et al.* (2008). Heterogeneous ice nucleation activity of bacteria: new laboratory experiments at simulated cloud conditions. *Biogeosciences* **5**: 1425–1435.
- Nemecek-Marshall M, LaDuca R, Fall R. (1993). High-level expression of ice nuclei in a *Pseudomonas syringae* strain is induced by nutrient limitation and low temperature. *J Bacteriol* **175**: 4062–4070.
- O'Brien RD, Lindow SE. (1988). Effect of plant species and environmental conditions on ice nucleation activity of *Pseudomonas syringae* on leaves. *Appl Environ Microbiol* **54**: 2281–2286.
- Pace NR. (1997). A molecular view of microbial diversity and the biosphere. *Science* **276**: 734–740.
- Price MN, Dehal PS, Arkin AP. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**: 1641–1650.
- Radosevich JL, Wilson WJ, Shinn JH, DeSantis TZ, Andersen GL. (2002). Development of a high-volume aerosol collection system for the identification of air-borne micro-organisms. *Lett Appl Microbiol* **34**: 162–167.
- Redford A, Bowers RM, Knight R, Linhart Y, Fierer N. The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environ Microbiol*, 2010; e-pub ahead of print 1 June 2010, doi:10.1111/j.1462-2920.2010.02258.x.
- Romantschuk M, Bamford DH. (1986). The causal agent of halo blight in bean, *pseudomonas syringae* pv. *phaseolicola*, attaches to stomata via its pili. *Microb Pathog* **1**: 139–148.

- Schwartz D, Thorne P, Yagla SJ, Burmeister LF, Olenchock SA, Watt JL *et al.* (1995). The role of endotoxin in grain dust-induced lung disease. *Am J Respir Crit Care Med* **152**: 603–608.
- Tong Y, Lighthart B. (1998). Effect of simulated solar radiation on mixed outdoor atmospheric bacterial populations. *FEMS Microbiol Ecol* **26**: 311–316.
- Tong Y, Lighthart B. (2000). The annual bacterial particle concentration and size distribution in the ambient atmosphere in a rural area of the Willamette Valley, Oregon. *Aerosol Sci Technol* **32**: 393–403.
- Vali G. (1971). Quantitative evaluation of experimental results on the heterogeneous freezing nucleation of supercooled liquids. *Journal of Atmospheric Sciences* **28**: 402–409.
- Vali G, Christensen M, Fresh RW, Galyan EL, Maki LR, Schnell RC. (1976). Biogenic ice nuclei. Part II: bacterial sources. *Journal of the Atmospheric Sciences* **33**: 1565–1570.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.
- Wery N, Monteil C, Pourcher A-M, Godon J-J. (2010). Human-specific fecal bacteria in wastewater treatment plant effluents. *Water Res* **44**: 1873–1883.
- Wilson KH, Wilson WJ, Radosevich JL, DeSantis TZ, Viswanathan VS, Kuczmariski TA *et al.* (2002). High-density microarray of small-subunit ribosomal DNA probes. *Appl Environ Microbiol* **68**: 2535–2541.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)