

Deciphering the Source of Primary Biological Aerosol Particles: A Pollen Case Study

Albert Rivas-Ubach,^{*} Bryan Stanfill, Swarup China, Ljiljana Paša-Tolić, Alex Guenther, and Allison L. Steiner



Cite This: *ACS Earth Space Chem.* 2021, 5, 969–979



Read Online

ACCESS |



Metrics & More



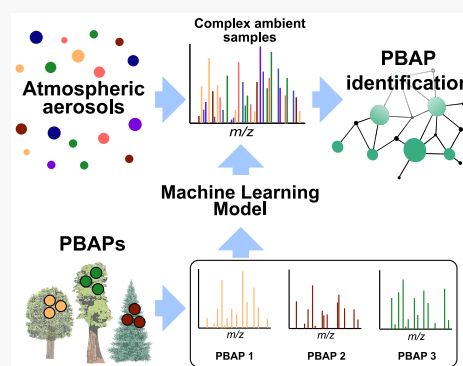
Article Recommendations



Supporting Information

ABSTRACT: Primary biological aerosol particles (PBAPs) are microscopic solids suspended in the atmosphere emitted by biological systems and play critical roles in the atmosphere and the atmosphere–biosphere system, impacting human health, climate, and the ecosystem function. Understanding the sources of PBAPs is necessary to decipher the mechanistic interactions between aerosols, climate, and other ecosystem components. However, the detection of specific PBAPs in complex ambient aerosol samples is challenging. We performed metabolomics analyses of pollen from three pollinating tree species and ambient samples collected during the peak pollination period of each species. Random Forest and sPLS-DA machine learning methods were employed to evaluate whether metabolic signatures of ambient samples can reveal the source of the main pollen particles present in the atmosphere. Our results suggest that atmospheric ecometabolomics techniques combined with sophisticated statistical methods can decipher the origin of abundant PBAPs from complex ambient samples. Developing complete libraries containing high-resolution metabolomic fingerprints of the major PBAPs present in the atmosphere would significantly advance future research to accurately understand the role of PBAPs in the atmosphere, ecosystems, and human health.

KEYWORDS: *atmo-ecometabolomics, metabolomics, bioaerosols, artificial intelligence, machine learning*



1. INTRODUCTION

Primary biological aerosol particles (PBAPs) are solid airborne particles of biological origin.¹ PBAPs are 0.01–1000 μm in size and can include pollen, fungal spores, protozoa, bacteria, algae, and biological debris emitted to the atmosphere. PBAPs are a key component of the atmosphere and can significantly impact human health, climate, and the ecosystem function.² PBAPs emitted from the terrestrial biosphere, especially pollen and fungal spores, have commonly received substantial attention within the human health community as they have been associated with important asthma and allergic rhinitis outbreaks.^{2–5} Despite their importance for health, the fate and impacts of PBAPs on the atmospheric and ecosystem function are still poorly understood.

The size of PBAPs is relatively large compared to other atmospheric particles [pollen grains (15–200 μm), fungal spores (2–4 μm)] in the atmosphere; however, studies have shown that PBAPs can rupture under high-humidity conditions resulting in the release of submicron fragments (<1 μm)^{6–11} to the atmosphere¹² and even reach the upper troposphere.¹³ Still, those fragments have not been accounted for in regional and global atmospheric models¹⁰ even though they have proven to act as cloud condensation nuclei⁹ and are also effective in nucleating ice.^{13,14} Anemophilous pollen accounts for a significant amount of biological material in the

atmosphere, especially during pollination seasons. Pollen emissions are seasonal and vary according to the distribution of different plant species, latitude, wind, frost-free days, and precipitation.¹⁵ In ecosystems with contrasting anemophilous pollination seasons (i.e., temperate and boreal areas), pollen concentrations have been found in the range between 0.04 and 0.8 $\mu\text{g m}^{-3}$ with a maximum concentration of pollen submicron particles (subpollen particles) during rainy events.¹² Interestingly, different pollen species, and their corresponding subpollen particles, have been demonstrated to significantly differ in their ice nucleation efficiency,¹³ having thus the potential to impact climate to different extents. Therefore, there is growing interest in pollen grains and subpollen particles, given their potential impacts on precipitation and the hydrological cycle.^{1,13,16} In addition, submicron particles can potentially impact the total radiation budget of the Earth as can they scatter and absorb sunlight. Although the concentration of

Special Issue: Chemical Interactions in the Plant–Atmosphere–Soil System

Received: October 27, 2020

Revised: February 24, 2021

Accepted: March 22, 2021

Published: March 30, 2021



ACS Publications

© 2021 American Chemical Society

969

<https://doi.org/10.1021/acsearthspacechem.0c00295>
ACS Earth Space Chem. 2021, 5, 969–979

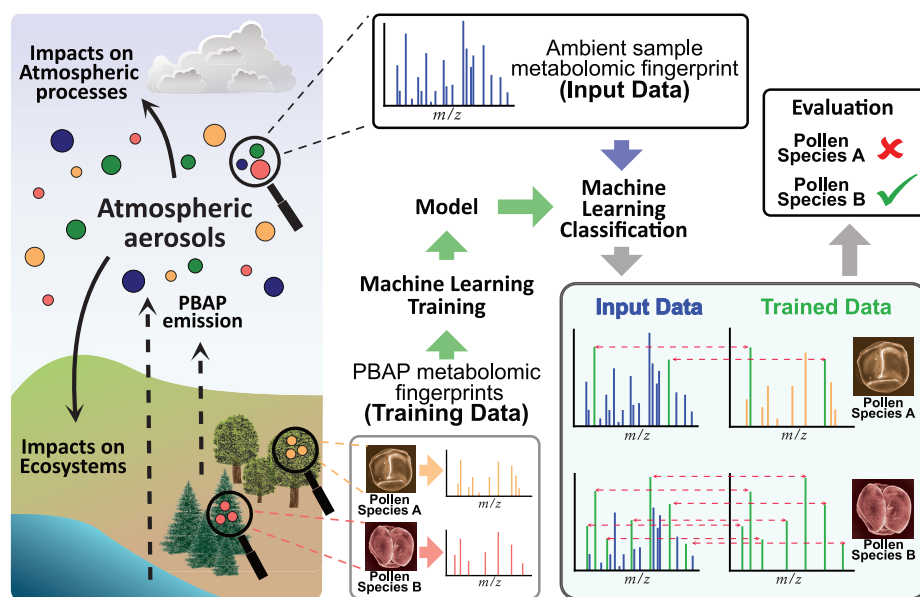


Figure 1. Illustration of the use of artificial intelligence methods (i.e., machine learning) on mass-spectrometry-based metabolomic fingerprints from aerosol samples to identify the source of primary biological aerosol particles (PBAPs).

subpollen particles in the atmosphere remains unknown, it has been suggested that they could be responsible for a reduction of precipitation events in clean continental environmental conditions and create a negative feedback to the subpollen particle production, thus affecting the total organic aerosol mass loading in the atmosphere.¹⁷

Deposition of aerosols can significantly impact terrestrial and aquatic environments.^{18–20} Pollen and other PBAPs can represent a rapid supply of nutrients [e.g., carbon (C), nitrogen (N), phosphorus (P), sulfur (S), etc.] to different components of ecosystems such as phytoplankton,^{18,21,22} plants,^{20,23–25} and the phyllosphere.^{26,27} Although there is no accurate data for terrestrial ecosystems yet, increases of 1.5 and 0.4 mg L^{−1} in soluble N and P, respectively, have been reported in lakes receiving pollen from surrounding *Pinus sylvestris* forests.²⁸ Significant shifts on the amount and types of PBAP deposition could lead to fluctuations of the ecosystem function by shifting the environmental C:N:P:S stoichiometry and nutrient cycling.^{29–31}

Given the potential impacts of PBAPs on atmospheric processes and ecosystem function, identifying the source of PBAPs from aerosol mixtures is of special importance to improve and increase accuracy of climate, environmental, and biogeochemical models. However, identifying the sources of PBAPs in the atmosphere remains a challenge, due to high PBAP diversity and the spatial and temporal complexity. Ambient organic aerosol samples can include a mixture of PBAPs, secondary organic aerosols (SOAs), and other organic compounds of anthropogenic origin.^{1,32,33}

DNA metabarcoding performs simultaneous identification of various taxa present in the same environmental sample by amplification of DNA and sequencing. Applying high-throughput metabarcoding techniques on complex ambient samples can provide good data on detection and quantification of certain PBAPs, but there are still challenges, especially in sample preparation and analysis. On one hand, the relatively long residence times of large particles in the atmosphere can lead to damage of DNA which can make amplification, sequencing, and taxa identification difficult. Small contami-

nants can also be amplified during polymerase chain reaction (PCR) methods, providing confusing results. On the other hand, metabarcoding methods cannot cope with smaller subparticles in suspension (i.e., subpollen) as they do not contain DNA, or the DNA has been substantially damaged. In addition, in terms of throughput, extracting DNA of tiny amounts of sample followed by polymerase chain reaction (PCR) methods and sequencing can still be complex and time-consuming. Fluorescence-based and spectroscopy techniques, including the wide-band integrated bioaerosol spectrometer (WIBS),³⁴ have been typically used for real-time detection of biological particles.

Different microscopy, spectroscopy, and mass spectrometry (MS) techniques have been used for detection of biological particles.³⁴ MS-based methods have been widely utilized in atmospheric research to detect specific molecular compounds (biomarkers) to track PBAPs in a rapid way. MS is currently one of the most sensitive analytical techniques available for tracing specific compounds, and only few molecules are necessary to be detected.³⁵ For example, glucose, fructose, and sucrose have been used as tracers for pollen grains^{12,36} as they represent a large proportion of their biomass.^{37,38} Nevertheless, sugars and other compounds are expected to be also found in other PBAPs such as fungi spores or plant debris^{36,39} making it difficult to directly identify pollen particles in the atmosphere using specific chemical tracers alone. We thus hypothesize that using a set of tracers at once instead of single compounds can increase the identification accuracy of PBAPs even at the species level. Novel chemical characterization methods, such as MS-based metabolomics, can produce a large number of markers (metabolomic features) that can be used to identify more precisely low-concentration PBAPs present in the atmosphere. Prior work has shown that the high sensitivity of MS instruments can detect the overall changes of organisms' metabolomes under biotic and nonbiotic stresses.^{40–42} In addition, MS-based metabolomic analyses have been proven to differentiate metabolic signatures between plant genotypes of the same species,^{43,44} between tree subspecies coexisting in the same environment,⁴⁵ and between insects of the same species

with different age and sex⁴⁶ or even feeding on distinct tree varieties.⁴⁷ In addition, MS-based metabolomics techniques have been also applied in aerosol research, i.e., atmospheric metabolomics,²⁶ and proven to efficiently differentiate chemical fingerprints of ambient samples collected in distinct seasons. Therefore, entire metabolic fingerprints, i.e., a collection of multiple metabolic features detected from samples, can act as unique complex signatures for individual samples that, together with recent bioinformatic methods (i.e., artificial intelligence), can serve to improve tracking of PBAPs. We postulate that applying advanced biostatistical methods of classification on metabolomics data can be a next-generation approach for PBAP identification at greater confidence (Figure 1).

This study represents a first attempt to test whether high-resolution MS-based metabolomic fingerprints coupled with machine learning approaches can trace specific PBAPs in the atmosphere. We performed metabolomics analyses of pollen from three different anemophilous pollinating tree species and from ambient samples collected during the pollination peak periods of each of those tree species. The generated metabolomic fingerprints from pollen samples were subsequently used to train different machine learning algorithms. Machine learning models were used to evaluate whether the signatures of each pollen species fingerprints are correctly found in the ambient samples collected during their corresponding pollination season. This study represents a first step for establishing novel methods of atmospheric aerosol identification using complex chemical data sets. Finally, we discuss specific limitations of this novel method and describe the main future perspectives on future applications of this technique.

2. MATERIAL AND METHODS

2.1. Study Site and Sampling. Samples were collected during three different tree pollination peak periods occurring in spring 2017 at the University of Michigan (UMICH; Ann Arbor, MI) and its surroundings. UMICH is located in the northeastern corner of Ann Arbor, MI (42° 29' N, 83° 70' W), and is a part of the Detroit metropolitan area with a total population of over 4.2 million. The area surrounding Ann Arbor is a mixture of landscapes represented mainly by agricultural, urban, and extended forested areas predominated by distinct tree taxa including *Pinus sp.*, *Quercus sp.*, *Betula sp.*, *Acer sp.*, *Fagus sp.*, and *Populus sp.* Climate is strongly influenced by the Great Lakes with an average maximum annual temperature of 28 °C and annual precipitation of 950 mm.

Pollen samples were directly collected from tree inflorescences of three common tree species in southeastern Michigan: *Betula papyrifera*. (hereafter birch), *Quercus bicolor* (hereafter oak), and *Pinus strobus* (hereafter pine) at the beginning of their pollination period in the Ann Arbor area in 2017. Tree inflorescences were harvested under conditions when pollen release was imminent and placed in a cool, dark location to release or “drop” the pollen, typically within 24 h of collection. Imminent release was determined on a day-by-day basis when pollen would be released from the tree with mild disturbance. Six samples were collected from different individuals of each tree species. Pollen samples were subsequently sieved through an 80 μ m sieve to remove impurities. Samples were kept at –80 °C in plastic tubes until lyophilization and compound extraction.

Aerosol samples were collected daily at UMICH campus during the pollination peak of each of the three tree species: birch pollination period (from April 24th to 26th, 2017; 3 days), oak pollination period (May 8th, 10th, 11th, 12th, 17th, 18th, and 22nd, 2017; 7 days), and pine pollination period (from May 30th to June 1st, 2017; 3 days). Environmental conditions during the sampling periods for each pollination peak are described in Table S1.

Aerosol particle samples were collected following an optimized protocol for atmospheric metabolomics analyses.²⁶ Briefly, samples were collected onto precombusted (5 h at 450 °C) high-purity quartz filters (WhatmanQM-A 37 mm, Whatman International Ltd., Maidstone, UK). Samples were collected at a flow rate of 30 L min^{–1} for 24 consecutive hours (from 9 am to 9 am of the next day) achieving a final air volume of ~32 400 L per sample. We collected total atmospheric particles without any size cutoff (see Rivas-Ubach et al.²⁶ for setup details). All ambient samples were stored wrapped in aluminum foil at –80 °C until compound extraction.

2.2. Metabolite Extraction of Pollen. Pollen samples were lyophilized for 48 h before metabolite extraction. Polar and semipolar metabolites from pollen were extracted following well-established protocols,⁴⁸ with minor modifications. Briefly, two set of 2 mL glass vials were labeled: set A to perform the extractions (18 vials, one per pollen sample) and set B to keep the extracts (18 vials, one per pollen extract/sample). Lyophilized pollen (10 mg) of each species was added into the corresponding vial of set A. Each vial received 0.8 mL of methanol/water (80:20) and was subsequently sonicated for 5 min. Vials were thus shaken at 1400 rpm in a Thermomixer instrument (Eppendorf AG, Hamburg, Germany) operating at 4 °C for 1 h, sonicated again for 10 min, and centrifuged at 2700g at 4 °C for 10 min. Supernatants from each sample were collected and added into their corresponding glass vial of set B. Extracts were stored at –80 °C until analyses.

2.3. Metabolite Extraction of Ambient Samples. Ambient samples were extracted following Rivas-Ubach et al.²⁶ Briefly, three different tube sets were labeled, sets of tubes A [8 mL combusted (450 °C for 5 h) glass tubes to perform the extractions], B [15 mL polypropylene tubes to keep the extracts], and C [2 mL glass vials for the concentrated extracts]. Each set was composed of 13 tubes: one per ambient sample. Each ambient sample (filter) was inserted into the corresponding tube of set A and received 5 mL of methanol/water (80:20). All samples were subsequently sonicated for 10 min at room temperature (~24 °C), and extracts (4 mL) were transferred to the tubes of set B. Two extraction procedures were performed to each filter but adding 4 mL of fresh methanol/water (80:20) for the second extraction in the tubes of set A. Second extractions were combined with the initial aliquots in the tubes of set B. Extracts in the tubes of set B were dried by using an ultrapure nitrogen evaporator, and 1 mL of fresh methanol/water (80:20) was added to each tube and vortexed for 30 s to resuspend the dried extracts. Concentrated ambient extracts were thus centrifuged at 4000g for 5 min, and supernatants were transferred to the tubes of set C. Combusted nonused filters were also extracted through all procedures and used later as experimental blanks to determine the background level of the ambient samples. Ambient sample extracts were stored at –80 °C until analyses.

2.4. LC-MS Data Acquisition and Chromatogram Processing. LC-MS analyses were performed with a high-

resolution mass spectrometer (HRMS; Orbitrap Velos) equipped with heated electrospray ionization (HESI) (Thermo Fisher Scientific, Waltham, MA) and coupled to an HPLC (high-pressure liquid chromatographer; Thermo Vanquish, Thermo Fisher Scientific, Waltham, MA). The LC was performed at 30 °C using a C18 reversed-phase column (Hypersil Gold 150 × 2.1 mm, 3 μm particle size). The flow rate was constant at 300 μL min⁻¹. Chromatographic solvents were composed of 0.1% formic acid in water (mobile phase A) and 0.1% formic acid in acetonitrile/water (90:10) (mobile phase B). The injection volume was 5 and 10 μL for pollen and ambient samples, respectively. The chromatographic method started at constant conditions of 90% A (10% B) for 5 min. Then, the elution linearly changed to 10% A (90% B) during the following 15 min (minute 20 of chromatography), and conditions were held for 2 more minutes (minute 22 of chromatography). The initial conditions were recovered over the following 2 min (minute 24 of chromatography). The column was thus washed and stabilized during 11 extra minutes at the initial conditions (90% A; 10% B) (minute 35 of chromatography; end). A total of 10 experimental blanks were injected during the sequence. A mixture of standards was analyzed during the sequence to test for mass accuracy of the instrument and for *m/z* calibration purposes. In order to avoid any carryover after the injection of standards, two consecutive injections of methanol/water (80:20) were analyzed after the standard mixture. The HRMS operated in Fourier transform mass spectrometry (FTMS) and full-scan mode at high resolving power [60 000 full width at half-maximum (fwhm)]. The acquired ions by the HRMS were between 50 and 1000 *m/z* in both negative and positive ionization modes.

LC-MS RAW files were processed and aligned in MZmine 2.38,⁴⁹ and the data set was exported to a CSV file (see Table S2 for detailed processing parameters in MZmine software). Metabolic feature assignment to specific metabolites was performed using an in-home LC-MS library containing over 600 common compounds from the primary and secondary metabolism, especially in plants, and it was based on exact mass and retention time (RT). This feature annotation corresponds to a second level of putative identification as proposed by the chemical analysis working group and the metabolomics standards initiative.⁵⁰ Using both RT and high-resolution MS reduces the number of false-positive feature annotations in low-molecular-weight metabolites (typically <400 Da).⁵¹ However, it is important to note that the main core of this study does not rely on metabolite identification but metabolomic fingerprinting, and using exact mass and RT for compound annotation still corresponds to putative identifications. Metabolite annotation for LC-MS data is detailed in Table S3 and Figure S1.

2.5. Metabolomics Data Set Filtering. The LC-MS data set, including both pollen and ambient samples and both ionization modes (positive and negative), was filtered at the cell level. A cell of a study corresponds to the group of samples defined by the unique combination of the different levels between categorical factors (see Figure S2 for a schematic representation of the “cell” concept). Peak areas of the same identified compounds in separate peaks and/or both ionization modes were summed into a single variable (see Table S3 and Rivas-Ubach et al.⁵¹ for details). Our study is composed of 6 different cells: birch pollen (6 replicates), oak pollen (6 replicates), pine pollen (6 replicates), ambient birch pollination peak (3 samples), ambient oak pollination peak

(7 samples), and ambient pine pollination peak (3 samples). Ambient samples could be considered as pseudoreplicates as they have been collected on different days within the same pollination peak period. Therefore, ambient samples can still present large intracell variability, and the data filter has been optimized to obtain a final data set with the most common information between sampling days within the same pollination peak period and to reduce the number of random single day-unique features. Accordingly, specific features derived from individual days will be removed during the following filtering pipeline, as they are not representative of a pollination peak period. Filtering was performed through 4 main steps:

- (1) *Minimum data.* Those variables (metabolite features) represented in less than 100% of the replicates across all cells for the given variable were removed from the data set. Only those variables with at least one cell containing values for all replicates were kept.
- (2) *Blank threshold.* When the number of experimental blanks with data (>0) for a given variable was equal to or lower than 2 out of 9 blank samples, all values from all experimental blanks for such a variable were considered as zero.
- (3) *Signal to noise (S/N).* Variables with *S/N* < 25 across all individual cells were removed from the data set. We kept only those variables with an average number in one or more cells that had a value at least 25 times the average number of the experimental blanks. Noise level was represented by the data obtained from the 9 experimental blanks.
- (4) *Zero filter.* To avoid spontaneous metabolic features shown in low number of samples of a cell, for each variable, when a cell had <60% of data (1 out of 3 replicates of a cell of ambient samples; 3 out of 6 replicates of a cell of pollen samples), all values for such cell replicate samples were considered zero.

2.6. Statistical Analyses. The filtered data set used in this study was composed of 43 319 metabolomic features (continuous variables); 6756 of these features were unique for pollen grains, and 2736 of these features were unique for ambient samples. From the aligned fingerprints, 259 features were assigned to a molecular compound at a second level of putative identification.⁵⁰

To visualize the overall metabolomic variability between all samples of the study, the entire metabolomic fingerprints of both pollen and ambient samples are represented in a heatmap with hierarchical clustering tree diagrams (dendrograms). Sample dendrograms represent a classification of samples according to their metabolome similarity between samples. Therefore, samples in close proximity in the tree diagram have more similar metabolome structures. Metabolomic features are classified in the variable dendrogram according to their relative abundance across the different samples. Prior to plotting the heatmap, individual variables were scaled to the same magnitude, with 1 as the largest value for each of the variables. Figure S3 provides additional details regarding the interpretation of the heatmap and hierarchical clustering for variables and study subjects. In addition, pollen and ambient samples' metabolomic fingerprints were separately subjected to permutational multivariate analysis of variance (PERMANOVA) using the Euclidian distance to test for overall differences between pollen species and between pollination peak periods,

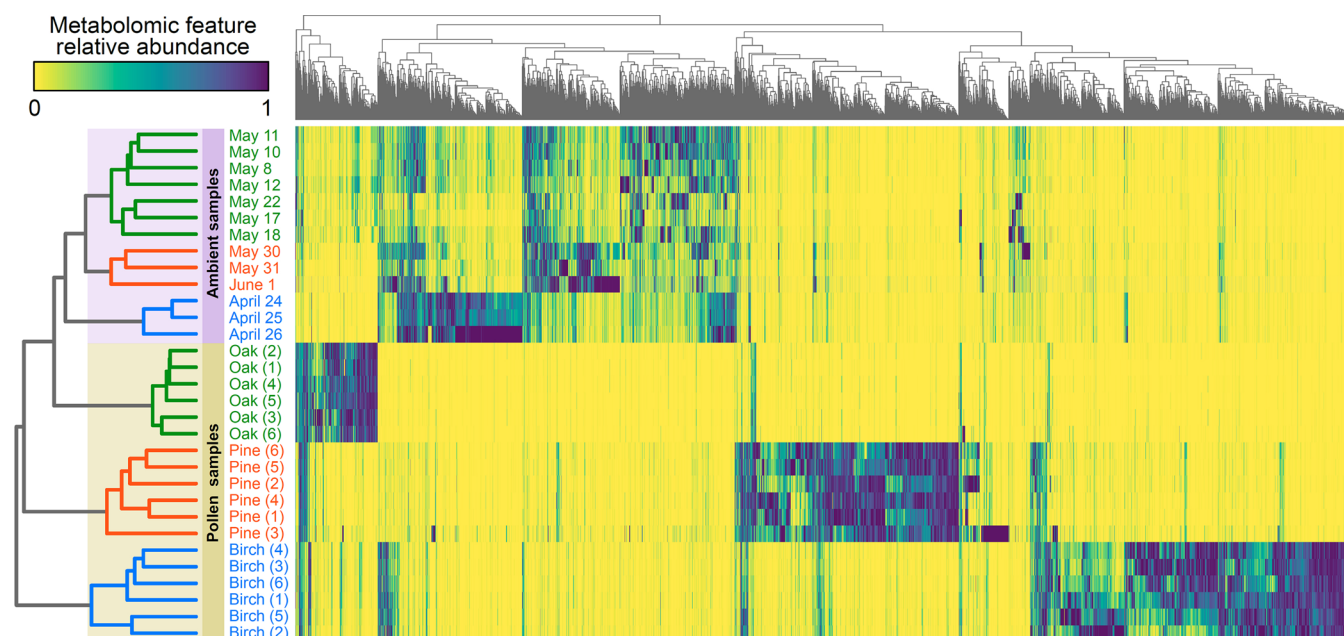


Figure 2. Hierarchical clustering tree diagrams (dendrograms) of samples (pollen and ambient samples; left-side of the heatmap) and metabolomic features (top of the heatmap) computed using all samples (31; 18 pollen and 13 ambient samples) and their metabolomic fingerprints (43 319 aligned features total). The dendrogram of samples illustrates the resemblance of the metabolomic fingerprints between all samples. Dendrograms for metabolomic variables illustrates the resemblance of metabolomic variable variation across all samples. Both dendrograms are integrated into a heatmap representing the relative abundance of all metabolomic features across samples. Both hierarchical clustering analyses and heatmaps were calculated according to the relative abundance of all metabolomic features of the data set. The color gradient of the heatmap represents the relative abundance for each detected metabolic feature (column of the heatmap) across all samples (row of the heatmap), and dark-blue represents the highest relative abundance. Pollen samples are represented in different colors (oak pollen in green, pine pollen in orange, and birch pollen in blue), and branches of dendrogram clusters of samples are represented with the same color as the sample. The replicate number of each sample is indicated between parentheses. Ambient samples are represented by the day of collection and different colors in accordance with the natural pollination peak of each tree species (oak pollination peak in green, pine pollination peak in orange, and birch pollination peak in blue). See the [Material and Methods](#) section and [Figure S3](#) for a more detailed explanation of hierarchical clustering and heatmap representation.

respectively. For this analysis, the permutation number was set at 10 000. PERMANOVA is a nonparametric multivariate statistical test which makes no explicit assumptions about the distribution of dissimilarities or the original variables.

Machine learning methods analyzed four different filtered sets of data. For computational consistency and to avoid redundant metabolic features (i.e., repeated signals found in both ionization modes), the entire metabolomic fingerprints obtained in positive (24 174 features) and negative (19 218 features) ionization modes were analyzed separately. At the same time, this allows testing whether specific ionization modes perform better for pollen species classification with machine learning techniques. For additional testing and to generate new hypotheses and questions, the *amino acid* and *central carbon metabolism* profiles (acquired by both positive and negative ionization modes, and putatively identified by RT and exact mass) were also analyzed (see [Table S4](#) for compounds included within each profile). Amino acids are monomers of proteins more directly linked to the unique genome of a species, and we tested whether amino acid profiles alone could serve for proper classification of pollen species. Central carbon metabolism compounds tend to be shared by most living organisms, and we also tested if shared metabolic signatures decoupled to genomic expression could properly classify the different pollen species.

Two different machine learning methods were utilized to determine whether a model trained on the pollen metabolite data (fingerprints, amino acids, and central carbon metabolism) could identify the same signatures in the ambient aerosol

samples collected during such species' peak pollination time. Machine learning regularized methods were applied on large data sets (negative and positive metabolomic fingerprints) because the variable selection step (sPLS-DA) protects the classifier from overfitting the training data, which likely contains many unnecessary features that generate "noise". For the smaller data sets (amino-acid profile and central carbon metabolism), however, regularization can hurt classifier performance because all the metabolites are needed to correctly classify the ambient samples because there are few or no unnecessary features, and in this case, we used Random Forest models as a more suitable tool for pollen classification. First, validation of the two trained models on pollen data was tested against the same pollen species. For that, out of the 6 replicates of each pollen species, 3 replicates were used to train a model that was subsequently used to classify the other 3 replicates of each of the pollen species. We observed that applied methods classified properly all pollen samples with a misclassification rate of 0% using all data sets [metabolomic fingerprints (positive and negative ionization modes), amino acids, and central carbon metabolism] ([Table S5](#)). We considered a correct classification when the pollination peak period of a specific ambient sample coincided with the highest probability match of its corresponding pollen species. Model validation was also used to choose the number of trees and features used in the Random Forest and the dimensionality/complexity of the reduced feature space in sPLS-DA (see the [Supporting Information](#) for R scripts for both sPLS-DA and Random Forest methods). All models were trained using the

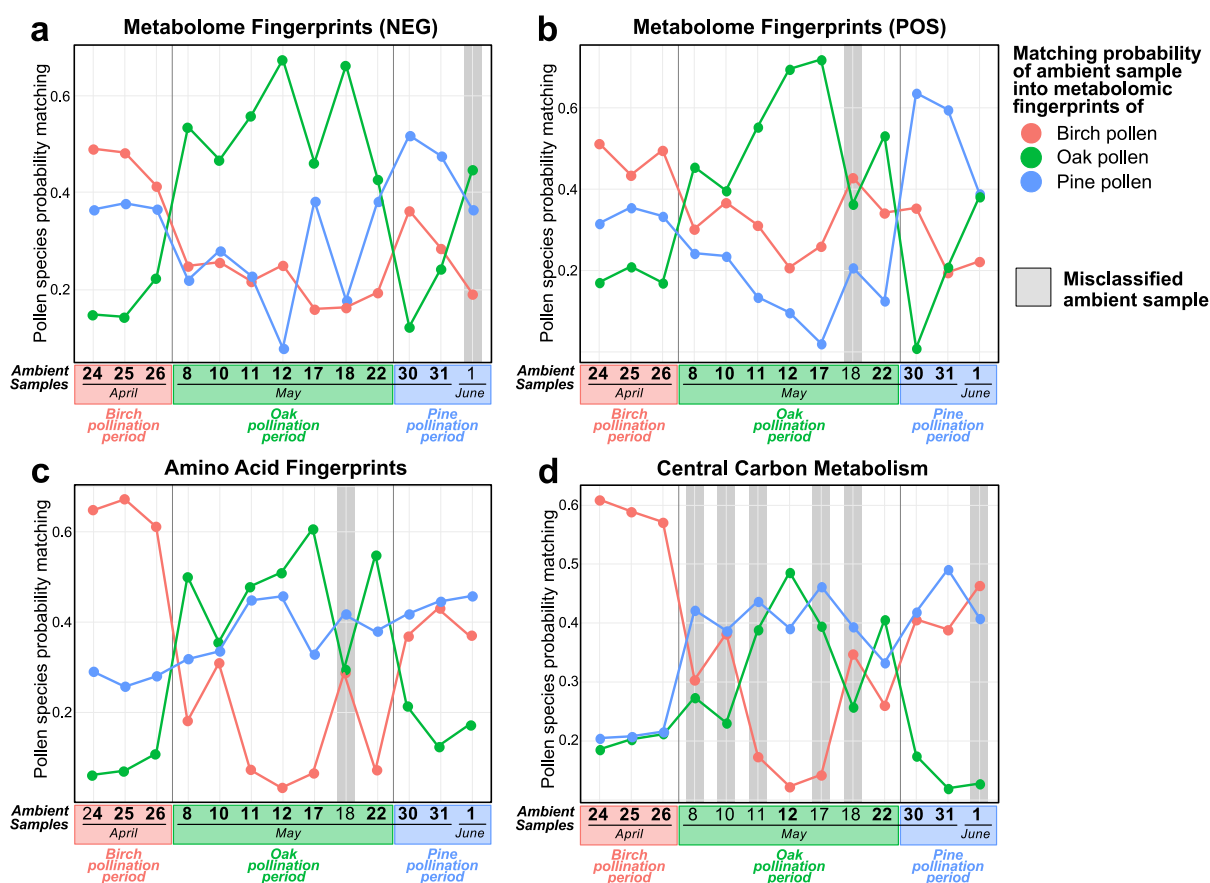


Figure 3. Metabolomic matching probability plots of metabolomic fingerprints of ambient samples into birch (red), oak (green), and pine (blue) pollen metabolomic fingerprints. Results are shown for each data set type used in the machine learning training: (a) metabolome fingerprints obtained in negative ionization mode, (b) metabolome fingerprints obtained in positive ionization mode, (c) amino acid fingerprints (combined ionization modes), and (d) metabolites from the central carbon metabolism (combined ionization modes). Results shown for metabolome fingerprints (a, b) are derived from the sPLS-DA classification method, and results for amino-acid (c) and central carbon metabolites (d) are derived from the Random Forest classification method. For each panel, the different ambient samples collected from April to June 2017 are indicated in the x-axis. Different colors grouping ambient samples indicate the pollination peak periods for each of the tree species for which pollen samples were collected: the birch pollination period is shown in light-red (April 24, 25, and 26), oak pollination period in green (May 8, 10, 11, 12, 17, 18, and 22), and pine pollination period in blue (May 30 and 31, and June 1). The matching probability of metabolomic fingerprints of ambient samples into each pollen species is represented along the Y-axis. Larger probability indicates larger similarity between the metabolome structure between the ambient sample and the indicated pollen species. Ambient samples considered as correctly classified are shown in bold. A sample was considered correctly classified when the pollination peak period of such a specific ambient sample coincided with the highest probability match of its corresponding pollen species. Misclassified samples are marked in gray along the Y-axis. Matching probability values are shown in Table S6 of the Supporting Information. See Figure S4 for more details about the matching probability plot interpretation.

centered and scaled metabolite fingerprints for the pollen samples. The ambient samples were centered and scaled based on the respective values computed from the pollen samples before being predicted with the models trained on the pollen samples. PERMANOVAs, graphs, and models were performed using R version 3.6.1.⁵² PERMANOVAs were conducted using the *adonis* function of the “vegan” package.⁵³ The Heatmap with hierarchical clustering for both variables and study subjects was performed using the *heatmap.2* function in the “gplots” package.⁵⁴ The hierarchical clustering calculation was performed using the “complete” method. Color gradient was plotted using the viridis palette.⁵⁵ The “randomForest” package⁵⁶ was used to fit Random Forest models, and sPLS-DA was fitted using the “mixOmics” package.⁵⁷ Figure 1 was created with Affinity Designer (Serif (Europe) Ltd., UK). Figures 2 and 3 were plotted in R-Studio and subsequently finalized using Affinity Designer (Serif (Europe) Ltd., UK).

3. RESULTS AND DISCUSSION

The metabolomics data set for the pollen and aerosol samples included 43 319 metabolomic features acquired with both negative and positive ionization modes. From those, 6756 (15.6%) were represented only in pollen grains, and 2763 (6.38%) were present only in ambient samples. When contrasting different pollen species, we found that 5910 (13.6% of the total), 4785 (11% of the total), and 805 (1.9% of the total) features were unique for birch, pine, and oak pollen, respectively. The overall metabolome structure differed significantly between pollen species, explaining almost 93% of the total variability (PERMANOVA; pseudo- $F = 97.65$, $R^2 = 0.929$, $P < 0.0001$). A hierarchical clustering tree for samples classified birch pollen with the most dissimilar metabolomes if compared to pine and oak pollen (Figure 2; see Figure S3 for details). Phylogenetically closely related organisms tend to share features of their phenotypes⁵⁸ and their metabolomes, and chemical phenotypes should not be an exception.

However, our metabolomic analyses show large species metabolic plasticity at the pollen level as the phylogenetic and metabolomics relationships between tree species do not match. For example, we would expect more similar metabolomes between oak and birch pollen as both species are angiosperms from the same order (Fagales) and, therefore, more closely related to each other than to pines (gymnosperms) from which they diverged at least 300 million years ago.⁵⁹ Given this large chemical variability between tree species, it would not be surprising to find different metabolomes in pollen from species of the same genus and, therefore, with the potential to impact the components of the biosphere (i.e., atmosphere, plants) in contrasted ways.

The metabolic fingerprints in ambient samples collected at distinct pollination peak periods showed a significant difference between them (PERMANOVA; pseudo- $F = 12.32$, $R^2 = 0.71$, $P < 0.0001$), which could be directly associated with the presence of specific pollen particles (birch, oak, or pine) during the distinct pollination peak periods. Those results suggest that ambient samples collected in ecosystems with high biodiversity composed of evolutionary distant organisms should facilitate the determination of the source of the main PBAPs via high-resolution MS analyses. More research is necessary to determine whether different pollen particles emitted from distinct tree species belonging to the same plant genus can be clearly differentiated in ambient samples using MS data and machine learning approaches. The complexity and vast diversity of compounds found in ambient samples can complicate the interpretation of the results when relying on single compounds for PBAP detection and classification because the abundance of those specific tracers can vary in different particles with diverse sources. For this reason, we hypothesize that using a large set of tracers (i.e., metabolomic features) can significantly improve the classification of PBAPs as identification relies on complex signatures unique to specific PBAPs. In addition, this approach can be efficient enough to detect PBAPs at the genus (i.e., pollen from oak, birch, or pine), and even species, level: a clear advantage over using specific features alone.

We found that the sPLS-DA regularized machine learning method correctly classified 12 of the 13 ambient samples trained on both the positive and negative metabolite fingerprints separately (Figure 3a,b; Table S6) (see Figure S4 for details). For negative ionization mode (Figure 3a), the ambient sample collected on June 1st, 2017, showed more similarity to oak than pine pollen, even though this day corresponds to the pollination peak period for pine. Conversely, fingerprints of ambient samples obtained in positive ionization mode (Figure 3b) matched the June 1st sample correctly with pine pollen, but there was very little difference with oak pollen (Table S6). The ambient sample collected on the 18th of May, 2017 (within the oak pollination peak period), was incorrectly attributed to birch pollen instead of oak pollen. Achieving only 1 misclassified sample out of 13 (7.7% of error) indicates that using machine learning methods on metabolomic fingerprint data from ambient samples can provide enough confidence to identify different genus (i.e., birch, oak, or pine) of PBAPs present in the atmosphere at a given moment. Similarly to the sPLS-DA method, a Random Forest model trained on the pollen samples correctly classified 12 of the 13 (7.7% of error) ambient samples when using the amino acid fingerprint, which contain a small number of metabolic features (Figure 3c). The ambient sample collected

on the 18th of May (within the oak pollination peak period) was incorrectly matched to pine pollen instead of oak pollen using solely the amino acid fingerprint. Amino acids, in addition to acting as the intermediaries of multiple metabolic pathways, are the monomers that comprise proteins and, therefore, the proteome of organisms. Unlike low-weighted molecular compounds, proteins are macromolecules directly linked to specific genes, and different organisms carry their specific set of DNA (genomes), including both genes and noncoding DNA sequences.⁶⁰ This result suggests that well-characterized amino-acid profiles could directly reflect the genomes of the different species and could potentially be used alone to decipher the origin of PBAPs. Further analyses using distinct biological material from diverse taxa are necessary to validate this hypothesis. Random Forest models on the central carbon metabolism signature of ambient samples (Figure 3d) acquired in both positive and negative ionization modes did a poor job of classifying the ambient samples into their corresponding pollination peaks; days 8, 10, 11, 17, and 18 of May (within the oak pollination peak period) were matched with pine pollen instead of oak pollen, and the June 1st ambient sample (within the pine pollination peak period) was matched to birch instead of pine pollen. The vast majority of metabolites of the central carbon metabolism are shared between all living systems,⁶¹ and contrary to proteins, the abundance and diversity of low-weighted compounds are not directly linked to genome translation. Ambient samples can contain different PBAPs which can exert a buffer effect on the overall metabolic signature of the central carbon metabolism alone due the large metabolic feature overlap between particles. Our results suggest that using metabolite features largely shared between PBAPs is not sufficient to properly distinguish the particle source. For this reason, using specific biomarkers alone (i.e., sugars) is not sufficient to clearly distinguish specific PBAPs (i.e., pollen, fungi) in complex ambient samples. In addition, as previously mentioned, single biomarkers cannot properly discriminate pollen at genus or species levels which can impact the atmosphere and ecosystems in particular ways.

3.1. Current Methodological Limitations and Future Perspectives. In this proof-of-concept study, we provided evidence that analyzing multiple chemical tracers (i.e., full metabolic fingerprints) along with advanced bioinformatic tools can not only detect PBAPs in ambient samples but also potentially identify them at the genus or species level. However, revealing the diversity and abundance of PBAPs in complex samples would still be necessary. Our framework used metabolomic data of monospecific PBAPs (birch, oak, or pine) to train machine learning algorithms, which limits the identification to the most abundant particles because those classification methods are based on similarity and therefore cannot reveal the relative abundance of particles in complex mixtures. However, identifying the most abundant PBAP of a sample can still be relevant in certain atmospheric, ecological, and health studies. Future research applying machine learning methods on real metabolomic fingerprints of PBAP complex mixtures is necessary to evaluate the accuracy to decipher the diversity and relative abundance of PBAPs from complex ambient samples. In addition, research is also required to determine the minimum amount of PBAPs present in complex ambient samples to generate useful chemical fingerprints for machine learning classification methods. The concentration of PBAPs in the lower troposphere can be relatively large;

however, this is not the case for the upper troposphere where concentrations can be as low as $1.7 \mu\text{g m}^{-3}$,⁶² which can complicate the detection if samples are not concentrated enough.

In terrestrial and aquatic ecosystems, most elements (C, N, P, S, ...) of organic matter are part of molecular compounds and do not act as individual elements.⁶³ For this reason, the supply of nutrients to the different living components of the biosphere is accomplished via numerous molecular structures. Understanding the detailed molecular composition of deposited PBAPs is of importance for several different aspects. For example, certain molecular structures are “preferred” over others in terms of how microbes acquire specific nutrients such as N.⁶⁴ In microbes, this preference seems to be driven by the higher affinity of specific-compound transporters.⁶⁵ Deciphering accurately the composition and concentration of PBAPs could facilitate the understanding and quantification of nutrient fluxes between different ecosystem components.⁶⁴ Therefore, we hypothesize that contrasted metabolic signatures from deposited PBAP (i.e., pollen from birch, oak, and pines) could impact the ecosystem function and environmental C:N:P stoichiometry in distinct ways. Determining PBAP diversity instead of considering single major groups alone (i.e., pollen, fungal spores) provides a more accurate picture of the nutrient deposition on ecosystems. In addition, understanding particle deposition on ecosystems at the molecular level can significantly help to fill key knowledge gaps of biogeochemical nutrient cycles and ecosystem function, especially under scenarios of climate change leading to significant plant phenological fluctuations. Long-term studies tracking specific pollen, or subpollen particles, in suspension could provide practical information to understand the link between phenological shifts of specific highly pollinating trees at both temporal and spatial resolution.

Given the contrasted efficiency as ice and cloud condensation nuclei,⁶⁶ as well as scattering and absorbing sunlight,^{67,68} understanding the diversity and abundance of different submicron particles in the atmosphere, such as subpollen, provides critical information to interpret the highly complex array of mechanisms of the atmospheric processes aiding to improve Earth system modeling. Given the high sensitivity of MS instruments, combining both metabolomics analyses with machine learning techniques could be a promising tool to detect and identify submicron biological particles within the aerosol mixture in the upper troposphere with high accuracy. This information will improve our understanding of the atmospheric composition and climate dynamics.

In addition, shedding light on the diversity and relative abundance of certain PBAPs at a given time provides critical data to understand the severe outbreaks of allergies and/or asthma. Automated bioinformatic methods can be developed to extract the chemical fingerprints from MS files, search for complex fingerprint mixtures, and provide quickly a summary of the diversity and abundance of pollen and other PBAPs in a specific ambient sample.

In summary, this proof-of-concept study used three high-pollen-emitting tree species that dominate the eastern USA with the goal of testing a new framework to identify PBAPs in the atmosphere. Our results provided evidence that high-resolution MS measurements coupled to sophisticated bioinformatic approaches can identify the most abundant pollen species in ambient samples when using the entire

metabolome, and even using the amino acid profiles (Figure 3). It is important to note that our metabolomics analyses of pollen were performed at the subpollen level for both pollen and ambient samples, as metabolite extraction procedures expose pollen grains to humidity and sonication, making pollen grains rupture into submicron particles. Using metabolomic fingerprints can be specifically beneficial to identify submicron PBAPs, i.e., subpollen particles or fungal spore fragments, as utilizing physical properties alone is challenging due to their similarity with other atmospheric particles.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsearthspacechem.0c00295>.

Environmental data during ambient sample collection periods; parameters applied to process LC-MS RAW files with MZmine 2.38; metabolite annotation data; metabolites included in *amino acid* and *central carbon metabolism* profiles; validation of machine learning models; metabolomic matching probability values; mass spectrometry spectra calibration chart; cell factor concept; hierarchical clustering and heatmap on metabolomics data; and machine learning pollen matching probability plots (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Albert Rivas-Ubach – Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99354, United States; orcid.org/0000-0003-1293-7127; Email: albert.rivas.ubach@gmail.com

Authors

Bryan Stanfill – Applied Statistics and Computational Modeling, Computing and Analytics Division, PNNL, Richland, Washington 99352, United States; orcid.org/0000-0003-0612-5333

Swarup China – Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99354, United States; orcid.org/0000-0001-7670-335X

Ljiljana Paša-Tolić – Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99354, United States

Alex Guenther – Department of Earth System Science, University of California, Irvine, Irvine, California 92697, United States

Allison L. Steiner – Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, Michigan 48109, United States; orcid.org/0000-0002-3823-1512

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acsearthspacechem.0c00295>

Author Contributions

A.R.-U., S.C., A.G., and A.L.S. conceptualized the study. A.L.S. collected the samples. A.R.-U. conducted the metabolomics analyses. A.R.-U. and B.S. analyzed the data. L.P.-T., A.G., and A.L.S. provided the resources for the research. L.P.-T., A.G.,

and A.L.S. supervised the research and administered the project. A.R.-U. wrote the original draft. All authors reviewed and edited the manuscript. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Data and R scripts of machine learning models are available in Github online repository (https://github.com/Rivas-Ubach/Pollen_MachineLearning). Authors thank Ronald Moore and Karl Weitz for their support in the laboratory. A.R.-U. thank Nancy Hess and Eva Baroni for funding management support. This work was supported by the Pacific Northwest National Laboratory (PNNL) directed research funds. Samples were analyzed in EMSL, a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research at PNNL. A.L.S. was supported in part by the National Science Foundation Grant 1821173 and the U.S. Department of Energy's Atmospheric System Research, an Office of Science Biological and Environmental Research program under Grant DE-SC0019084.

REFERENCES

- (1) Després, V. R.; Alex Huffman, J.; Burrows, S. M.; Hoose, C.; Safatov, A. S.; Buryak, G.; Fröhlich-Nowoisky, J.; Elbert, W.; Andreae, M. O.; Pöschl, U.; Jaenicke, R. Primary Biological Aerosol Particles in the Atmosphere: A Review. *Tellus, Ser. B* **2012**, *64*, 15598.
- (2) Fröhlich-Nowoisky, J.; Kampf, C. J.; Weber, B.; Huffman, J. A.; Pöhlker, C.; Andreae, M. O.; Lang-Yona, N.; Burrows, S. M.; Gunthe, S. S.; Elbert, W.; Su, H.; Hoor, P.; Thines, E.; Hoffmann, T.; Després, V. R.; Pöschl, U. Bioaerosols in the Earth System: Climate, Health, and Ecosystem Interactions. *Atmos. Res.* **2016**, *182*, 346–376.
- (3) Pöschl, U. Atmospheric Aerosols: Composition, Transformation, Climate and Health Effects. *Angew. Chem., Int. Ed.* **2005**, *44* (46), 7520–7540.
- (4) Mauderly, J. L.; Chow, J. C. Health Effects of Organic Aerosols. *Inhalation Toxicol.* **2008**, *20* (3), 257–288.
- (5) Darnhofer, B.; Tomin, T.; Liesinger, L.; Schittmayer, M.; Tomazic, P. V.; Birner-Gruenberger, R. Comparative Proteomics of Common Allergenic Tree Pollens of Birch, Alder, and Hazel. *Allergy* **2021**, 14694.
- (6) Taylor, P. E.; Flagan, R. C.; Valenta, R.; Glovsky, M. M. Release of Allergens as Respirable Aerosols: A Link between Grass Pollen and Asthma. *J. Allergy Clin. Immunol.* **2002**, *109* (1), 51–56.
- (7) Taylor, P. E.; Flagan, R. C.; Miguel, A. G.; Valenta, R.; Glovsky, M. M. Birch Pollen Rupture and the Release of Aerosols of Respirable Allergens. *Clin. Exp. Allergy* **2004**, *34* (10), 1591–1596.
- (8) Grote, M.; Valenta, R.; Reichelt, R. Abortive Pollen Germination: A Mechanism of Allergen Release in Birch, Alder, and Hazel Revealed by Immunogold Electron Microscopy. *J. Allergy Clin. Immunol.* **2003**, *111* (5), 1017–1023.
- (9) Steiner, A. L.; Brooks, S. D.; Deng, C.; Thornton, D. C. O.; Pendleton, M. W.; Bryant, V. Pollen as Atmospheric Cloud Condensation Nuclei. *Geophys. Res. Lett.* **2015**, *42* (9), 3596–3602.
- (10) Pummer, B. G.; Budke, C.; Augustin-Bauditz, S.; Niedermeier, D.; Felgitsch, L.; Kampf, C. J.; Huber, R. G.; Liedl, K. R.; Loerting, T.; Moschen, T.; Schauer, M.; Tollinger, M.; Morris, C. E.; Wex, H.; Grothe, H.; Pöschl, U.; Koop, T.; Fröhlich-Nowoisky, J. Ice Nucleation by Water-Soluble Macromolecules. *Atmos. Chem. Phys.* **2015**, *15*, 4077–4091.
- (11) China, S.; Wang, B.; Weis, J.; Rizzo, L.; Brito, J.; Cirino, G. G.; Kovarik, L.; Artaxo, P.; Gilles, M. K.; Laskin, A. Rupturing of Biological Spores As a Source of Secondary Particles in Amazonia. *Environ. Sci. Technol.* **2016**, *50* (22), 12179–12186.
- (12) Rathnayake, C. M.; Metwali, N.; Jayarathne, T.; Kettler, J.; Huang, Y.; Thorne, P. S.; O'shaughnessy, P. T.; Stone, E. A. Influence of Rain on the Abundance of Bioaerosols in Fine and Coarse Particles. *Atmos. Chem. Phys.* **2017**, *17*, 2459–2475.
- (13) Pummer, B. G.; Bauer, H.; Bernardi, J.; Bleicher, S.; Grothe, H. Suspendable Macromolecules Are Responsible for Ice Nucleation Activity of Birch and Conifer Pollen. *Atmos. Chem. Phys.* **2012**, *12* (5), 2541–2550.
- (14) Dreischmeier, K.; Budke, C.; Wiehemeier, L.; Kottke, T.; Koop, T. Boreal Pollen Contain Ice-Nucleating as Well as Ice-Binding 'Antifreeze' Polysaccharides. *Sci. Rep.* **2017**, *7* (1), 41890.
- (15) Zhang, Y.; Bielory, L.; Mi, Z.; Cai, T.; Robock, A.; Georgopoulos, P. Allergenic Pollen Season Variations in the Past Two Decades under Changing Climate in the United States. *Glob. Chang. Biol.* **2015**, *21* (4), 1581–1589.
- (16) Morris, C. E.; Conen, F.; Alex Huffman, J.; Phillips, V.; Pöschl, U.; Sands, D. C. Bioprecipitation: A Feedback Cycle Linking Earth History, Ecosystem Dynamics and Land Use through Biological Ice Nucleators in the Atmosphere. *Glob. Chang. Biol.* **2014**, *20* (2), 341–351.
- (17) Wozniak, A. S.; Bauer, J. E.; Sleighter, R. L.; Dickhut, R. M.; Hatcher, P. G. Technical Note: Molecular Characterization of Aerosol-Derived Water Soluble Organic Carbon Using Ultrahigh Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Atmos. Chem. Phys.* **2008**, *8* (17), 5099–5111.
- (18) Baker, A. R.; Kelly, S. D.; Biswas, K. F.; Witt, M.; Jickells, T. D. Atmospheric Deposition of Nutrients to the Atlantic Ocean. *Geophys. Res. Lett.* **2003**, *30*, 24.
- (19) Mahowald, N. M.; Artaxo, P.; Baker, A. R.; Jickells, T. D.; Okin, G. S.; Randerson, J. T.; Townsend, A. R. Impacts of Biomass Burning Emissions and Land Use Change on Amazonian Atmospheric Phosphorus Cycling and Deposition. *Global Biogeochem. Cycles* **2005**, *19*, GC4030.
- (20) Seco, R.; Peñuelas, J.; Filella, I. Short-Chain Oxygenated VOCs: Emission and Uptake by Plants and Atmospheric Sources, Sinks, and Concentrations. *Atmos. Environ.* **2007**, *41* (12), 2477–2499.
- (21) Paytan, A.; Mackey, K. R. M.; Chen, Y.; Lima, I. D.; Doney, S. C.; Mahowald, N.; Labiosa, R.; Post, A. F. Toxicity of Atmospheric Aerosols on Marine Phytoplankton. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (12), 4601–4605.
- (22) Wang, R.; Balkanski, Y.; Bopp, L.; Aumont, O.; Boucher, O.; Ciais, P.; Gehlen, M.; Peñuelas, J.; Ethé, C.; Hauglustaine, D.; Li, B.; Liu, J.; Zhou, F.; Tao, S. Influence of Anthropogenic Aerosol Deposition on the Relationship between Oceanic Productivity and Warming. *Geophys. Res. Lett.* **2015**, *42* (24), 10745–10754.
- (23) Fageria, N. K.; Filho, M. P. B.; Moreira, A.; Guimarães, C. M. Foliar Fertilization of Crop Plants. *J. Plant Nutr.* **2009**, *32* (6), 1044–1064.
- (24) Uzu, G.; Sobanska, S.; Sarret, G.; Muñoz, M.; Dumat, C. Foliar Lead Uptake by Lettuce Exposed to Atmospheric Fallouts. *Environ. Sci. Technol.* **2010**, *44* (3), 1036–1042.
- (25) Wedding, J. B.; Carlson, R. W.; Stukel, J. J.; Bazzaz, F. A. Aerosol Deposition on Plant Leaves. *Environ. Sci. Technol.* **1975**, *9* (2), 151–153.
- (26) Rivas-Ubach, A.; Liu, Y.; Steiner, A. L.; Sardans, J.; Tfaily, M. M.; Kulkarni, G.; Kim, Y.-M.; Bourrianne, E.; Paša-Tolić, L.; Peñuelas, J.; Guenther, A. Atmo-Ecometabolomics: A Novel Atmospheric Particle Chemical Characterization Methodology for Ecological Research. *Environ. Monit. Assess.* **2019**, *191* (2), 78.
- (27) Vorholt, J. A. Microbial Life in the Phyllosphere. *Nat. Rev. Microbiol.* **2012**, *10* (12), 828–840.
- (28) Rösel, S.; Rychla, A.; Wurzbacher, C.; Grossart, H.-P. Effects of Pollen Leaching and Microbial Degradation on Organic Carbon and Nutrient Availability in Lake Water. *Aquat. Sci.* **2012**, *74* (1), 87–99.
- (29) Sterner, R.; Elser, J. *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere*; Princeton University Press: Princeton, NJ, 2002.

- (30) Sardans, J.; Rivas-Ubach, A.; Peñuelas, J. The Elemental Stoichiometry of Aquatic and Terrestrial Ecosystems and Its Relationships with Organismic Lifestyle and Ecosystem Structure and Function: A Review and Perspectives. *Biogeochemistry* **2012**, *111* (1–3), 1–39.
- (31) Sardans, J.; Rivas-Ubach, A.; Peñuelas, J. The C:N:P Stoichiometry of Organisms and Ecosystems in a Changing World: A Review and Perspectives. *Perspect. Plant Ecol. Evol. Syst.* **2012**, *14* (1), 33–47.
- (32) Fröhlich-Nowoisky, J.; Pickersgill, D. A.; Després, V. R.; Pöschl, U. High Diversity of Fungi in Air Particulate Matter. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (31), 12814–12819.
- (33) Womack, A. M.; Bohannan, B. J. M.; Green, J. L. Biodiversity and Biogeography of the Atmosphere. *Philos. Trans. R. Soc., B* **2010**, *365* (1558), 3645–3653.
- (34) Huffman, J. A.; Perring, A. E.; Savage, N. J.; Clot, B.; Crouzy, B.; Tummon, F.; Shoshanim, O.; Damit, B.; Schneider, J.; Sivaprakasam, V.; Zawadowicz, M. A.; Crawford, I.; Gallagher, M.; Topping, D.; Doughty, D. C.; Hill, S. C.; Pan, Y. Real-Time Sensing of Bioaerosols: Review and Current Perspectives. *Aerosol Sci. Technol.* **2020**, *54* (5), 465–495.
- (35) Naik, A. K.; Hanay, M. S.; Hiebert, W. K.; Feng, X. L.; Roukes, M. L. Towards Single-Molecule Nanomechanical Mass Spectrometry. *Nat. Nanotechnol.* **2009**, *4* (7), 445–450.
- (36) Yttri, K. E.; Dye, C.; Kiss, G. Ambient Aerosol Concentrations of Sugars and Sugar-Alcohols at Four Different Sites in Norway. *Atmos. Chem. Phys.* **2007**, *7* (16), 4267–4279.
- (37) Fu, P.; Kawamura, K.; Kobayashi, M.; Simoneit, B. R. T. Seasonal Variations of Sugars in Atmospheric Particulate Matter from Gosan, Jeju Island: Significant Contributions of Airborne Pollen and Asian Dust in Spring. *Atmos. Environ.* **2012**, *55*, 234–239.
- (38) Speranza, A.; Calzoni, G. L.; Pacini, E. Occurrence of Mono- or Disaccharides and Polysaccharide Reserves in Mature Pollen Grains. *Sex. Plant Reprod.* **1997**, *10* (2), 110–115.
- (39) Bauer, H.; Claeys, M.; Vermeylen, R.; Schueller, E.; Weinke, G.; Berger, A.; Puxbaum, H. Arabitol and Mannitol as Tracers for the Quantification of Airborne Fungal Spores. *Atmos. Environ.* **2008**, *42* (3), 588–593.
- (40) Sardans, J.; Gargallo-Garriga, A.; Urban, O.; Klem, K.; Walker, T. W. N.; Holub, P.; Janssens, I. A.; Peñuelas, J. Ecometabolomics for a Better Understanding of Plant Responses and Acclimation to Abiotic Factors Linked to Global Change. *Metabolites* **2020**, *10* (6), 239.
- (41) Gargallo-Garriga, A.; Preece, C.; Sardans, J.; Oravec, M.; Urban, O.; Peñuelas, J. Root Exudate Metabolomes Change under Drought and Show Limited Capacity for Recovery. *Sci. Rep.* **2018**, *8* (1), 12696.
- (42) Rivas-Ubach, A.; Barbeta, A.; Sardans, J.; Guenther, A.; Ogaya, R.; Oravec, M.; Urban, O.; Peñuelas, J. Topsoil Depth Substantially Influences the Responses to Drought of the Foliar Metabolomes of Mediterranean Forests. *Perspect. Plant Ecol. Evol. Syst.* **2016**, *21*, 41.
- (43) Sánchez, D. H.; Schwabe, F.; Erban, A.; Udvardi, M. K.; Kopka, J. Comparative Metabolomics of Drought Acclimation in Model and Forage Legumes. *Plant, Cell Environ.* **2012**, *35* (1), 136–149.
- (44) Handakumbura, P. P.; Stanfill, B.; Rivas-Ubach, A.; Fortin, D.; Vogel, J. P.; Jansson, C. Metabotyping as a Stopover in Genome-to-Phenome Mapping. *Sci. Rep.* **2019**, *9* (1), 1858.
- (45) Rivas-Ubach, A.; Sardans, J.; Hódar, J. A.; Garcia-Porta, J.; Guenther, A.; Paša-Tolić, L.; Oravec, M.; Urban, O.; Peñuelas, J. Close and Distant: Contrasting the Metabolism of Two Closely Related Subspecies of Scots Pine under the Effects of Folivory and Summer Drought. *Ecol. Evol.* **2017**, *7* (21), 8976–8988.
- (46) Hoffman, J. M.; Soltow, Q. A.; Li, S.; Sidik, A.; Jones, D. P.; Promislow, D. E. L. Effects of Age, Sex, and Genotype on High-Sensitivity Metabolomic Profiles in the Fruit Fly, *Drosophila melanogaster*. *Aging Cell* **2014**, *13* (4), 596–604.
- (47) Rivas-Ubach, A.; Peñuelas, J.; Hódar, J.; Oravec, M.; Paša-Tolić, L.; Urban, O.; Sardans, J. We Are What We Eat: A Stoichiometric and Ecometabolomic Study of Caterpillars Feeding on Two Pine Subspecies of *Pinus Sylvestris*. *Int. J. Mol. Sci.* **2019**, *20* (1), 59.
- (48) t'Kindt, R.; De Veylder, L.; Storme, M.; Deforce, D.; Van Bocxlaer, J. LC-MS Metabolic Profiling of Arabidopsis Thaliana Plant Leaves and Cell Cultures: Optimization of Pre-LC-MS Procedure Parameters. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2008**, *871* (1), 37–43.
- (49) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC Bioinf.* **2010**, *11* (1), 395.
- (50) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.-M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R. Proposed Minimum Reporting Standards for Chemical Analysis. *Metabolomics* **2007**, *3* (3), 211–221.
- (51) Rivas-Ubach, A.; Sardans, J.; Hódar, J. A.; Garcia-Porta, J.; Guenther, A.; Oravec, M.; Urban, O.; Peñuelas, J. Similar Local but Different Systemic Metabolomic Responses of Closely Related Pine Subspecies to Folivory by Caterpillars of the Processionary Moth. *Plant Biol.* **2016**, *18* (3), 484–494.
- (52) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013; <https://www.R-project.org/>.
- (53) Oksanen, J.; Guillaume-Blanchet, F.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P. R.; O'Hara, R. B.; Simpson, G. L.; Solymos, P.; Stevens, M. H. H.; Szöcs, E.; Wagner, H. *Vegan: Community Ecology Package*, R package version 2.5–7; 2020; <https://CRAN.R-project.org/package=vegan>.
- (54) Warnes, G. R.; Bolker, B.; Bonebakker, L.; Gentleman, R.; Huber, W.; Liaw, A.; Lumley, T.; Maechler, M.; Magnusson, A.; Moeller, S.; Schwartz, M.; Venables, B. *Gplots: Various R Programming Tools for Plotting Data*, R package version 3.1.1; 2015; <https://CRAN.R-project.org/package=gplots>.
- (55) Garnier, S. *Viridis: Default Color Maps from "Matplotlib"*, R package version 0.5.1; 2018; <https://CRAN.R-project.org/package=viridis>.
- (56) Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2* (3), 18–22.
- (57) Rohart, F.; Gautier, B.; Singh, A.; Lê Cao, K.-A. MixOmics: An R Package for 'omics Feature Selection and Multiple Data Integration. *PLoS Comput. Biol.* **2017**, *13* (11), e1005752.
- (58) Blomberg, S. P.; Garland, T.; Ives, A. R. Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile. *Evolution (N. Y.)* **2003**, *57* (4), 717–745.
- (59) Zhang, P.; Tan, H. T. W.; Pwee, K.-H.; Kumar, P. P. Conservation of Class C Function of Floral Organ Development during 300 Million Years of Evolution from Gymnosperms to Angiosperms. *Plant J.* **2004**, *37* (4), 566–577.
- (60) Zivy, M.; Vienne, D. de. Proteomics: A Link between Genomics, Genetics and Physiology. *Plant Mol. Biol.* **2000**, *44* (5), 575–580.
- (61) Smith, E.; Morowitz, H. J. Universality in Intermediary Metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (36), 13168–13173.
- (62) Cozic, J.; Verheggen, B.; Weingartner, E.; Crosier, J.; Bower, K. N.; Flynn, M.; Coe, H.; Henning, S.; Steinbacher, M.; Henne, S.; Collaud Coen, M.; Petzold, A.; Baltensperger, U. Chemical Composition of Free Tropospheric Aerosol for PM1 and Coarse Mode at the High Alpine Site Jungfraujoch. *Atmos. Chem. Phys.* **2008**, *8* (2), 407–423.
- (63) Rivas-Ubach, A.; Sardans, J.; Pérez-Trujillo, M.; Estiarte, M.; Peñuelas, J. Strong Relationship between Elemental Stoichiometry and Metabolome in Plants. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (11), 4181–4186.
- (64) Warren, C. R. Organic N Molecules in the Soil Solution: What Is Known, What Is Unknown and the Path Forwards. *Plant Soil* **2014**, *375* (1–2), 1–19.

- (65) Roberts, P.; Jones, D. L. Microbial and Plant Uptake of Free Amino Sugars in Grassland Soils. *Soil Biol. Biochem.* **2012**, *49*, 139–149.
- (66) Gute, E.; Abbatt, J. P. D. Ice Nucleating Behavior of Different Tree Pollen in the Immersion Mode. *Atmos. Environ.* **2020**, *231*, 117488.
- (67) Liu, C.; Yin, Y. Inherent Optical Properties of Pollen Particles: A Case Study for the Morning Glory Pollen. *Opt. Express* **2016**, *24* (2), A104.
- (68) Noh, Y. M.; Müller, D.; Lee, H.; Choi, T. J. Influence of Biogenic Pollen on Optical Properties of Atmospheric Aerosols Observed by Lidar over Gwangju, South Korea. *Atmos. Environ.* **2013**, *69*, 139–147.