1. Supplementary Methods

1.1 DNA-Containing Cell Enumeration

Cells in the precipitation were preserved by adding sodium borate-buffered formalin (pH=8.2, stored at room temperature) to a final concentration of 5% v/v. Triplicate samples from each precipitation event were processed by filtering 10 mL of sample onto 0.22µm, 25mm black polycarbonate filters (Millipore) and staining with a final concentration of 25X SYBR-Gold (Invitrogen) for 15 min. in the dark. Cell density estimates were obtained using an epifluorescence microscope (Olympus bx51) and data from 60 fields of view (1 field of view=34636 μm^2).

1.2 DNA Extraction, Sequencing, and Analysis

The thawed 47mm Supor PES membrane filters were transferred to a laminar flow hood and cut into small pieces using sterile scissors. The filter pieces were transferred to a bead beating tube from the FastDNA™ SPIN Kit for Soil (MP Biomedicals, Santa Ana, CA). The DNA was extracted according to the manufacturer's protocol, with the following modifications: in step 4, a mini bead beater was used to homogenize extracts for 70 seconds; step 5 was performed for 8 minutes; a 15 mL tube was used for step 7; and 100 µL of DNase/pyrogen-free water was used to elute DNA in step 16. The extracts obtained were further purified using steps 14-22 of the manufacturer's protocol for the MoBio Power Soil Kit (MoBio Laboratories, Carlsbad, CA). In these steps, 160 µL of solution C4 was added to the DNA extraction. The manufacturer's protocol was followed for all other steps with the exception of adding 25

23 µL of solution C6 and incubating at room temperature prior to centrifugation. The DNA
24 was stored at −20 °C prior to polymerase chain reaction (PCR) amplification.
25 The V4 region of the bacterial 16S rRNA gene was PCR amplified in triplicate
26 from each DNA extract for sequencing on the Illumina MiSeq (Illumina, Inc., San Diego,
27 CA, USA). The PCR was carried out using the barcoded primers and methods of
28 Caporaso et al. (1). The 25µL reaction contained the following components: 5 Prime
29 Master Mix (1X), 0.5 µM 515F, 0.5 µM 806R and nuclease free water (13 µL) and 2.0
30 µL of DNA. Thermal-cycling was carried out in an Eppendorf PRO S Master Cycler
31 (Eppendorf North America, Hauppauge, NY, USA) under the following conditions:
32 initial denaturation at 94 °C for 3 min. followed by 32 cycles of denaturation at 94 °C for
33 1 min., annealing at 50 °C for 1 min. extension at 72 °C for 1 min. 45 s, and followed by
34 a final extension at 72 °C for 10 min. The amplicons obtained were evaluated by
35 electrophoretic separation on a 1% agarose gel buffered with Tris-acetate-EDTA and
36 stained with ethidium bromide. PCR products from the triplicate reactions were pooled,
37 purified, and concentrated using the Qiagen MinElute PCR Cleanup kit (Qiagen,
38 Valencia, California, USA) with the optional 35% guanidine-HCl wash step to ensure
39 removal of large primer-dimers. Amplicons were stored at −20 °C until sequencing at
40 Idaho State University's Molecular Research Core Facilty (Pocatello, ID, USA) on the
41 Illumina MiSeq Platform (Illumina Inc., San Diego, CA) using the V2 500 bp kit.
42 Sequence data were analyzed using the mothur software package (2). Contigs
43 were assembled and parsed on the basis of unique barcodes attached to the 806R primer
44 (1). Sequences that did not contain exact matches to the primer and barcodes utilized in
45 the PCR amplification were discarded. Sequences were filtered for quality with a 50-base

sliding window and a minimum average quality score of 25, and those containing ambiguous bases, homopolymers (> 7 bases), or having lengths > 259 bases were eliminated from the dataset.

1.3 Inorganic Chemical Analyses

Conductivity and pH were measured using a multi-parameter PCSTest probe (Oakton Instruments, Vernon Hills, IL). Samples of deionized water were routinely analyzed and served as procedural blanks. A Dionex ICS-3000 ion chromatography system was used to determine the concentration of major ions in the precipitation samples. The system was equipped for anion separation with a 4×250 mm RFIC™ IonPac® AS18 column (Dionex Corporation, CA, USA), using a water:potassium hydroxide eluent, and a 4×250 mm RFIC™ IonPac® CS16 column (Dionex Corporation, CA, USA) with a water:methanesulfonic acid eluent. Guard columns (Dionex Corporation, CA, USA) preceded each column (4×50 mm RFIC™ IonPac® AG18 guard column was used for the anion channel and a 4×50 mm RFIC™ IonPac® CG16 guard column for the cation channel). Each sample was analyzed in triplicate. With this method, 2 carboxylic acids (formate and oxalate) and 13 inorganic ions (F^-, Cl^-, NO^{2-}, Br^-, NO^{3-}, SO_4^{2-}, PO_4^{3-}, Li^+, Na^+, NH_4^+, K^+, Mg^{2+} and Ca^{2+}) could be quantified. The limit of detection, calculated as three times the standard deviation of the field blanks, was between 0.1 to 0.8 µM for all ions reported.

1.4 Organic Carbon Concentration, Fluorescent Characterization, and PARAFAC Modeling

DOC concentrations were obtained from a GE Sievers 900 Total Organic Carbon Analyzer. An average DOC measurement was calculated from three measurements of organic carbon concentrations for each precipitation filtrate sample (sample size ~25 mL). Blank samples of Milli-Q Water were measured between each sample to monitor successive sample-to-sample contamination throughout instrument use. Acidification was not necessary prior to experimentation due to an internal acidification step within the instrument.

A Horiba Jobin Yvon Fluoromax-4 Spectrofluorometer generated the Excitation Emission Matrices (EEMs) of the fluorescent dissolved organic matter (DOM) in the precipitation samples. This instrument is equipped with a Xenon lamp light source and a 1 cm path length quartz cuvette was used for all measurements. Excitation (Ex) wavelengths were scanned from 240-450 nm in 10 nm intervals and emission (Em) was recorded between 300-560 nm in 2 nm increments. Data integration time was 0.25 s and data acquisition was carried out in signal/reference mode using a 5 nm bandpass on both Ex and Em monochromators, normalizing the fluorescence Em signal with the Ex light intensity. Absorbance spectra (190-1100nm) was incorporated into the spectral correction calculations of primary and secondary inner filter effects for post-processing the fluorescence data to generate EEMs (3, 4). Spectra were blank corrected against purified water from a Milli-Q system each day. A Parallel factor analysis (PARAFAC) model was generated in MATLAB by drEEM and the N-way toolbox scripts (5) to determine individual DOM fluorescing components in the EEMs.

1.5 Meteorological Data Collection and Analysis

On-site meteorological data was collected continuously with a weather station (Vantage Pro, Davis Instruments), and supplemented with data obtained locally through the Louisiana Agriclimatic Information System (automated weather station at Ben Hur Agricultural Fields, ~5 km SE of the main sampling location).

Storm classification was based on cloud top data retrieved from The National Weather Service (NWS) archive of Geostationary Operational Environmental Satellite-East (GOES-East) infrared and visible satellite imagery and Next Generation Radar (NEXRAD) Level III radar reflectivity provided by the National Climatic Data Center's (NCDC) website (https://www.ncdc.noaa.gov/nexradinv/). Troposphere temperature profiles were retrieved from The NWS radiosonde data archive of stations located in Lake Charles, Louisiana (LCH, station number 72240) and Slidell Muni, Louisiana (LIX, station number 72233).

Classification of convective and nimbostratus precipitation based on radar reflectivity and satellite imagery was carried out according to previous methods (6, 7). Tropospheric stability indices from NWS soundings (8, 9) were used to confirm the presence or lack of Convective Available Potential Energy (CAPE), which indicates the presence of convection. Convection occurs when the surface of the earth is heated unevenly, leading to the warming of air directly above the heated surface. This warmer air is more buoyant than the surrounding air and begins to rise. Once this "parcel" of warm air rises to the "convective condensation level" (CCL), water vapor will begin to condense and form water droplets, and subsequently, a cloud develops. If precipitation came from a cloud which was formed in the presence of convection the Convective Condensation Level (CCL) was used to estimate the height of the cloud base (Supplementary Figure S1 a).

Convergence (Supplementary Figure S1 b) occurs when a low pressure system is present. Air in high pressure regions moves towards lower pressure regions, leading to the convergence of air masses, forcing the air to move up in the atmosphere. Warm front (Supplementary Figure S1 c) lifting occurs when a warm front advances and the less dense, warm air within that warm front is displaced upward over cooler, denser air ahead of it. Cold front (Supplementary Figure S1 d) lifting occurs when a cold front advances and displaces the warmer air ahead of it upward. Orographic lifting was not observed in this study and is not depicted. If precipitation came from a cloud which was formed in the absence of convection and by one of the prior three methods listed, the Lifted Condensation Level (LCL) was used to estimate the height of the cloud base (Supplementary Figure S1 b-d).

As an example of how trajectories were analyzed, panels f and e in Supplementary Figure S1 show the altitudes and trajectories used for a particular rain event that occurred on August 25, 2014. The cloud base for this event was at approximately 1160 mAGL, and the cloud top was at approximately 16800 mAGL. Given that the cloud system developed through convection (Supplementary Figure S1 a), the six altitudes chosen (depicted in panels a-d as the gray dotted lines) for HYSPLIT backward trajectory analysis were *below* the cloud at the time of precipitation in Baton Rouge. The six backward trajectories were then examined for previous interactions with the MBL or surface using the tdump csv files generated by HYSPLIT, in which trajectory height, MBL height, and surface height are listed at hourly intervals for each trajectory. If at any point along the trajectory history (histories of which ranged from 120-168 hours) the air masses descended into the MBL or interacted with the ground, the geographic coordinates these interactions were recorded for

that event and plotted in R to determine the corresponding ecoregion (Supplementary Figure S1 f). Panel f shows the geographic coordinates of the trajectories for this precipitation event where interactions with the surface of MBL occurred. These coordinates were then mapped to the ecoregions outlined in Figure 1. Trajectory and ecoregion interactions are listed in Supplementary Dataset S1.

Cloud top heights were estimated using the Equilibrium Levels (EL) and Maximum Parcel Levels (MPL), in addition to NCDC's Level III echo top data, which estimates cloud height based on recorded pressure and temperature levels. Herein, stratiform precipitation is defined specifically as precipitation that was collected from stratus and nimbostratus-like cloud systems independent of trailing stratiform regions from convective storm formations (10).

Once cloud formation type was determined, six unique altitudes were chosen to be analyzed for 120-168 h backward trajectory analysis based on the CCL or LCL of the precipitation event (Supplementary Figure S1a-d). Trajectories were analyzed for previous interactions with the surface and/or mixed boundary layer (MBL) (Supplementary Figure S1e). This was accomplished by downloading the "tdump.csv" files produced by HYSPLIT, which detailed the recorded height above ground level of the trajectory being analyzed, in addition to the height of the MBL.

1.6 Statistical Analyses

The statistical procedures (Exploratory Factor Analysis, Multiple Imputation, Analysis of Variance, Multivariate Analysis of Variance, Mann-Whitney U-Test, Welch's Tests, Kruskall-Wallis Test, Pearson's and Spearman's Correlations, and

Tukey's Honest Significant Difference post-hoc analysis) were performed using SAS software, Version 9.4 of the SAS System for Windows. Graphs and plots were produced using R Software Version 3.2.1 (The R Core Team 2015).

Prior to hypothesis testing with analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA), the raw data were screened for univariate and multivariate normality using The Shapiro-Wilk Test and through visual inspection of Q-Q plots. The assumption of homoscedasticity was verified using Levene's test, univariate outliers were examined based on z-score distributions, multivariate outliers identified using Mahalanobis distance, and linearity/collinearity evaluated through visual inspection of bivariate scatter plots. Log and arcsine transformations were used on distributions that violated assumptions of normality. For distributions not corrected by data transformations, hypothesis tests that assume non-normality or heteroscedasticity were used (i.e., Mann-Whitney U-test, Kruskall-Wallis one-way ANOVA, and Welch's Test). For extreme outliers (values more than three times the interquartile range), raw data values were adjusted according to the method of Tabachnick and Fidell (11).

Multiple imputation was used to provide missing INP concentration data prior to Exploratory Factor Analysis (EFA) and hypothesis testing. The differential INP concentrations were grouped and summed based on the results of the EFA. For example, the INP concentrations used to represent $Bio_{-5\ to\ -10}$ is the summed differential concentrations of INPs active between -5 and $-10^{o}C$.

For hypothesis testing, the dependent variables were always the summed differential INP concentrations for each INP category determined by EFA, which were continuous variables. When the independent variables were continuous (Cell abundance,

pH, conductivity, major ion concentrations, DOC concentrations, PARAFAC Components C1-C3 intensities, OTU sequence reads), Pearson's R and Spearman's Rank correlational analyses were used. Note that for correlations between INP concentrations and bacterial taxa, the number of OTU sequence reads was used for the analysis. When the independent variables were categorical (ecoregion classification, cloud type, season, and precipitation type), ANOVA and MANOVA were used. For post-hoc analysis of ANOVA and MANOVA, Tukey's Honest Significant Difference (HSD) analysis was used. Tukey's HSD test is a post-hoc analysis that compares the means of each group to find significant differences between groups (11).

References

1. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6:1621–1624.

2. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Horn DJV, Weber CF. 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Appl Environ Microbiol 75:7537–7541.

3. Jr WEA, Tucker SA, Fetzer JC. 1991. Fluorescence Emission Properties of Polycyclic Aromatic Compounds in Review. Polycyclic Aromatic Compounds 2:75–105.

4. Tucker SA, Acree WE. 1992. Excitation Versus Emission Spectra as a Means to Examine Selective Fluorescence Quenching Agents. Appl Spectrosc, AS 46:1388–1392.

5. R. Murphy K, A. Stedmon C, Graeber D, Bro R. 2013. Fluorescence spectroscopy and multi-way techniques. PARAFAC. Analytical Methods 5:6557–6566.

6. Anagnostou EN. 2004. A convective/stratiform precipitation classification algorithm for volume scanning weather radar observations. Meteorological Applications 11:291–300.

7. Biggerstaff MI, Listemaa SA. 2000. An Improved Scheme for Convective/Stratiform Echo Classification Using Radar Reflectivity. J Appl Meteor 39:2129–2150.

8. Schultz P. 1989. Relationships of Several Stability Indices to Convective Weather Events in Northeast Colorado. Wea Forecasting 4:73–80.

9. Blanchard DO. 1998. Assessing the Vertical Distribution of Convective Available Potential Energy. Wea Forecasting 13:870–877.

10. Houze Jr. RA. 2014. Cloud Dynamics. Academic Press.

11. Tabachnick BG, Fidell LS, Ullman JB. 2019. Using multivariate statisticsSeventh edition. Pearson, Boston.

Supplementary Table S1. Results of Multiple Imputation for missing INP data. Missing data column lists the variables which did not contain an observation. All missing data followed a monotone missing data pattern.

Missing data	Sample Size	Missing Data Pattern Frequency[a]	Number of Imputations Performed	Relative Efficiency[b]	Pr > |t|[c]
total INPs					
≤−13°C	4	6.56%	5	0.98	<0.0001
≤−14°C	8	13.11%	5	0.92	<0.0001
−15°C	3	4.92%	5	0.92	<0.0001
biological INPs					
≤−13°C	6	9.84%	5	0.95	<0.0001
≤−14°C	6	9.84%	5	0.93	<0.0001
−15°C	4	6.56%	5	0.89	<0.0001
bacterial INPs					
≤−13°C	3	13.46%	5	0.98	<0.0001
≤−14°C	7	13.46%	5	0.92	<0.0001
−15°C	7	5.77%	5	0.95	<0.0001

[a]The percent of observations containing the specified missing data

[b]Measure of how well the imputation calculations converged, as described in Li et al., JAMA 314:1966–1967, 2015.

[c]P-value for t-test of H_0: mean=0

Supplementary Table S2. MANOVA results of INP concentrations, interactions of air masses, and ecoregions. PM, Pacific Maritime; NAM, North Atlantic Maritime; SAM, South Atlantic Maritime; NFM, Northwest Forested Mountains; DSAH, Desert and Semi-Arid Highlands; HNL, High Northern Latitudes; GP, Great Plains; EWW, Eastern Woodlands and Wetlands; EA, East Asia.

INP class	**Ecoregions of significance[a]**		**ANOVA test results**
	Highest INP concentrations[a]	*Lowest INP concentrations*[a]	Prob > F
$total_{-5\ to\ -11}$	EA, HNL	NAM, SAM	P < 0.0001
$total_{-11\ to\ -14}$	EA	NAM, SAM, EWW, DSAH	P < 0.001
$bio_{-5\ to\ -10}$	EA, HNL	NAM, SAM	P < 0.0001
$bio_{-13\ to\ -14}$	EA	NAM, SAM, EWW, DSAH	P < 0.0001
$bio_{-11\ to\ -12}$	NFM	NAM	P < 0.05
$bac_{-5\ to\ -10}$	EA, HNL	SAM, EWW, DSAH	P < 0.001

[a] Based on significant differences in INP concentrations between ecoregions determined via Tukey-Kramer HSD Connecting Letters Report (MANOVA post-hoc analysis) at a confidence level of alpha=0.05

Supplementary Table S3. Characteristics of fluorescent dissolved organic matter PARAFAC components in precipitation from air masses interacting with distinct ecoregions. Ecoregions are based on Level 1 Ecoregions defined by the EPA and CEC. The PARAFAC component means are shown as Raman Units. Numbers following ecoregion name correspond to numbers listed in the ecoregion column of Supplementary Dataset S1.

Ecoregion	Avg. Fluorescence Intensity Maximum
Pacific Maritime (1)	C1: 0.0035 C2: 0.0072 C3: 0.0017
North Atlantic Maritime (3)	N/A
South Atlantic Maritime (4)	C1: 0.0036 C2: 0.0062 C3: 0.0040
Northwest Forested Mountains (5)	C1: 0.0041 C2: 0.0065 C3: 0.0035
Desert and Semi-Aric Highlands (6)	C1: 0.0043 C2: 0.0066 C3: 0.0037
High Northern Latitudes (7)	C1: 0.0071 C2: 0.0083 C3: 0.0032
Great Plains (8)	C1: 0.0035 C2: 0.0053 C3: 0.0015
Eastern Woodlands and Wetlands (9)	C1: 0.0041 C2: 0.0062 C3: 0.0036
East Asia (11)	C1: 0.0071 C2: 0.0083 C3: 0.0032

335 **Supplementary Table S4.** Results of multivariate analysis of variance (MANOVA) for
336 all INP (total, biological, and bacterial) concentrations as a function of season, cloud
337 type, and precipitation type.

Meteorological Parameter	**MANOVA Test Results**
Season	F(15, 122) = 3.12 **p = 0.0003**
Cloud Type	F(5, 46) = 2.47 **p=0.0462**
Precipitation Type	F(5, 46) = 2.45 **p=0.0472**
Bolded *p*-values indicate statistically significant differences between air masses tested at alpha=0.05 level Whole Model, F Test, Prob>F	

338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353

Supplementary Table S5. Correlations between ice nucleating particle (INP) factors and local meteorological conditions. Pearson correlation coefficients (*r*) calculated between INP factors and locally recorded meteorological data. Relative humidity (RH%). Significance levels of Pearson correlation coefficients: **p* < .05, ***p* < .01, ****p* < .001. Cloud top temperature, surface temperature, surface wind speed N=61; Relative humidity, rain amount N=60.

INP Factor	Cloud top temperature (ºC)	Surface Temperature (ºC)	RH %	Rain Amount (mm h^{-})	Surface wind speed (mph)
Total					
$total_{-5\ to\ -11}$	.07	-.53***	-.24	.06	.45***
$total_{-11\ to\ -14}$	.01	.30*	-.08	-.18	-.17
Biological					
$bio_{-5\ to\ -10}$	.06	-.53***	-.24	.05	.45***
$bio_{-13\ to\ -14}$	-.05	-.01	-.03	-.07	-.12
$bio_{-11\ to\ -12}$	.18	-0.29*	-.005	-.14	.04
Bacterial					
$bac_{-5\ to\ -10}$	-.01	-0.51***	-.30*	-.12	.32*

Supplementary Table S6. Significant Spearman's rank correlation coefficients (for which rho $\rho \geq 0.40$; and significance $p<0.05$) between ice nucleating particle (INP) factors and taxon abundance. Significance levels of Spearman's rank correlation coefficients: $*p < .05$, $**p < .01$, $***p < .001$. Only taxa with relative abundance >0.1% for total number of sequence reads across all precipitation events were analyzed. Total number of sequence reads across all precipitation events are listed in last column.

	total$_{-5 \text{ to } -11}$	**total**$_{-11 \text{ to } -14}$	**bio**$_{-5 \text{ to } -10}$	**bio**$_{-11 \text{ to } -12}$	**bio**$_{-13 \text{ to } -14}$	**bac**$_{-5 \text{ to } -10}$	**No. Sequence Reads**
Taxon (Order; Family; Genus)	Spearman's rho (ρ)	Spearman's rho (ρ)	Spearman's rho (ρ)	Spearman's rho (ρ)	Spearman's rho (ρ)	Spearman's rho (ρ)	
Acidobacteria							
Acidobacteriales; Acidobacteriaceae; Candidatus Chloracidobacterium	--	--	0.41*	--	--	--	5674
Bacteroidetes							
Bacteroidales; Rikenellaceae; N/A	0.62***	0.43*	0.59***	--	0.55***	0.54**	81
Sphingobacteriales; N/A; N/A	0.51**	--	0.53**	--	--	--	67269
Cytophagales; Cyclobacteriaceae; N/A	--	0.40*	--	--	--	--	514
Cytophagales; Cyclobacteriaceae; Algoriphagus	--	--	--	0.45*	--	--	69
Cytophagales; Cytophagaceae; N/A	0.51**	--	0.55***	--	--	0.43*	28658
Cytophagales; Hymenobacteraceae; Hymenobacter	0.56***	0.42*	0.59***	--	--	0.48**	19161
Cytophagales; Cytophagaceae; Spirosoma	--	--	0.41*	--	--	--	4486
Chitinophagales; Chitinophagaceae; Segetibacter	0.55***	0.45*	0.53**	--	0.44*	0.55**	294
Sphingobacteriales; env.OPS_17; N/A	0.52**	0.45*	0.44*	--	0.52**	0.46*	672
Cytophagales; Cytophagaceae; Flexibacter	0.42*	0.53**	0.44*	0.48**	--	--	277
Sphingobacteriales; Sphingobacteriaceae; N/A	0.41*	--	0.40*	--	--	--	20330

	total$_{-5\ to\ -11}$	**total**$_{-11\ to\ -14}$	**bio**$_{-5\ to\ -10}$	**bio**$_{-11\ to\ -12}$	**bio**$_{-13\ to\ -14}$	**bac**$_{-5\ to\ -10}$	**No. Sequence Reads**
Candidate Division TM6	--	--	--	--	0.42*	--	81
Chlorobi							
Chlorobiales; N/A; N/A	0.42*	--	--	--	--	--	144
Cyanobacteria	--	--	0.40*	--	--	--	707954
Firmicutes							
Bacillales; Planococcaceae; Planococcus	--	0.53**	--	--	--	--	888
Bacillales; Staphylococcaceae; Macrococcus	--	0.57***	--	--	0.41*	--	271
Clostridiales; Lachnospiraceae; Blautia	0.43*	0.56***	0.43*	--	--	--	75
Clostridiales; Lachnospiraceae; Roseburia	--	0.41*	--	--	--	--	53
Erysipelotrichales; Erysipelotrichaceae; Turicibacter	--	0.47**	--	--	--	--	679
Lactobacillales; Carnobacteriaceae; N/A	0.47**	0.49**	0.46*	--	--	0.40*	642
Lactobacillales; Carnobacteriaceae; Carnobacterium	0.41*	0.48**	0.43*	--	--	--	157
Lactobacillales; Carnobacteriaceae; Desemzia	--	--	0.40*	--	--	--	91
Lactobacillales; Lactobacillaceae; N/A	--	0.42*	--	--	--	--	16067
Lactobacillales; Leuconostocaceae; Leuconostoc	--	--	--	--	0.47**	--	457
Planctomycetes							
Planctomycetales; Planctomycetaceae; Planctomyces	--	0.47**	--	--	--	--	105
Proteobacteria							
Campylobacterales; Campylobacteraceae; Arcobacter	--	0.45*	--	0.43*	--	--	300
Bacteriovoracales; Bacteriovoraceae; N/A	--	0.45*	--	--	--	--	770
Bacteriovoracales; Bacteriovoraceae; Peredibacter	--	0.47**	--	--	--	--	730

	total$_{-5\ to\ -11}$	**total**$_{-11\ to\ -14}$	**bio**$_{-5\ to\ -10}$	**bio**$_{-11\ to\ -12}$	**bio**$_{-13\ to\ -14}$	**bac**$_{-5\ to\ -10}$	**No. Sequence Reads**
Rhizobiales; Methylocystaceae; Methylosinus	--	--	--	0.45*	--	--	523
Sphingomonadales; Erythrobactereaceae; N/A	--	--	--	--	0.41*	--	1190
Oceanospirillales; Oceanospirillaceae; N/A	--	0.44*	--	--	0.42*	--	66
Pseudomonadales; Moraxellaceae; Perlucidibaca	--	0.41*	--	--	--	--	2013
Burkholderiales; Comamonadaceae; Polaromonas	--	--				0.42*	108
Chromatiales; Chromatiaceae; Rheinheimera	--	0.50*	--	--	--	--	912
Rhodocyclales; Rhodocyclaceae; N/A	0.47**	--	--	--	0.52**	0.48*	8105
Rhodospirillales; wr0007; N/A	--	--	--	0.40*	--		1094
Xanthomonadales; Xanthomonadaceae; N/A	0.47**	0.44*	0.49**	--	--	--	14094
Spirochaetes							
Spirochaetales; N/A; N/A	--	0.41*	--	--	0.43*	--	93
Verrucomicrobia							
Chthoniobacterales; Chthoniobacteraceae; Chthoniobacter	--	--	--	--	0.42*	--	510
Unclassified							
Unclassified; OTU20	--	--	--	--	--	0.40*	425
Unclassified; OTU32	--	--	--	--	--	0.42*	175
Unclassified; OTU43	--	--	--	--	0.43*	--	1020
Unclassified; OTU51	--	--	--	0.43*	--	--	1051
Unclassified; OTU73	0.52**	0.45*	0.44*	--	0.52**	0.46*	351
Unclassified; OTU74	--	--	--	0.42*	--	--	672
Unclassified; OTU88	--	--	--	0.44*	--	--	254
Unclassified; OTU108	--	--	--	--	0.46*	--	120
Unclassified; OTU13	0.47*	--	0.52**	--	--	--	232
Unclassified; OTU18	--	--	--	--	0.43*	--	146
Unclassified; OTU22	0.44*	--	0.51**	--	--	--	110

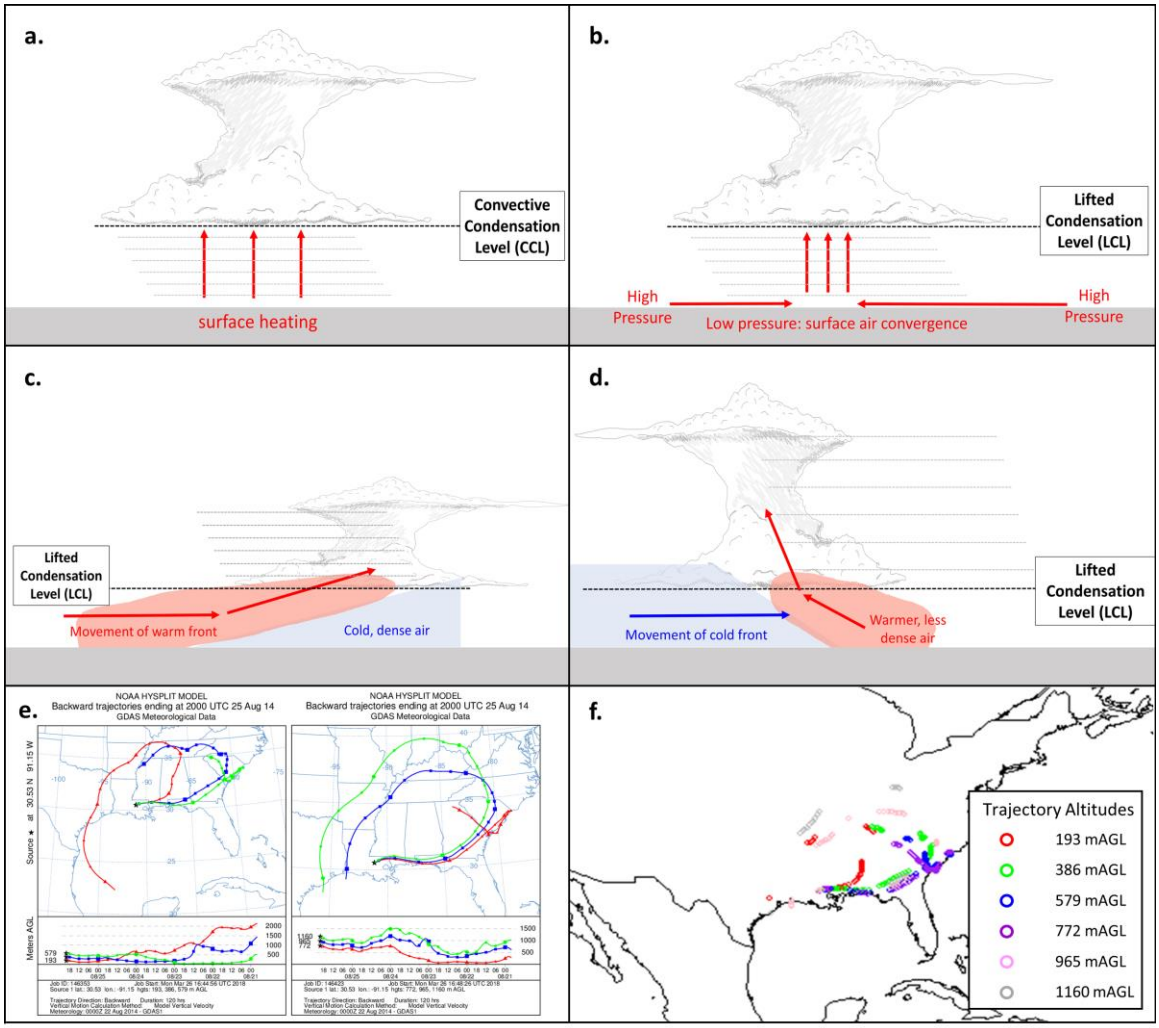
391
392
393
394

Supplementary Table S7. Taxa with significantly different abundances based on cloud type and season. Table lists the test performed (Mann-Whitney U-Test (MT), Welch's T-Test (WT), Kruskall-Wallis ANOVA (KA), and Welch's ANOVA (WA)) and corresponding *p*-value. Only significant *p*-values ($p<0.05$) are shown. The taxa analyzed are listed in Supplementary Table S6.

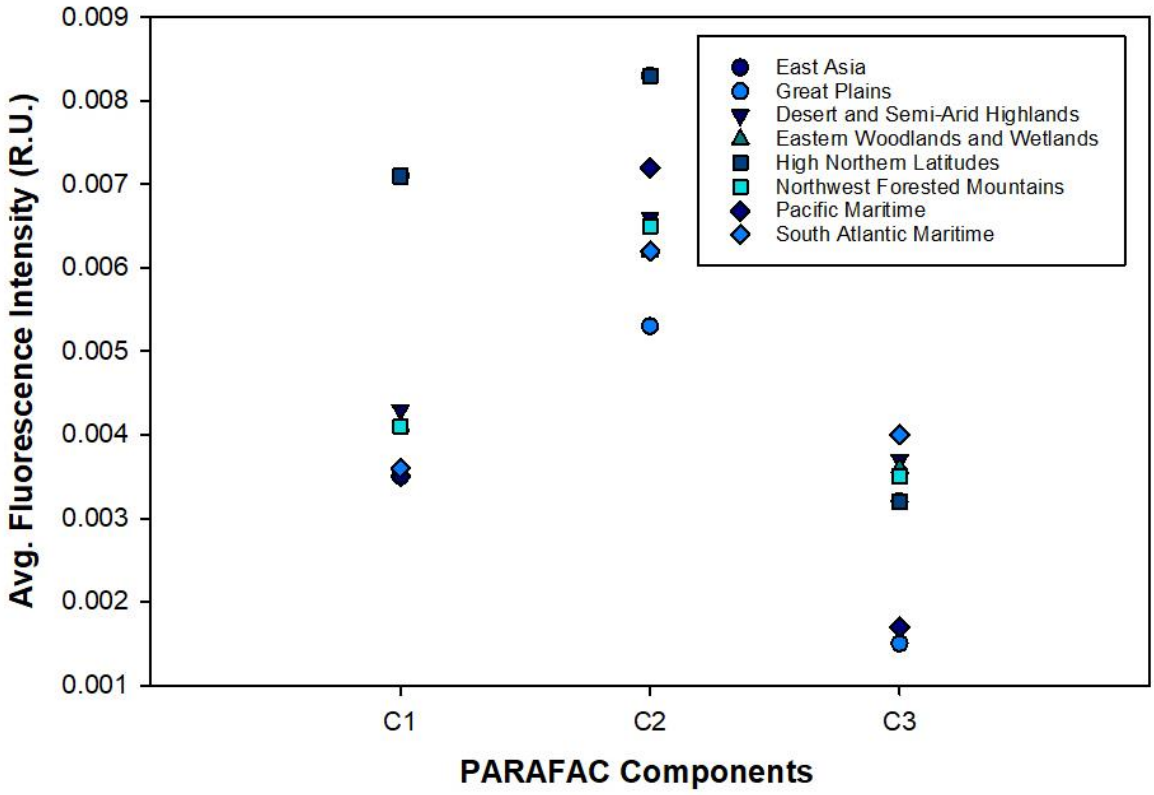
Taxon (Order; Family; Genus)	Cloud Type	Season
Bacteroidetes		
Cytophagales; Cytophagaceae; N/A	WT 0.0291	WA 0.0112
Cytophagales; Cytophagaceae; Flexibacter	WT 0.0469	--
Cytophagales; Hymenobacteraceae; Hymenobacter	WT 0.0231	WA 0.0108
Bacteroidales; Rikenellaceae; N/A	WT 0.0384	--
Sphingobacteriales; Sphingobacteriaceae; N/A	MT 0.0393	WA 0.0209
Sphingobacteriales; N/A; N/A	MT 0.0056	WA 0.0029
Firmicutes		
Clostridiales; Lachnospiraceae; Blautia	MT 0.0005	--
Lactobacillales; Carnobacteriaceae; N/A	WT 0.0233	WA 0.0400
Lactobacillales; Carnobacteriaceae; Carnobacterium	WT 0.0003	--
Planctomycetes		
Planctomycetales; Planctomycetaceae; Planctomyces	--	WA 0.0353
Proteobacteria		
Campylobacterales; Campylobacteraceae; Arcobacter	MT 0.0031	--
Rhodocyclales; Rhodocyclaceae; N/A	--	WA 0.0111
Verrumicrobia		
Chthoniobacterales; Chthoniobacteraceae; Chthoniobacter	MT 0.0118	WA 0.0326
Unclassified		
Unclassified; OTU20	MT 0.0088	KA 0.0007
Unclassified; OTU32	MT 0.0432	WA 0.0299
Unclassified; OTU74	MT 0.0344	--
Unclassified; OTU13	WT 0.0278	WA 0.0049
Unclassified; OTU18	WT 0.0391	--
Unclassified; OTU22	MT 0.0187	--

Supplementary Table S8. Spearman correlations between ice nucleating particle (INP) factors. Correlations calculated for differential concentrations of INPs between factor groupings. Significance levels of Pearson correlation coefficients (top number in each cell) and Spearman's rho (bottom number in each cell): *p < .05, **p < .01, ***p < .001.

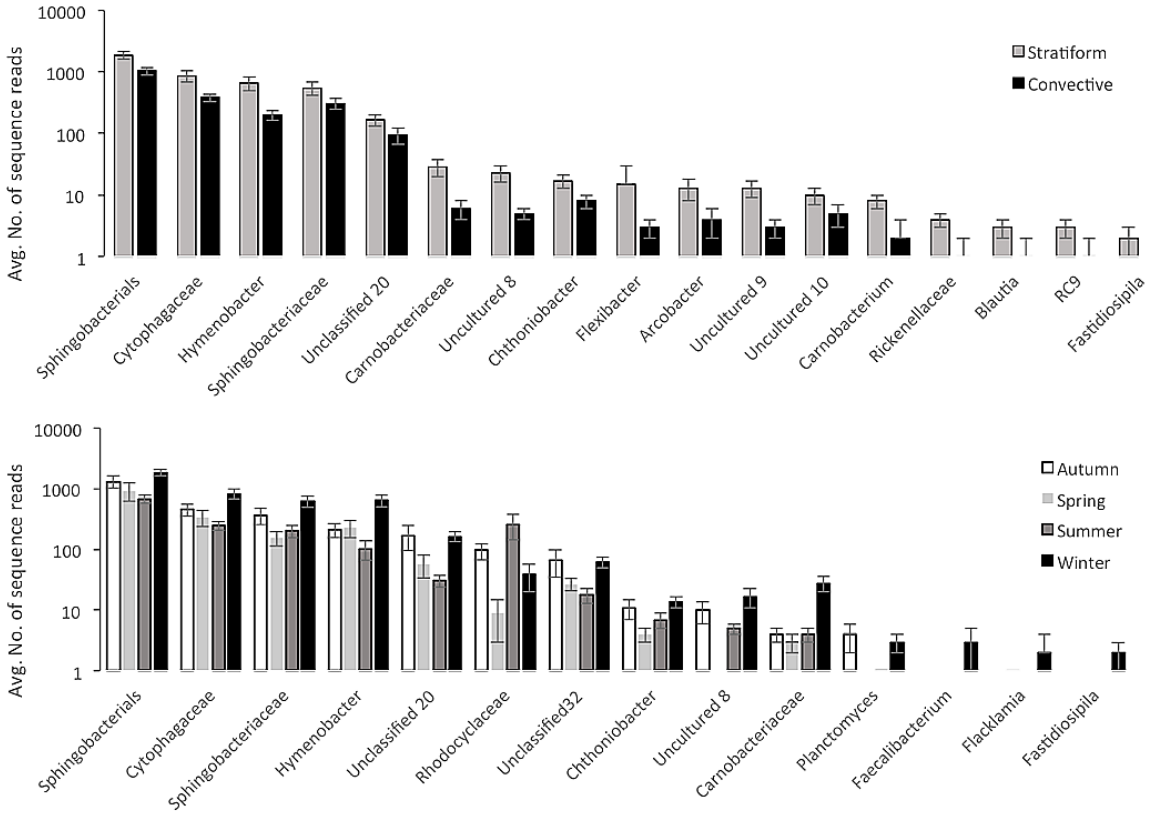
	total$_{-5 \text{ to } -11}$	**total**$_{-11 \text{ to } -14}$	**bio**$_{-5 \text{ to } -10}$	**bio**$_{-13 \text{ to } -14}$	**bio**$_{-11 \text{ to } -12}$	**bac**$_{-5 \text{ to } -10}$
total$_{-5 \text{ to } -11}$	--	--	--	--	--	--
total$_{-11 \text{ to } -14}$	.08 .03	--	--	--	--	--
bio$_{-5 \text{ to } -10}$	.96*** .94***	.20 .16	--	--	--	--
bio$_{-13 \text{ to } -14}$	-.01 -.07	.47** .62***	.00 .01	--	--	--
bio$_{-11 \text{ to } -12}$	.04 -.02	.53** .55***	.00 -.05	.00 .06	--	--
bac$_{-5 \text{ to } -10}$	.83*** .83***	.26 .32*	.84*** .83***	.06 .07	.10 .15	--



Supplementary Figure S1. Cloud formation mechanisms and HYSPLIT trajectory analysis. The mechanism of lifting for each precipitation event is important in determining the altitudes for backward trajectory analysis. There are five general mechanisms of large-scale air movement that lead to cloud formation, which are dealt with as described in the supplemental methods. (a) Convection (b) Convergence (c) Warm front lifting (d) Cold front lifting. Orographic lifting was not observed in this study and is not depicted. Panels e and f show an example of how trajectories were analyzed, and are described in detail in the supplemental methods. Trajectory and ecoregion interactions are listed in Supplementary Dataset S1.



Supplementary Figure S2. PARAFAC Components fluorescence intensity profiles based on The North American Ecoregion classifications used in this study. Average Fluorescence Intensity was calculated based on ecoregion and is plotted on the y-axis in Raman Units (R.U.). PARAFAC Components C1-C3 are plotted as categories on the x-axis.



Supplementary Figure S3. Significant differences in DNA operational taxonomic unit (OTU) abundances as a function of cloud type and season. The mean number of sequence reads for each OTU is plotted, with bars indicating the standard error of the mean. Each taxon is represented by a single unique OTU. Top: OTUs that correlated with ice nucleating particle (INP) concentrations and had significantly different abundances in precipitation from stratiform (N=10) and convective (N=35) cloud formations. Bottom: OTUs that correlated with INP concentrations and had significantly different abundances based on season (Autumn, N=12; Spring, N=6; Summer, N=14; Winter, N=13).