# Characterizing the contribution of bioaerosols diversity from complex aerosol particle samples.

September 2021

Albert Rivas-Ubach (PI)
Swarup China
Paritosh Pande
Xueyun Zheng

**DISCLAIMER**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, **makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights**. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

# Characterizing the contribution of bioaerosols diversity from complex aerosol particle samples.

September 2021

Albert Rivas-Ubach (PI)
Swarup China
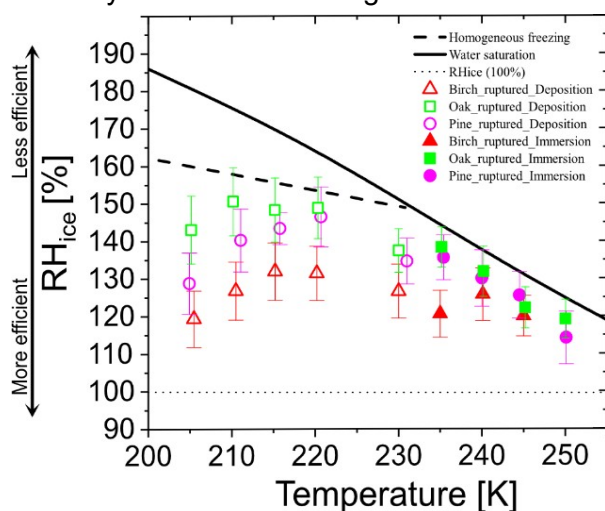Paritosh Pande
Xueyun Zheng

Pacific Northwest National Laboratory
Richland, Washington 99354

**Abstract.**

Current global change scenarios expect to significantly increase the contribution of pollen to the total organic aerosol budget. In addition, pollen can burst into smaller fragments that are highly efficient as cloud condensation nuclei. However, current climate models still do not incorporate the effects of these particles due to detection and quantification challenges in complex aerosol mixtures. Chemical biomarkers (i.e. fructose) have commonly been used to trace pollen but this approximation can generate false positive results given the large number of shared compounds between different bioaerosols. A previous pilot EMSL experiment performed on pure pollen samples suggested that using a set of metabolites as a fingerprint of bioaerosols could serve to substantially increase detection accuracy over single biomarker approaches. In this proposal we developed novel bioinformatic strategies to characterize and quantify complex pollen mixtures present in the atmosphere. For that, we used diverse machine learning algorithms to analyze metabolic fingerprints from complex mixtures of different pollen acquired with high-resolution mass spectrometry (Orbitrap Q-Exactive). Our preliminary results demonstrated the potential to accurately discern and identify at specific relative abundances different pollen species in complex mixtures. Our research will lead to a significant advancement in the atmospheric chemistry and provide much needed data to improve the accuracy of climate models.

**Background.**

Primary biological aerosol particles (PBAPs) are solid airborne particles of biological origin and are mainly represented by pollen, fungal spores, protozoa, bacteria, algae, and biological debris emitted to the atmosphere (Després et al., 2012). PBAPs have commonly received substantial attention within the human health community as they have been associated with important asthma and allergic rhinitis outbreaks (Fröhlich-Nowoisky et al., 2016; Mauderly & Chow, 2008) but also critically affect climate and the ecosystem function (Fröhlich-Nowoisky et al., 2016). Climate change has proven to impact plant phenology by lengthening the green cover period of ecosystems and altering flower abundance and pollination time of plants (Llorens & Peñuelas, 2005; van Vliet et al., 2002). Anemophilous pollen accounts for a significant amount of biological material in the atmosphere, especially during pollination seasons. Pollen can rupture under high humidity conditions resulting in the release of sub-micron fragments (Steiner et al., 2015; Taylor et al., 2002) that can be released to the atmosphere (Rathnayake et al., 2017) and, although still to be proven, reach the upper troposphere. In controlled experiments, pollen fragments have proven to act as cloud condensation nuclei (Steiner et al., 2015) and are also effective in nucleating ice (Dreischmeier et al., 2017). In fact, previous EMSL ice-nucleation experiments corroborated the potential of pollen fragments to significantly impact climate (Fig. 1). Therefore, pollen fragments have the potential to significantly impact precipitation and the hydrological cycle (Després et al., 2012; Morris et al., 2014). Although the concentration of pollen fragments 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 sub-pollen



***Fig 1.*** *Ice nucleation efficiency for three different pollen fragments (birch, oak and pine).*

particle production, thus affecting the total organic aerosol mass loading in the atmosphere (Wozniak et al., 2008).

Pollen fragments are still not possible to detect and quantify in a direct way. Applying metabarcoding techniques on complex ambient samples can provide good data on detection and quantification of certain bioaerosols but DNA of particles in the atmosphere can often be damaged which difficult amplification, sequencing and identification. In addition, metabarcoding cannot cope with sub-micron particles (i.e., pollen fragments) as they do not contain DNA or has been substantially damaged. Similar to sequencing techniques, microscopy cannot identify sub-micron size fragments as most organic aerosols can be very similar in morphology and size. A vast majority of the atmospheric community currently use mass spectrometry (MS) instruments as one of the main techniques for detection of biological particles in the atmosphere, such as pollen, given its sensitivity (Huffman et al., 2020). Single molecular biomarkers, such as glucose and sucrose, are in large proportion in pollen (Fu et al., 2012; Speranza et al., 1997) and have been used as pollen tracers (Rathnayake et al., 2017). However, the complexity and diversity of compounds in ambient samples complicates the interpretation of the results when relying on single compounds alone for bioaerosol detection and classification because those compounds are also present in other organic aerosols. For this reason, it has not been possible to differentiate fragmented bioaerosols within an ambient sample using biomarkers alone yet. We thus hypothesize that using a large set of metabolic features (i.e., metabolic fingerprints) should significantly improve bioaerosol identification over single biomarkers because it relies on complex signatures and no single features. Previous pilot EMSL MS-based metabolomics experiments on pollen suggested that metabolic fingerprints may serve in an accurate way to detect and distinguish the most abundant pollen species in the atmosphere (Rivas-Ubach et al., 2021). However, using metabolomic data of unique bioaerosols to train machine learning algorithms limits the identification to the most abundant particles because those bioinformatic classification methods are based on similarity, and therefore, they can't reveal the relative abundance of particles in complex mixtures. For this reason, combining mass spectrometry data of complex bioaerosol mixtures with machine learning algorithms can shed light into the diversity and relative contribution of different PBAPs (pollen fragments in this case) in the atmosphere. We hypothesize that analyzing complete metabolic signatures of pollen fragments instead of using single compounds will significantly help to characterize and quantify bioaerosol diversity in the atmosphere. In this project, experiments were conducted using pollen samples collected in 2017 from three major forestry species of trees (oak, pine, and birch) in Michigan.

Metabolites of pure and several different pollen mixture combinations were extracted using standard and well described protocols (t'Kindt et al., 2008). This first step generated already pollen fragments. For deep metabolite characterization, samples were analyzed in high-resolution LC-MS Orbirap Q-Exactive operating at a resolving power of 140,000 FHWM. The data from pure pollen and known mixtures of pollen of Oak, Pine and Birch will be using as training and testing data in different machine learning models (Fig 2). We explored two different machine learning (ML) approaches on pollen mixture fingerprints. In the first approach, the metabolite signatures of known pollen species will be used in a supervised learning framework to train various models, including but not limited to Random Forests, sparsity-based discriminant analysis, and support vector machines, to classify ambient mixtures of complex pollen particles into known pollen species. Such supervised learning approaches classify a mixture into constituent species by assigning a probabilistic classification score to each species that represents the degree of similarity between the metabolic fingerprints of the mixture and the individual species. As a result, the most abundant species in a mixture is assigned the highest score, followed by the next most abundant species and so on. The second approach belongs to a class of algorithms generally known as "source separation" or "spectral unmixing". Unlike supervised ML methods, where the model during the training phase, learns a rule to best separate unknown test samples into known classes, in unmixing algorithms, the bulk signal (metabolic fingerprints of the ambient sample) is

explicitly modeled as a linear or nonlinear mixing process of the underlying source signals (metabolic signatures of individual pollen species). The model coefficients are interpreted as the relative contributions of the source signals to the mixed signal. Since the supervised learning and unmixing model the contributions of individual pollen species to the bulk ambient samples, the information about the relative abundances of pollen species in the bulk sample obtained from these approaches can be expected to provide different insights.
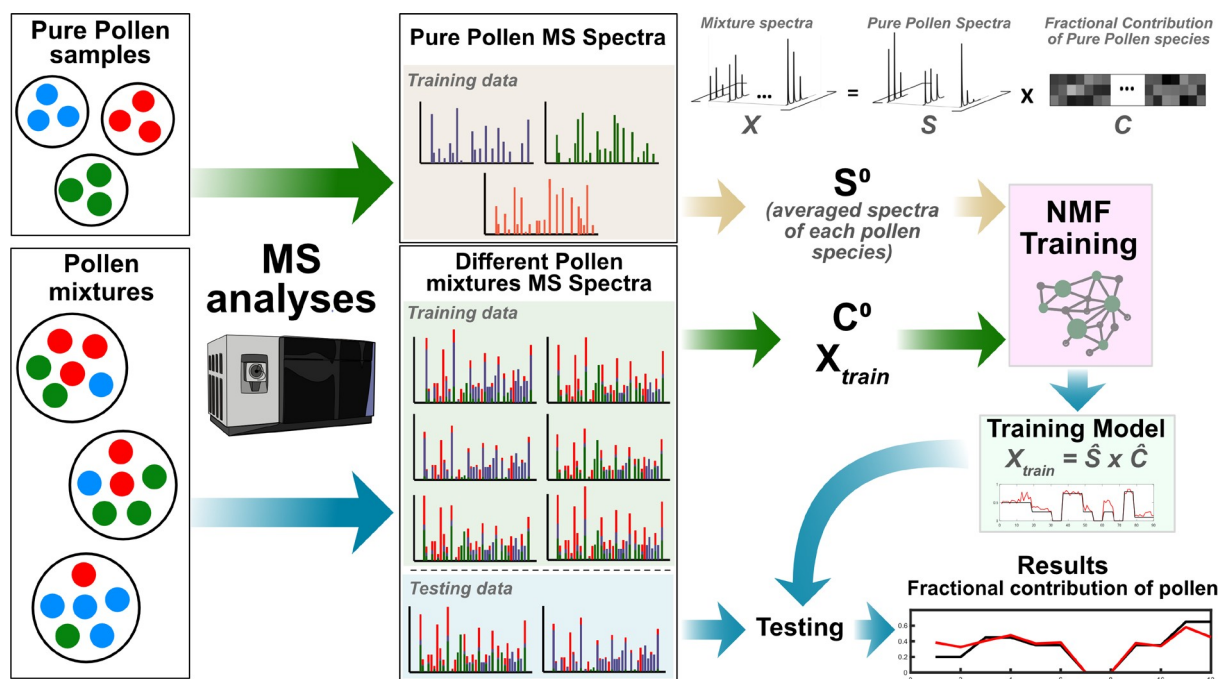


**Fig.3** *Illustration of the use of machine learning methods (Nonnegative Matrix Factorization (NMF)) on Mass Spectrometry-based metabolomic fingerprints of pure pollen and pollen mixtures samples analyzed in positive ionization mode. The averaged spectrum of each pure pollen species ($S^0$) spectra (6 replicates each species) was used as the reference fingerprint identifying each of the species (training data). The metabolomic fingerprints of complex samples ($X_{train}$) and their pollen species relative contribution (birch, oak, and pine) ($C^0$) (training data; 90 samples) were used for the NMF training to generate a prediction model ($X_{train} = \hat{S} \times \hat{C}$) for testing unknown mixtures of pollen (testing data).*

**Preliminary results.**
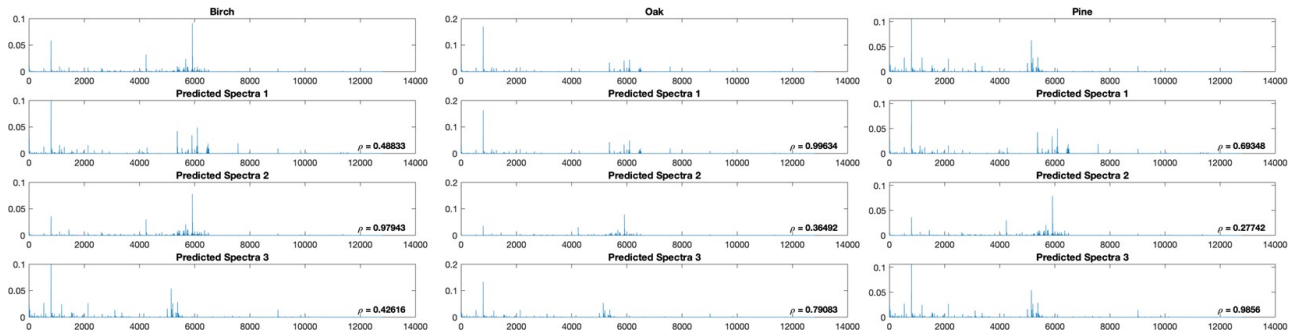


*Fig.3 Correlation between fractional predicted spectra (Predicted spectra 1, 2, and 3) of mixture samples ($X_{train}$) and the averaged spectra of each of the pollen species (Birch, Oak, and Pine) ($S^0$). Clearly, the training model is able to accurately predict a spectra of a pure pollen species within a complex pollen mixture (correlations of p = 0.979 for Birch pollen; p = 0.996 for Oak pollen; p = 0.986 for Pine pollen).*
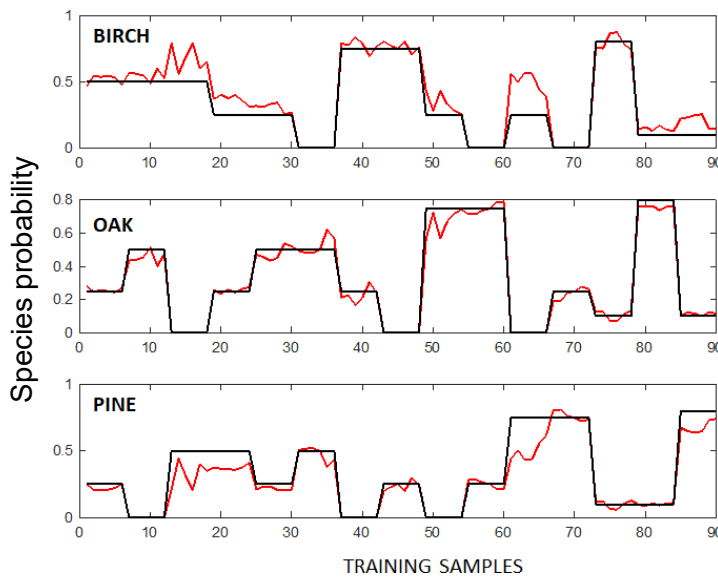


**Fig. 4.** Nonnegative matrix Factorization (NMF) model for all mixture samples (x axis). Black lines correspond to the real contribution ($C^0$) of each pollen species in each of the mixtures. The red line represents the predicted contribution (species probability) of each pollen species on the mixture.
The prediction model adjusted very well to the real contribution of the pollen species for each of the sample mixtures.
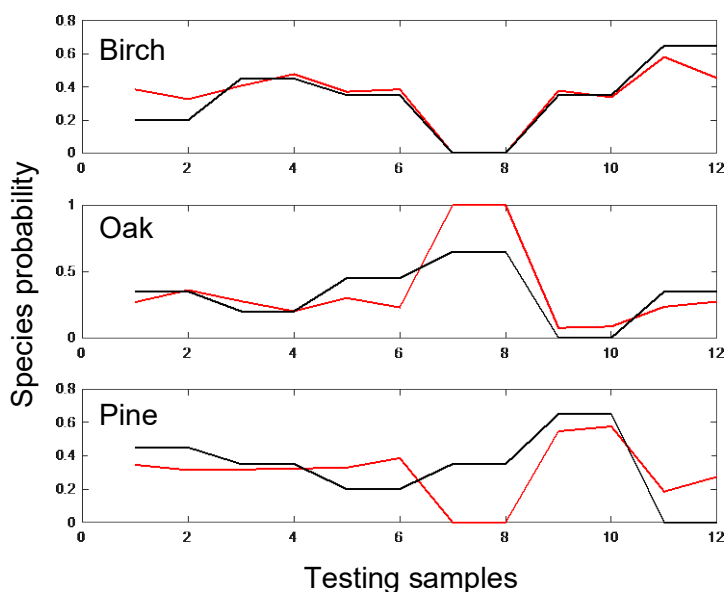
**Fig. 5.** Testing the generated NMF model into a set of samples (total of 12) which information was not used to train the model. We observed a good prediction of the relative contribution of each pollen species to the real mixtures. The most adjusted species was birch, followed by oak and pine which metabolome profiles were slightly confused (especially for samples 7 & 8). This result is in accordance with our past results (Rivas-Ubach et al., 2021) showing a higher resemblance between oak and pine metabolome polarn and semi-polar extractions.

**Research in progress.**

· Determine the relative contribution of more than 3 different pollen species when two of them are from the same genus.
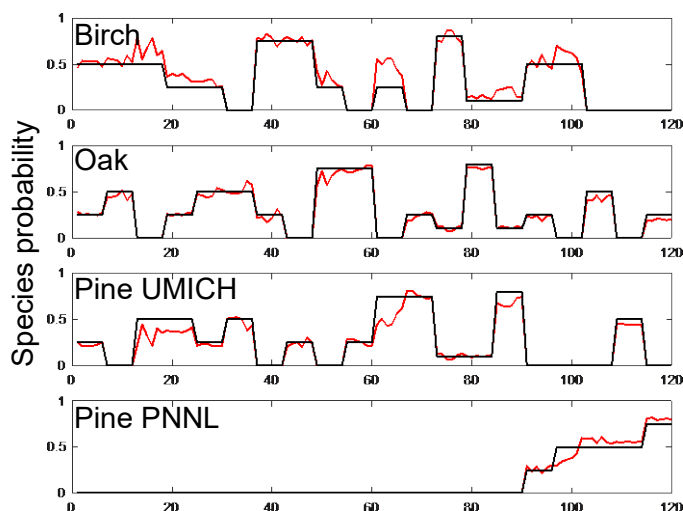


**Fig. 5.** Preliminary NMF model on complex pollen mixtures including up to 4 different species of trees. Two of the species correspond to the same genus (Pinus: "Pine UMICH" and "Pine PNNL").

· Predict the contribution of mixtures of 3 different pollen species based on training of mixtures including 4 different pollen species.
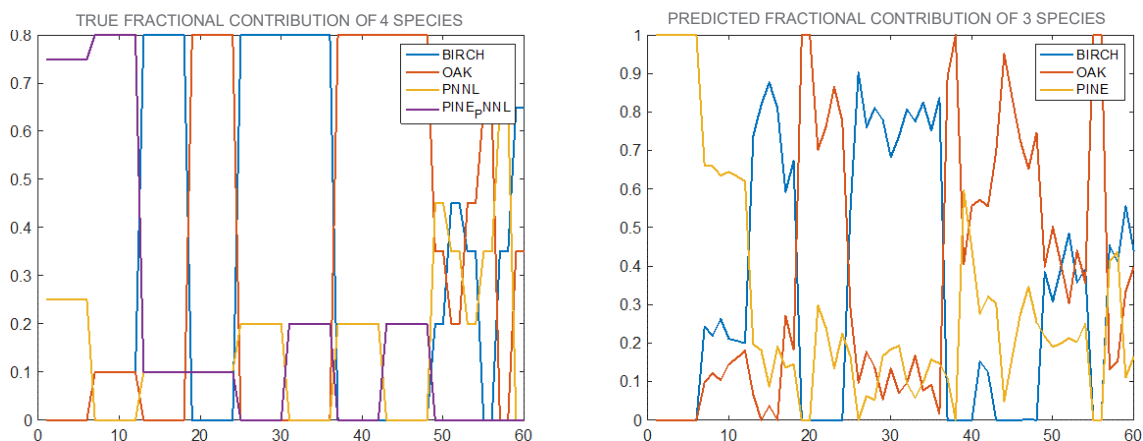


**Fig. 6.** Left panel - Preliminary NMF model trained on complex pollen mixtures including up to 4 different species of trees. Two of the species correspond to the same genus (Pinus: "Pine UMICH" and "Pine PNNL"). Right panel – Preliminary results of the prediction of mixture samples containing 3 pollen species. The model can't discern properly between the two pine species suggesting that machine learning models to decipher the relative contribution of different PBAPs in the atmosphere could be limited at genus level. Future research using different models is necessary to definitely confirm whether machine learning models are capable to distinguish very similar PBAPs in the atmosphere.

Future Exploration:

- How do predictions change when an incorrect number of components (pollen species) are used?

- Perform dimensionality reduction/feature selection to identify most relevant markers for pure pollen species (i.e., sPLS-DA)

- Compare current results in positive ionization mode with negative ionization mode metabolomic fingerprints of pure pollen and mixtures of pollen.

## References.

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. (2012). Primary biological aerosol particles in the atmosphere: a review. *Tellus B*, *64*(15598), 1–58. https://doi.org/10.3402/tellusb.v64i0.15598

Dreischmeier, K., Budke, C., Wiehemeier, L., Kottke, T., & Koop, T. (2017). Boreal pollen contain ice-nucleating as well as ice-binding 'antifreeze' polysaccharides. *Scientific Reports*, *7*(1), 41890. https://doi.org/10.1038/srep41890

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. (2016). Bioaerosols in the Earth system: Climate, health, and ecosystem interactions. *Atmospheric Research*, *182*, 346–376. https://doi.org/10.1016/J.ATMOSRES.2016.07.018

Fu, P., Kawamura, K., Kobayashi, M., & Simoneit, B. R. T. (2012). Seasonal variations of sugars in atmospheric particulate matter from Gosan, Jeju Island: Significant contributions of airborne pollen and Asian dust in spring. *Atmospheric Environment*, *55*, 234–239. https://doi.org/10.1016/J.ATMOSENV.2012.02.061

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. (2020). Real-time sensing of bioaerosols: Review and current perspectives. *Aerosol Science and Technology*, *54*(5), 465–495. https://doi.org/10.1080/02786826.2019.1664724

Llorens, L., & Peñuelas, J. (2005). Experimental Evidence of Future Drier and Warmer Conditions Affecting Flowering of Two Co-occurring Mediterranean Shrubs. *International Journal of Plant Sciences*, *166*(2), 235–245. https://doi.org/10.1086/427480

Mauderly, J. L., & Chow, J. C. (2008). Health Effects of Organic Aerosols. *Inhalation Toxicology*, *20*(3), 257–288. https://doi.org/10.1080/08958370701866008

Morris, C. E., Conen, F., Alex Huffman, J., Phillips, V., Pöschl, U., & Sands, D. C. (2014). Bioprecipitation: a feedback cycle linking Earth history, ecosystem dynamics and land use through biological ice nucleators in the atmosphere. *Global Change Biology*, *20*(2), 341–351. https://doi.org/10.1111/gcb.12447

Rathnayake, C. M., Metwali, N., Jayarathne, T., Kettler, J., Huang, Y., Thorne, P. S., O&amp;apos;Shaughnessy, P. T., & Stone, E. A. (2017). Influence of rain on the abundance of bioaerosols in fine and coarse particles. *Atmospheric Chemistry and Physics*, *17*(3), 2459–2475. https://doi.org/10.5194/acp-17-2459-2017

Rivas-Ubach, A., Stanfill, B., China, S., Paša-Tolić, L., Guenther, A., & Steiner, A. L. (2021). Deciphering the Source of Primary Biological Aerosol Particles: A Pollen Case Study. *ACS Earth and Space Chemistry*, *5*(4), 969–979. https://doi.org/10.1021/ACSEARTHSPACECHEM.0C00295

Speranza, A., Calzoni, G. L., & Pacini, E. (1997). Occurrence of mono- or disaccharides and

polysaccharide reserves in mature pollen grains. *Sexual Plant Reproduction*, *10*(2), 110–115. https://doi.org/10.1007/s004970050076

Steiner, A. L., Brooks, S. D., Deng, C., Thornton, D. C. O., Pendleton, M. W., & Bryant, V. (2015). Pollen as atmospheric cloud condensation nuclei. *Geophysical Research Letters*, *42*(9), 3596–3602. https://doi.org/10.1002/2015GL064060

t'Kindt, R., De Veylder, L., Storme, M., Deforce, D., & Van Bocxlaer, J. (2008). LC-MS metabolic profiling of Arabidopsis thaliana plant leaves and cell cultures: optimization of pre-LC-MS procedure parameters. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, *871*(1), 37–43. https://doi.org/10.1016/j.jchromb.2008.06.039

Taylor, P. E., Flagan, R. C., Valenta, R., & Glovsky, M. M. (2002). Release of allergens as respirable aerosols: A link between grass pollen and asthma. *The Journal of Allergy and Clinical Immunology*, *109*(1), 51–56. https://doi.org/10.1067/MAI.2002.120759

van Vliet, A. J. H., Overeem, A., De Groot, R. S., Jacobs, A. F. G., & Spieksma, F. T. M. (2002). The influence of temperature and climate change on the timing of pollen release in the Netherlands. *International Journal of Climatology*, *22*(14), 1757–1767. https://doi.org/10.1002/joc.820

Wozniak, A. S., Bauer, J. E., Sleighter, R. L., Dickhut, R. M., & Hatcher, P. G. (2008). Technical Note: Molecular characterization of aerosol-derived water soluble organic carbon using ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Atmospheric Chemistry and Physics*, *8*(17), 5099–5111. https://doi.org/10.5194/acp-8-5099-2008

## Pacific Northwest
## National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354
1-888-375-PNNL (7665)

*www.pnnl.gov*