Journal of proteome research

pubs.acs.org/jpr

Article

Metaproteomics as a Powerful Tool for an Extensive Characterization of Ambient Bioaerosols

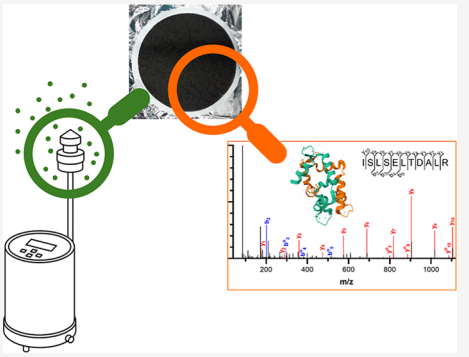
Mathieu Galmiche,*,# Lauriane Kuhn,# Maurice Millet, and Yannis-Nicolas François

Cite This: https://doi.org/10.1021/acs.jproteome.4c00936

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Bioaerosols are of great health and environmental concern. Current techniques for their characterization are generally designed to detect individual species or oppositely unspecific molecular tracers. Metaproteomics on the other hand features the possibility to cover a broad range of taxonomies in a single analysis. This work presents a successful application of metaproteomics to characterize the biological fraction of airborne particulate matter (PM). A bottom-up proteomic strategy was employed, including protein extraction by ultrasonication in aqueous buffer, in-solution tryptic digestion, and nanoflow liquid chromatography−high-resolution mass spectrometry analysis. Extraction parameters were optimized to enhance proteins' recovery. The method was validated on *Escherichia coli* extracts before its application on ambient PM_{10} samples collected over 12 weeks in Strasbourg, France. A total of 1,087 peptides were detected across all samples, with a weekly average of 223 ± 104 peptides corresponding to 111 ± 40 proteins. Peptides from species belonging to animals, plants, fungi, bacteria, and archaea kingdoms were inventoried. Many of them proved to be very relevant, as they were related to human allergens and pathogens, plant pathogens, or ecological indicators. In this work, the major benefits of metaproteomics, yet rather unexploited, as well as its pitfalls and challenges for a broader application in atmospheric chemistry, are discussed.



KEYWORDS: *environmental proteomics, aerobiology, primary biological aerosol particles, PM_{10}, sample preparation, liquid chromatography, mass spectrometry, bottom-up proteomics*

INTRODUCTION

Airborne particulate matter (PM) is generally made up of inorganic, organic, and biological fractions. The latter is called "bioaerosol" or primary biological aerosol particles (PBAP).[1−3] This fraction can account for up to 25% of total PM and even up to 70% if only the coarsest particles are considered.[4]

These bioaerosols are very relevant not only for the understanding of atmospheric and meteorological processes[1,2,4,5] but also for their consequences on human health due to the pathogens they can convey.[2,5−9] Human exposure to allergens and pathogens from airborne particles can occur via inhalation of respirable particles or oral and dermal contact with settled dust.[10]

Bioaerosols host a wide variety of cellular and molecular components released by viruses, bacteria, fungi, plants, or animals, among others.

Viruses are generally treated as distinct from all other PBAP components because of their smaller size (from 0.02 to 0.3 μm),[2,3,9] even though they tend to be associated with bigger particles.[2,11] Viruses are commonly transmitted by airborne particles, as demonstrated by the recent pandemic caused by the SARS-CoV-2 virus.[11]

Bacteria may be released from many sources such as volatilization from wastewater and soils, as well as animal and vegetal fragments.[2,12] Their airborne size can vary among ultrafine, fine, and coarse particles, ranging from 0.5 to 10 μm.[3,9] *Firmicutes*, *Proteobacteria*, and *Actinobacteria* are the most frequently detected bacterial phyla in bioaerosols.[2,12,13]

In addition to virus and bacteria, fungi are the major airborne pathogens for plant, animal, and human health.[14] They are known to be responsible for many respiratory allergies and pathologies.[15−18] Phylogenetic diversity among airborne fungal spores can be enormous, but *Cladosporium*, *Alternaria*, *Penicillium*, and *Aspergillus* species are the most common in bioaerosols.[2,12,14,15] Their size distribution encompasses particles from 0.5 to 50 μm.[3]

Pollen grains are also a class of biological constituents that are frequently released and transported in the atmospheric compartment. Their size generally ranges from 10 to 100 μm, but they can also be disrupted, releasing airborne cytoplasmic granules and other fragments between 30 nm and 5 μm.[2,3]

Received: October 17, 2024
Revised: March 12, 2025
Accepted: May 13, 2025

Outside pollens, other fragments of vegetal origin such as humic substances or biogenic polymers can make up a major fraction of total bioaerosols.[1,2]

Finally, animals, especially domestic animals, are a known source of allergens, especially indoors, via their fur.[15,19] Animal allergenic proteins are commonly associated with sub-10 μm particles.[19] Other animals such as mice are also a common source of allergens in the bioaerosols of our daily life,[20] and dust mites are considered to be the most significant source of indoor allergens,[15,21] via their corpses and fecal particles, generating 10−40 μm particles.[3,15,19]

Among the previous works that were designed to characterize bioaerosols, optical microscopy has been regularly implemented, mostly to fungal spores and pollen.[2,22,23] Its more advanced counterpart, fluorescence in situ hybridization, can also be applied for a broader quantitative application,[2,4] but microscopic techniques are generally restricted to this kind of constituents.

On the other hand, it is possible to quantitatively determine the total protein content in bioaerosols via colorimetric assays such as Bradford, BCA, and NanoOrange,[24−27] or via HPLC-UV for amino acids.[25,27] Organic molecular tracers, such as mannitol and levoglucosan, can also be used as proxies for airborne fungal spores and biomass pyrolysis, respectively. These compounds are usually quantified by chromatographic methods.[28−30] Therefore, these types of analytical techniques provide rough estimates of the biological material amount in aerosols, but they are short of species identification.

Another possibility to assess the abundance of some allergens of specific interest is their targeted quantification with enzymatic immunoassays.[21,31,32]

The last approach to tackle the challenging characterization of PBAP relies on metagenomic and metatranscriptomic techniques that belong to the emerging field of environmental DNA (eDNA) such as PCR, 16S rRNA sequencing, whole-genome shotgun sequencing, or metabarcoding (next-generation sequencing).[9,13,14,33−37] They are widely used and may even bring quantitative information[2,33] but might be limited in breadth by the availability of sensitive and specific primers and the difficulty of expansion toward new targets.[2,38] They seem to be the most advanced tool to date for the taxonomic and functional description of bioaerosols but might still be short of taxonomic depth.[39]

Therefore, it is remarkable that none of the previously cited approaches has the capability to identify the entire complex array of species in the biological fraction of aerosols.

Metaproteomics could well be a crucial approach to deciphering the complex composition of bioaerosols. It has been defined by Wilmes and Bond as "the large-scale characterization of the entire protein complement of environmental microbiota at a given point in time".[40] It comes as a complement to other "meta-omics" sciences (metagenomics, metatranscriptomics, and metametabolomics) as an appropriate strategy for investigating the functional activity of all species within a given environment along with their phenotypes. Metaproteomics has demonstrated its efficiency for the characterization of proteins in wastewater and sewage sludge,[41−43] in dissolved organic matter from rivers[44] and oceans,[45,46] in agricultural soils,[47] and recently in aerosols.[39,48,49]

However, to the best of our knowledge, the number of metaproteomic studies on aerosols remains limited to three. Liu et al. investigated multiple samples of fine, coarse, and total PM, collected over a one-week timespan in an outdoor peri-urban environment. They identified 33 proteins in a total PM sample, with twenty originating from plants (including several pollen allergens) and ten from molds.[48] Piovesana et al. investigated various work environments, including a composting plant (179 identified proteins), a wastewater treatment station (15 identified proteins), and a farm (444 identified proteins). Molds were the predominant source of proteins in this study, followed by proteobacteria.[49] Meyer et al. studied swine confinement buildings, where they identified about 11,872 peptides in a porker barn and 4395 in a sow barn, mostly originating from bacteria.[39]

In the present work, we propose to further increase the knowledge of ambient bioaerosols. This study was focused on outdoor urban aerosol samples to provide meaningful results for the whole population, beyond acute occupational exposure. The enhancement of the extraction of proteins from PM was particularly optimized along with the sample treatment before proteomic analysis. Compared to similar environmental samples from the study of Liu et al., about 3.4 times more proteins were reported on average in weekly samples of the present work, and 13.5 times more when all weekly samples were combined.[48] The peptides identified herein indicate the occurrence of known allergens from plants and fungi in ambient aerosols, confirming the need for improvements in the monitoring and remediation of these biological pollutants.

EXPERIMENTAL SECTION

Chemicals and Reagents

Standard bovine serum albumin (BSA) was obtained from Sigma-Aldrich (Merck KGaA). Bradford and bicinchoninic acid (BCA) colorimetric assays for protein quantification were performed using commercially available kits (Bio-Rad Protein Assay from Bio-Rad Laboratories, Inc. and Pierce BCA Protein Assay Kit from Thermo Fisher Scientific, respectively). Ultrapure water was obtained by using a Milli-Q purification system (Merck KGaA). High-purity LC−MS (Chromasolv)-grade solvents were used for the nanoLC-HRMS/MS analysis (Riedel-de-Haën, Honeywell International Inc.), along with LC−MS-grade formic acid (Fluka, Honeywell International Inc.).

Sampling

The samples of interest of this study on bioaerosols were collected on the roof of the Botanical Institute in Strasbourg (France, 48°35′2″ N, 7°46′1″ E), located close to the city center, thus representative of an urban environment. An additional contribution from the nearby botanical garden was expected to release vegetal fragments and pollen in the ambient aerosol.

PM_{10} samples were collected thanks to a Sven Leckel LVS6-RV medium-volume air sampler (Sven Leckel Ingenieurbüro GmbH) operated at 2.3 m^3/h for a whole week, during 12 consecutive weeks from April 21, 2021 to July 15, 2021. All PM_{10} samples were collected on glass fiber filters (GFFs) (Ø 47 mm, GF6, Whatman, GE Healthcare). Before sampling, the GFFs were rinsed with water (H_2O) and acetonitrile (ACN) and conditioned at 50 °C for at least 48 h to remove contaminants.

Between collection and extraction, PM_{10} samples were stored in an aluminum foil at −20 °C.

Complementary information on the meteorological conditions associated with these 12 one-week samples were retrieved from the Web site infoclimat.fr.[50] Concurrent pollen levels in this period were evaluated via the allergy risk alerts of the French aerobiology network.[51]

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX

Sample Preparation

The optimization of the sample preparation procedure was performed thanks to GFFs that were previously used for PM_{10} sampling. They were first washed in a water/acetonitrile bath (1:1, v/v) to remove traces of naturally occurring proteins and then dried overnight at 50 °C. Thus, only the carbonaceous particulate matrix on the filter was preserved. Then, each GFF was spiked with 50 μL of 2 mg/mL solution of BSA in water (i.e., 100 μg BSA per filter). The filters were cut into pieces or ground in liquid nitrogen and transferred in 5 mL Eppendorfs with the extraction solution. Extraction was performed in an ultrasonic bath. Ten microliters of the concentrated extracts was subject to protein quantification using Bradford and/or BCA colorimetric assays based on standard protocols.[52−54] The evaluation of each extraction method parameter optimized in this work is presented later in this paper, and the procedure for this evaluation is illustrated in Figure S1.

The optimized extraction conditions implemented for the analysis of real-world PM_{10} samples, and confirmed with the proof-of-concept experiment were as follows: filters were first ground in liquid nitrogen, and then the resulting fragments were transferred in a 5.0 mL Eppendorf tube (Eppendorf SE). Proteins were extracted in 2.0 mL of UTCT buffer (0.8 M urea, 0.2 M thiourea, 0.1% CHAPS, 2 mM Tris) in an ultrasonic bath, for seven consecutive cycles of 30 min ultrasonication, followed by 5 min without ultrasound generation to allow the ultrasonic bath temperature to remain below 30 °C, with the addition of ice in the bath if necessary. The total extraction time was 4 h.

Extract supernatants were transferred in 1.5 mL Eppendorfs and centrifuged for 2 × 20 min at 21,000 rcf. After each centrifugation step, the solid phase (remaining filter fragments) was removed. The resulting extract was then concentrated to 0.1 mL under vacuum (Genevac, miVac).

For further analysis of the environmental samples, an in-solution sample preparation procedure was selected, similar to the approach of Piovesana et al., while Liu et al. rather went for in-gel digestion.[48,49] Proteins were precipitated overnight at −20 °C with 5 volumes of glacial 0.1 M ammonium acetate in 100% methanol. After centrifugation at 12,000*rcf* and 4 °C for 15 min, the resulting pellets were washed twice with glacial 0.1 M ammonium acetate in 80% methanol and further dried under vacuum. The protein pellet was resuspended in 100 μL of 50 mM ammonium bicarbonate, and a second round of protein precipitation was performed as previously described. After two rounds of protein precipitation, the final protein pellets were resuspended in 100 μL of 50 mM ammonium bicarbonate and submitted to reduction with 5 mM dithiothreitol (10 min at 95 °C) and alkylation with 10 mM iodoacetamide (20 min at room temperature). Proteins were finally digested via an initial 2 h incubation with 150 ng sequencing-grade trypsin (Promega), followed by an overnight second incubation with 150 ng trypsin. The resulting peptides were centrifuged for 45 min at 12,000*g*, and the supernatant was dried under vacuum.

LC−MS/MS Analysis

The dried tryptic digests obtained from the standard samples as well as the *Escherichia coli* proof-of-concept samples were resuspended in water containing 0.1% (v/v) formic acid (solvent A). The peptide mixtures were analyzed using a U3000-RSLC chromatographic system (Thermo Fisher Scientific, Bremen, Germany) coupled to a TripleTOF 5600 mass spectrometer (AB Sciex, Concord, Canada) operating in positive mode with a nanoelectrospray source. Five microliters of each sample was loaded on a C_{18} precolumn (75 μm ID × 20 mm nanoViper, 3 μm Acclaim PepMap; Thermo Fisher Scientific). After 10 min of desalting and concentration, the precolumn was switched online with the analytical C_{18} column (75 μm ID × 250 mm nanoViper, 3 μm Acclaim PepMap; Thermo Fisher Scientific) equilibrated in solvent A: solvent B (95:5; v/v). Peptides were eluted using a 5%−40% gradient of solvent B (0.1% formic acid in ACN) for 240 min at a flow rate of 300 nL/min. The TripleTOF 5600 system was operated in data-dependent acquisition mode (DDA) with Analyst 1.7 software. Up to 20 of the most intense multiply charged ions (≥+2) were selected for CID fragmentation with a rolling collision energy: this so-called "Top20" method with a constant cycle time of 3.3 s was set to a high-sensitivity mode. A dynamic exclusion time of 10 s was applied during the peak selection process. The mass spectrometric data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository[55] with the data set identifier PXD056038.[56]

The dried tryptic digests obtained from the environmental aerosol samples were resuspended in 10 μL of water containing 0.1% (v/v) formic acid (solvent A). The peptide mixtures were analyzed using an Easy-nanoLC-1000 system coupled to a Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific) operating in positive mode with a nanoelectrospray source. Five microliters of each sample was loaded on a C_{18} precolumn (75 μm ID × 20 mm nanoViper, 3 μm Acclaim PepMap; Thermo Fisher Scientific) at 800 bar in solvent A. After desalting and concentration, the precolumn was switched online with the analytical C_{18} column (75 μm ID × 25 cm nanoViper, 3 μm Acclaim PepMap; Thermo Fisher Scientific) equilibrated in solvent A: solvent B (95:5; v/v). Peptides were eluted at a flow rate of 300 nL/min using a gradient from 5% B to 20% B in 120 min, from 20% B to 32% B in 15 min, from 32% B to 95% B in 1 min, and 95% B to 95% B during 24 min. The Q-Exactive Plus instrument was operated in DDA with Xcalibur software (Thermo Fisher Scientific). Survey MS scans were acquired at a resolution of 70K at 200 m/z (mass range 350−1250), with a maximum injection time at 100 ms and an automatic gain control (AGC) set at 3×10^6. Up to 10 of the most intense multiply charged ions (≥2) were selected for HCD fragmentation, with the normalized collision energy set at 27, at 17.5K resolution, with a maximum injection time at 100 ms and AGC set at 1×10^3. A dynamic exclusion time of 20 s was applied during the peak selection process. The mass spectrometric data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository[55] with the data set identifier PXD056044.[57]

Data Processing

MS data sets generated by the TripleTOF 5600 mass spectrometer for the proof of concept of the extraction method on *E. coli*-spiked samples were searched against the Swiss-Prot database restricted to the *E. coli strain K12/MG1655* taxonomy (release 2020_03, 4392 sequences) using the Mascot algorithm (version 2.8, Matrix Science, London, UK). Peptide modifications allowed during the search were: *N*-acetyl (protein N-term), carbamidomethylation (C), and oxidation (M). Mass tolerances in MS and MS/MS were set to 10 ppm and 0.5 Da, respectively, while the instrument setting was specified as ESI-QUAD-TOF. The +2, +3, and +4 charged peptides were considered. The trypsin/P parameter was used for enzyme specificity with 2 allowed missed cleavages. The resulting .dat Mascot files were then imported into Proline v2.1 package for

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX

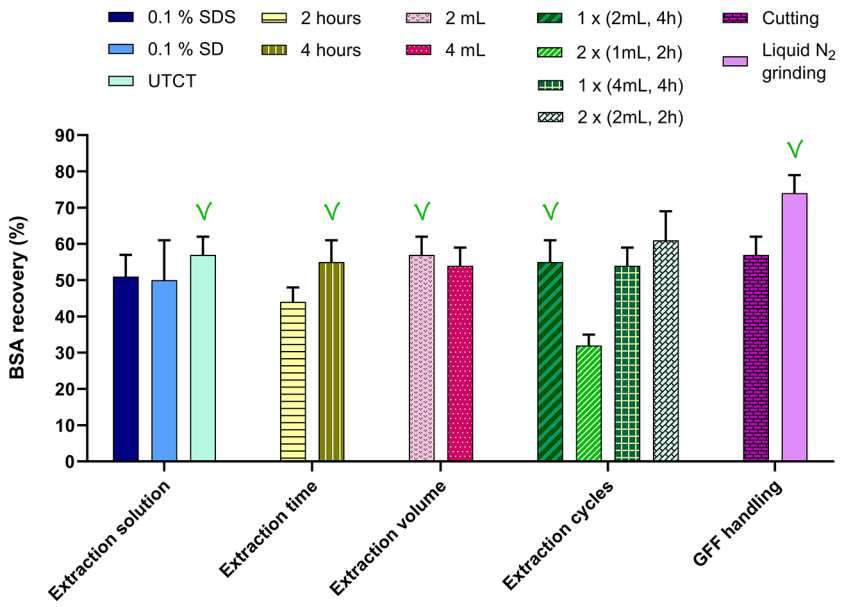


Figure 1. Extraction recoveries of 100 μg bovine serum albumin spiked onto PM-covered GFFs, with different extraction parameters, measured by colorimetric protein quantification assays.

further postprocessing.[58] Proteins were validated if they were identified by at least one proteotypic (unique) peptide, with a Mascot pretty rank equal to 1, a PSM (peptide spectrum match) score above 25, and a 1% FDR (false discovery rate) on both PSM and protein sets (based on the score). Additionally, manual curation of MS/MS fragmentation spectra was performed to enhance the quality of the results, particularly for unique peptides, enabling unambiguous protein identification despite the presence of peptides shared by multiple accession numbers. This manual curation involved identifying a minimum of 5 consecutive amino acids considering either y- or b-ions and attributing the major peaks of a fragmentation spectrum to y- or b-ions.

Environmental bioaerosol data sets generated by the Q-Exactive Plus mass spectrometer were analyzed using the Mascot algorithm (version 2.8, Matrix Science, London, UK) in a two-round searching strategy against the SwissProt database. This approach, widely used now in metaproteomic studies,[59,60] included a first round of database searching based on the entire SwissProt database without any taxonomic restrictions (release 2022_05, 568,744 sequences). The second round then focused on a refined database containing only the protein sequences from taxa identified in the first round (264 taxonomies, 174,777 sequences). Peptide modifications allowed during the search were: *N*-acetyl (protein N-term), carbamidomethylation (C), and oxidation (M). Mass tolerances in MS and MS/MS were set to 10 ppm and 0.02 Da, respectively, whereas the instrument setting was specified as Q-EXACTIVE. The +2, +3, and +4 charged peptides were considered. The trypsin/P parameter was used for enzyme specificity with 2 allowed missed cleavages. The resulting .dat Mascot files were then imported into Proline v2.1 package for further postprocessing.[58] Proteins were validated if they were identified by at least one proteotypic (unique) peptide, with a Mascot pretty rank equal to 1, PSM score above 25, and 1% FDR on both PSM and protein sets (based on the score). These criteria were applied to the raw Mascot results from both search rounds: the first round, conducted without taxonomic restrictions, and the second round, which used a refined database containing 264 taxa.

All the proteins validated at FDR < 1% were then classified using their Swissprot-reported taxonomic lineage into the following categories: Eukaryota−Metazoa, Eukaryota−Viridiplantae, Eukaryota−Fungi, Other Eukaryota, Archae, and Bacteria. When the whole set of peptides identifying a protein was attributed to several accession numbers (so-called same-set proteins), the taxonomic lineage of all the accession numbers was taken into account: if one of the same-set proteins was belonging to the Eukaryota−Metazoa category, then this category was attributed to the reference protein (protein chosen by the Mascot algorithm to be the representative of the same-set accession numbers). Most of the time, all the same-set proteins belonged to the same superkingdom (Eukaryota, Archae, and Bacteria) or kingdom (Metazoa, Viridiplantae, and Fungi) annotation. All the validated peptides from the Swiss-Prot database search were combined into a single list for each of the 12 environmental samples. Proteins confidently identified by nanoLC-MS/MS in the SwissProt database in each kingdom were then analyzed with respect to their functional classification using the PANTHER classification system (http://www.pantherdb.org/).[61] The latest PANTHER 19.0 release includes data from 144 complete genomes encompassing the taxonomies identified in our study. This system was used to retrieve annotations from the gene ontology (GO) knowledge base (biological process, molecular function, and cellular component) as well as the PANTHER protein class and pathway.

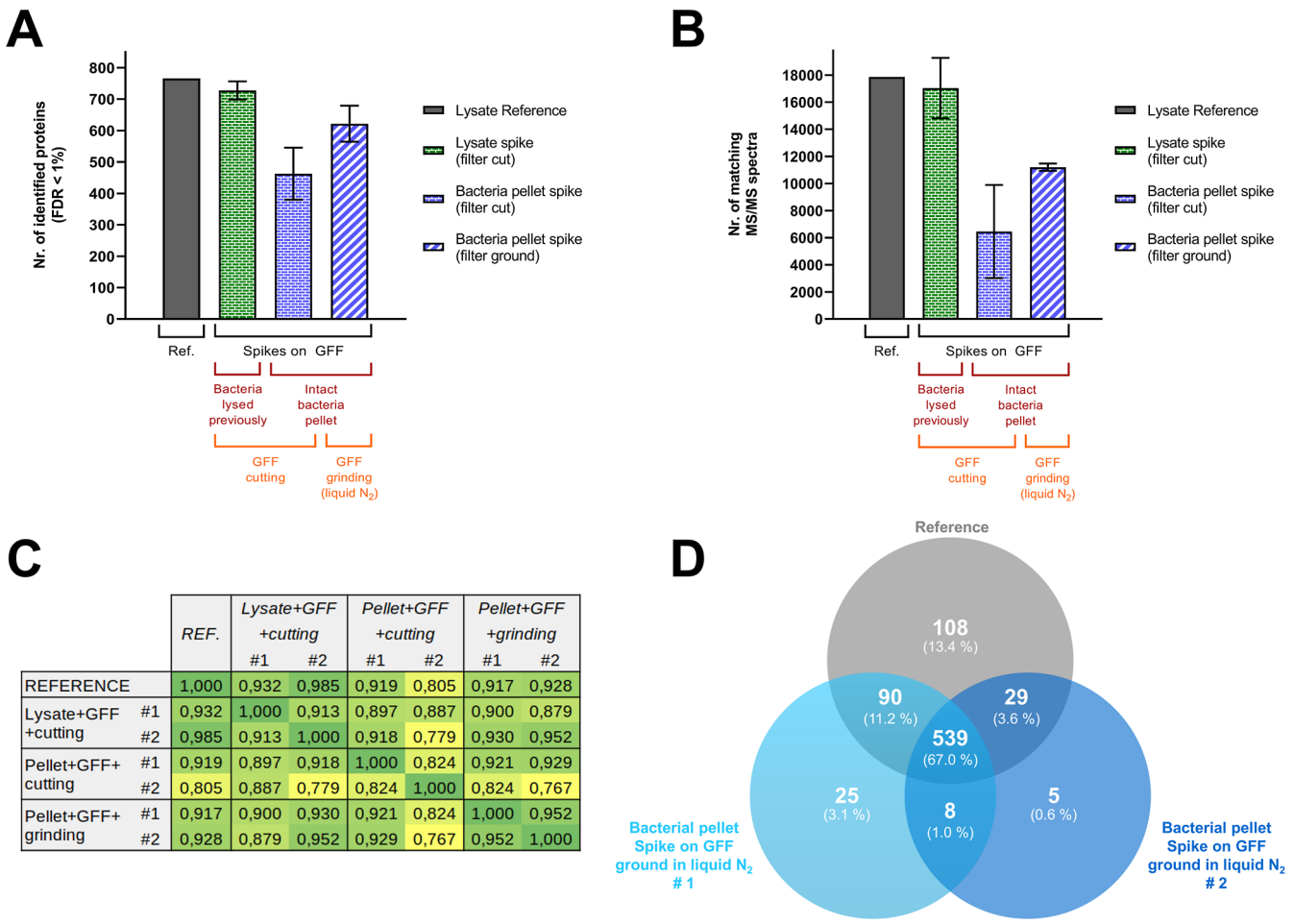
RESULTS AND DISCUSSION

Optimization of the Sample Preparation Workflow

Protein extraction recoveries were assessed quantitatively in this work for various extraction conditions. An illustration of the protocol for the evaluation of the extraction yield is presented in the Supporting Information (Figure S1), and the latter is detailed in the "Sample Preparation" paragraph of the Experimental Section in this manuscript.

A major optimization that was searched in this work was the identification of an optimal extraction buffer for environmental proteins sorbed onto GFFs. An aqueous solution of the anionic

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX



		REF.	Lysate+GFF +cutting		Pellet+GFF +cutting		Pellet+GFF +grinding	
			#1	#2	#1	#2	#1	#2
REFERENCE		1,000	0,932	0,985	0,919	0,805	0,917	0,928
Lysate+GFF +cutting	#1	0,932	1,000	0,913	0,897	0,887	0,900	0,879
	#2	0,985	0,913	1,000	0,918	0,779	0,930	0,952
Pellet+GFF+ cutting	#1	0,919	0,897	0,918	1,000	0,824	0,921	0,929
	#2	0,805	0,887	0,779	0,824	1,000	0,824	0,767
Pellet+GFF+ grinding	#1	0,917	0,900	0,930	0,921	0,824	1,000	0,952
	#2	0,928	0,879	0,952	0,929	0,767	0,952	1,000

Figure 2. Extraction recoveries of *E. coli* spiked onto PM-covered GFFs were compared across various sample preparation protocols against a reference *E. coli* proteome. (A) Total number of *E. coli* proteins identified using the Mascot algorithm, respecting the FDR < 1% threshold. (B) Total number of MS/MS spectra identifying *E. coli* proteins by the Mascot algorithm, meeting the FDR < 1% threshold. (C) Correlation analysis among all 7 samples to evaluate the influence of the starting material (lysate or pellet) and compare the cutting and grinding procedures. (D) Venn diagram comparing the proteins identified between the reference *E. coli* proteome and the GFF spiked with a 1×10^9 *E. coli* pellet and subjected to a liquid nitrogen grinding protocol.

detergent sodium dodecyl sulfate (SDS) at 0.1% and another of its alternative sodium dodecanoate (SD) at the same concentration were evaluated in this study. SD was tested because it has the advantage of not interfering with the Bradford colorimetric assay. A commercial buffer "SDS−Tris−Gly" containing 0.1% SDS, 25 mM Tris, and 192 mM glycine was also involved in this trial, as recommended by Liu et al.[48] However, it was not possible to retrieve results from the latter as it interfered with both colorimetric assays. SDS is known to be incompatible with the Bradford assay but could be evaluated by the BCA assay. However, a shift of color was also observed with "SDS−Tris−Gly" in BCA toward light blue instead of the light green−dark purple calibration range, impairing colorimetric quantification of the extracts obtained from this buffer.

Finally, a more complex buffer was also investigated in this work, based on previous studies in proteomics.[62] It was made up of 0.8 M urea, 0.2 M thiourea (chaotropic agents), 0.1% CHAPS (zwitterionic detergent), and 2 mM Tris (pH buffer), also called UTCT buffer. The concentration of the various components in water was decreased compared with previous works to avoid exceeding the solubility threshold during the vacuum concentration step. An improvement was obtained when using the latter "UTCT" buffer (57 ± 5% extraction yield) instead of 0.1% SDS (51 ± 6%) or 0.1% SD (50 ± 11%) (see Figure 1). The efficiency of this solution was thus confirmed.

The number and duration of the extraction cycles were also evaluated. A significant improvement was observed when the extraction time was increased from 2 h (44 ± 4%) to 4 h (55 ± 6%). On the other hand, when this unique cycle of 4 h extraction in 2 mL solution was split up in two 2 h cycles with 1 mL extraction solution each, BSA recovery decreased from 55 ± 6% to 32 ± 3%, likely due to insufficient extraction volume (see Figure 1).

Therefore, the experiment was repeated with a 4 mL total extraction volume. The two-cycle procedure (2 × 2 h, 2 × 2 mL) gave 61 ± 8% recovery, vs 54 ± 5% for its one-cycle counterpart. The application of two distinct extraction cycles could thus be efficient to recover proteins on GFFs but only with sufficiently high total volumes (≥4 mL). Considering the limited gain in extraction yield when such volumes were used and the unnecessary increase in duration of the extract evaporation, this condition was not selected, and one 4 h extraction cycle in 2 mL of buffer was considered as optimal (see Figure 1).

However, with such conditions, we still failed to obtain more than 60% recovery of BSA. A hypothesis for this shortcoming was the excessive loss of proteins still sorbed on filters after

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX

extraction as well as the amount of extract which was absorbed by the filter and not retrieved afterward. Therefore, the grinding of filters in liquid nitrogen was investigated to increase the exchange surface with the solution while reducing its absorption by the filter, compared to the simple "fragment-cutting" procedure. This approach proved itself very efficient, as it increased BSA recovery from 57 ± 5% to 74 ± 5%, which was our final optimized extraction yield.

Proof of Concept: Application to *E. coli* Spikes on Aerosol Filters

As the method was developed on the commonly accepted standard protein BSA, it was decided to further confirm the applicability of our metaproteomic workflow with an optimal control sample whose matrix components and complexity should match those of the real samples of interest. In this study, we decided to subject a 1×10^9 *E. coli* bacteria pellet to our protocol as a reference material. The methodology employing nanoLC−MS/MS with a TripleTOF 5600 mass spectrometer, searching against an *E. coli* K12 taxonomy-restricted database, and validating under a stringent FDR < 1% threshold utilizing the target-decoy approach, allowed us the establishment of a reference *E. coli* proteome, encompassing 766 proteins identified through a total of 17,866 matching spectra. Subsequently, this reference proteome coverage was compared with other *E. coli* proteomes obtained via the spike and recollection of *E. coli* on PM-covered GFFs, yet using identical nanoLC−MS/MS parameters and Mascot search and validation protocols.

To assess our protocol for the characterization of PBAP in urban PM_{10} samples, we investigated the extraction efficiency of this complex sample by comparing the total number of *E. coli* proteins and spectra identified through nanoLC−MS/MS analysis when *E. coli* lysate (from a 1×10^9 *E. coli* UTCT-based lysis) was directly analyzed, versus when it was spiked on GFF and re-extracted afterward. First trials were performed with the simple approach of GFF fragment-cutting. Following the loading of GFFs with lysate, we successfully retrieved 95 ± 4% of *E. coli* proteins in comparison to the reference proteome (728 vs 766 proteins, see Figure 2A), and 95 ± 12% of the peptide-spectrum matches from the control were recovered (17,034 vs 17,866 spectra, see Figure 2B). In addition to this overall recovery rate, we aligned protein identifications, revealing a correlation indicating that the two *E. coli* proteome replicates correlated at 93.2% and 98.5%, respectively, with the reference bacterial proteome in the absence of any filter (Figure 2C). These findings underscore the capacity of our procedure to efficiently extract bacterial proteins manually spiked on GFF.

However, lysed bacteria do not necessarily reflect an environmental biological material, as bacteria would rather be found intact in bioaerosols. In order to better mimic environmental samples, we endeavored to extract *E. coli* proteins from GFFs spiked with a consistent bacterial pellet of 1×10^9 cells reconstituted with ammonium bicarbonate, preserving the integrity of the starting material, instead of utilizing a lysed bacterial pellet. The extraction of an intact bacterial pellet spiked onto the GFF unsurprisingly yielded fewer proteins, with a recovery of 60 ± 11% (see Figure 2A), as well as fewer spectra (see Figure 2B). Notably, this protocol exhibited a significant variability between two replicates, particularly due to the low recovery rate in the second replicate (only 80.5% correlation with the reference, vs 91.9% for the first replicate; see Figure 2C). Our hypothesis regarding this low recovery rate was that an intact bacterial pellet necessitated a more efficient mechanical extraction procedure than the previously employed cutting-based protocol.

To investigate deeper this hypothesis, we spiked again a 1×10^9 *E. coli* pellet onto two independent GFF filters and subsequently ground them in a mortar with liquid nitrogen to obtain a fine nitrogen powder, as advised by the results obtained in the Optimization of the Sample Preparation Workflow section. The remainder of the protocol was unchanged for the sample preparation and nanoLC−MS/MS analysis. This grinding protocol enabled the recovery of 622 proteins compared to the 766 proteins in the reference proteome (81 ± 7%), as well as 11,199 spectra compared to the 17,866 spectra in the reference proteome (63 ± 2%), which confirmed the major interest of the liquid nitrogen grinding step. The comparison of the protein identifications between the reference proteome and the two replicates from the latter optimized protocol of this work (liquid nitrogen grinding of the bacterial intact pellet) is illustrated in Figure 2D using a Venn diagram. Notably, 82% of identifications were found in both proteomes, while 13% of proteins observed exclusively in the reference proteome consisted primarily of low-abundant proteins, identified on average by 3 spectra. Given the 75% quantitative recovery rate observed with BSA, this proof of concept on an intact *E. coli* bacterial pellet indicated that the extraction procedure was powerful enough for the analysis of more complex samples such as the environmental samples of this work.

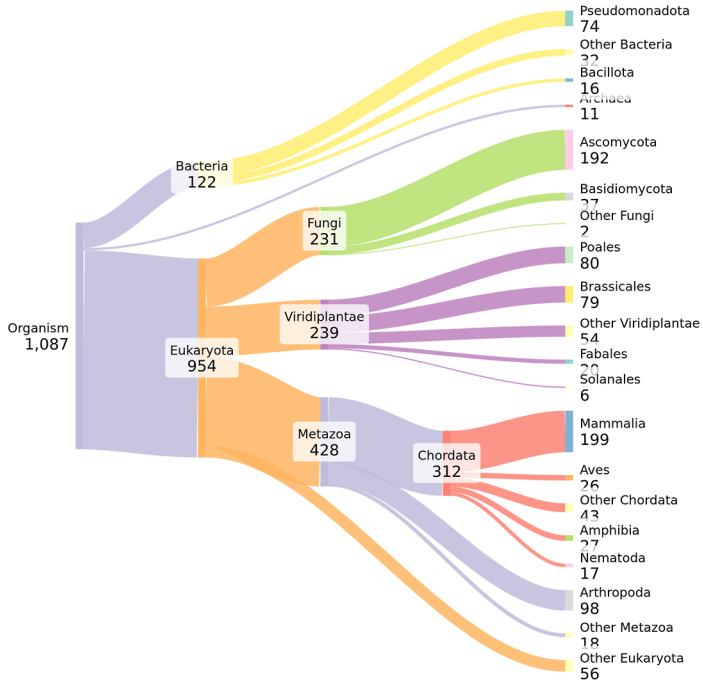
Metaproteomic Analysis of Ambient Outdoor PM_{10} Samples in Strasbourg, France

Global Observations on the Protein Content of Ambient Aerosols. The natural environment can play a significant role in the epidemiology of infectious diseases and allergies, as pathogens and allergens persist and evolve in environmental niches from where they can be transferred to new hosts. The atmospheric compartment and its particulate fraction are one of those environmental niches. It is thus of major importance to investigate the nature and variety of the biogenic elements in this matrix, with specific attention toward allergens and pathogens.

The analysis of the biological components extracted from 12 PM_{10} filters collected weekly on the rooftop of the Botanical Institute in Strasbourg presented significant challenges. These included the limited quantity of proteins available on the filters, necessitating quantitative extraction and sensitive detection by mass spectrometry, and the need to adapt the database search strategy to accommodate the wide variety of ecological systems and proteins that can be found in atmospheric samples but are not known in advance. To address this, we used a two-step database search method. The first step involved searching against a large sequence database based on the complete Swiss-Prot database (release 2022_05), which contained 568,744 sequences with no taxonomic restrictions. In the second step, we conducted a refined search using a smaller database composed of the 264 taxa identified in the initial search. This approach was chosen to mitigate the risk of false positives, which tend to increase when searching against a large database and can only be controlled by applying stringent thresholds. Peptide-spectrum matches (PSMs) and proteins were validated using a stringent FDR threshold of <1%, as recommended by current proteomic guidelines,[63] and manually verified (see Experimental Section—Data Processing).

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX

A



B

Taxonomy					S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Archae					3	2	3			1	2	2		1	2	
Bacteria	Bacillota				3	3	3	2	3	2	1	2		1	5	
	Pseudomonadota				9	3	9	6	25	18	7	8	3	8	30	
	Other Bacteria				7	3	10	1	12	6		1	1		2	
Eukaryota	Fungi	Ascomycota			24	26	40	43	54	56	26	17	10	21	103	34
		Basidiomycota			6	2	15	17	16	12	2	4	1	4	28	17
		Other Fungi					1		1					1		
	Metazoa	Arthropoda			38	26	33	10	27	21	19	17	14	28	57	12
		Chordata	Amphibia		15	11	13	3	9	6	4	6	5	6	11	3
			Aves		3		2	7	8	2	1	4		6	14	3
			Mammalia	Artiodactyla	38	16	32	14	40	27	4	14	22	10	20	9
				Rodentia	8	6	4	1	6	3		6	15	3	5	
				Other Mammalia	16	7	10	4	6	5		4	1	4	1	
			Other Chordata		11	12	17	10	13	12	3	3	6	14	27	13
		Nematoda								2	3	4		1	4	
		Other Metazoa			2	2	7	4	6	5		1		6	12	8
	Viridiplantae	Chlorophyta			2	2	4	3	8	8	1	1		2	7	6
		Streptophyta			52	37	63	43	93	78	26	36	39	32	89	44
	Other Eukaryota				15	15	22	15	21	19	12	10	4	7	33	16
Total number of peptides					**252**	**173**	**288**	**183**	**348**	**283**	**111**	**140**	**121**	**155**	**450**	**165**

Figure 3. Metaproteomic analysis conducted on 12 environmental samples collected weekly over a span of 12 consecutive weeks from the rooftop of the Botanical Institute in Strasbourg. (A) Sankey diagram representing the taxonomic distribution of the 1,087 peptides corresponding to Swiss-Prot accession numbers identified when combining all 12 samples. (B) Heat-map representation illustrating the distribution of the 1,087 distinct peptides across the 12 week duration of data collection, spanning the taxonomic diversity observed.

Respecting the FDR < 1% threshold and employing manual curation, a total of 1,087 peptides, corresponding to 447 nonredundant Swiss-Prot entries, were detected across the whole sample set. The complete taxonomic data of this data set are presented in Table S1: proteins were assigned to the superkingdoms of Eukaryota, Bacteria, and Archae. Each peptide resulting in the identified proteins from Table S1 is detailed in Table S2. Figure 3A presents a Sankey diagram depicting the taxonomic distribution of this comprehensive peptide set, whereas Figure S2 illustrates the taxonomic distribution at the protein level. In this perspective, we chose to report primarily the identified peptide set to provide an overview of the taxonomies present in our ambient PM_{10} samples. Indeed, in such a complex environmental proteome, closely related proteins sharing very similar peptide sequences can coexist, and their distinction throughout the protein inference step has been described as a major pitfall in bottom-up metaproteomics which we wished to circumvent.[46,64]

The major part of the peptides (954/1,087, 87%) were attributed to sequences with at least one eukaryotic correspondence. As some peptides are highly conserved among all organisms, when several protein entries were all identified by a similar set of peptides (so-called same-set proteins) and when one of these same-set proteins was

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX

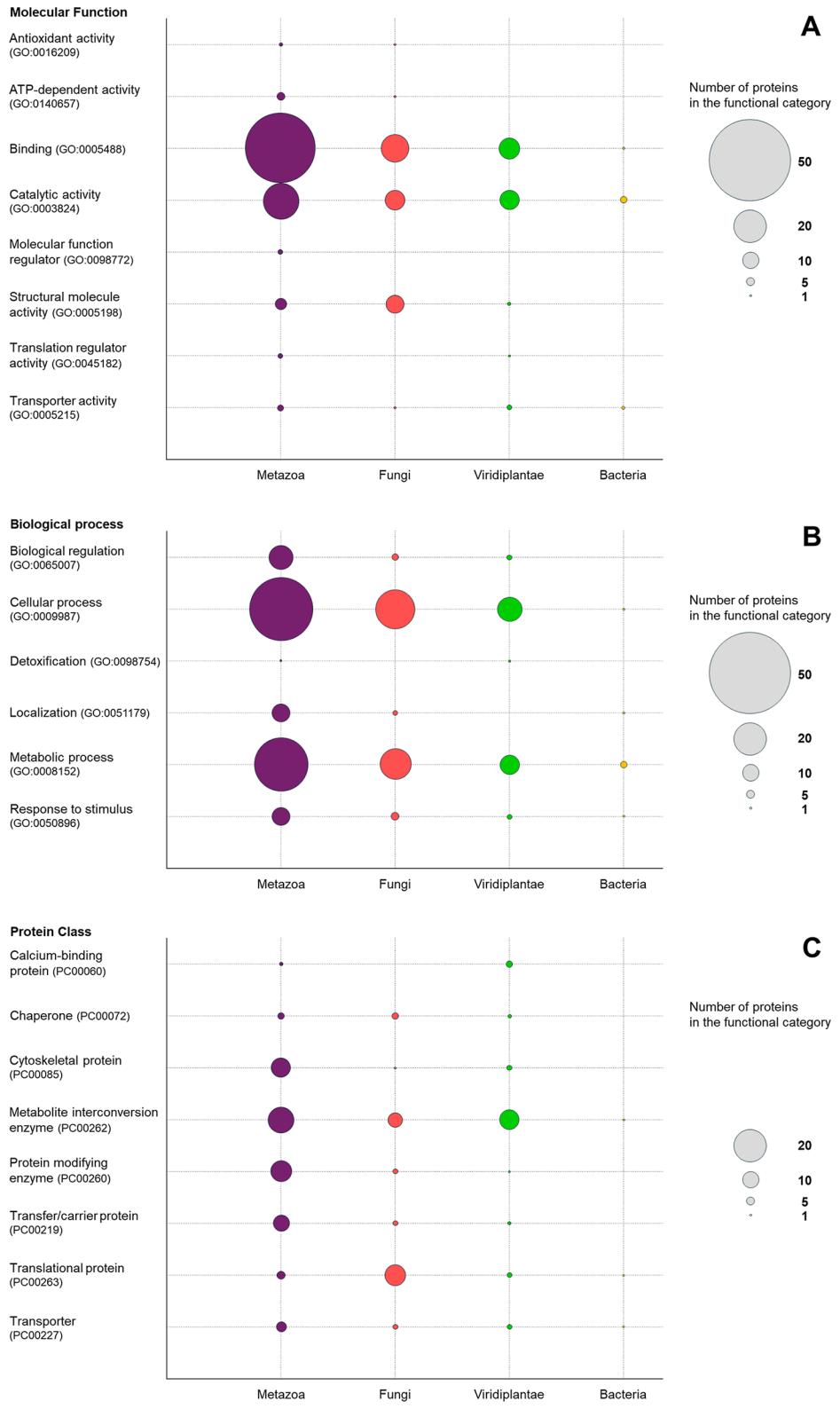


Figure 4. Functional analysis of the metaproteomic data set derived from the analysis of 12 weekly samples of outdoor PM_{10}. Bubble plots characterize the partition of proteins detected via at least one unique peptide across (A) GO molecular functions, (B) GO biological processes, and (C) PANTHER protein classes. The size of the circles is proportional to the number of entries associated with each functional category for each kingdom.

eukaryotic, we then decided to specifically choose the eukaryotic same-set protein as the representative protein from the protein set (36/954 eukaryotic peptides).

Figure 3B provides a heat-map illustration of the distribution of the 1,087 distinct peptides over the 12 week duration of data collection, showcasing the observed taxonomic diversity across this period. It is noteworthy that some weeks returned a considerably higher number of peptides (W5, W6, and W11) compared to other weeks (W7 to W9). This discrepancy might be due to different meteorological conditions, which are

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX

summarized in Table S3: W7 and W9 were the rainiest weeks of our sampling campaign (45.2 and 55.7 mm of cumulative rainfall, respectively), whereas W5, W6, and W11 were among the driest (6.1 ± 5.3 mm weekly rainfall averaged over these 3 weeks). While PBAP could accumulate during W5, W6, and W11 and was thus successfully sampled, it may be assumed that, oppositely, wash-out of PM_{10} and its proteinaceous material occurred during the rainiest weeks. Consistent with our observation, Petersson Sjögren et al. reported a higher abundance of bioaerosols with a dryer weather.[65] However, in their study, the increased amount of bioaerosol material was also associated with warmer weather, which is not verified here because temperature and humidity could not be fully resolved during these 12 weeks. These results suggest that there may be some interesting relationships between the total amount of aerosolized proteins and the meteorological conditions in aerobiological studies.

Metaproteomics Reveal the Presence of Specific Species in Bioaerosols. *Metazoa.* Among the 954 peptides derived from Eukaryota, the majority were associated with the kingdom Metazoa, comprising 428 peptides. Upon excluding peptides assigned to *Homo sapiens* and farm animals, those validated with Mascot at FDR < 1% predominantly belonged to the classes *Insecta* (83), *Rodentia* (48), *Amphibia* (27), *Nematoda* (17), and *Branchiopoda* (11). The prevalence of these five classes is in line with the expected composition of Metazoa species in outdoor environmental samples.

According to a functional analysis based on the GO annotations, the proteins originating from the presence of animals were mostly involved in cellular and metabolic processes. Most of them were either binding proteins or proteins with catalytic activity. On the other hand, they were shared between many protein classes, including metabolite interconversion enzymes and protein-modifying enzymes responsible for biochemical reaction catalysis, but also transfer proteins critical for lipid transport, and cytoskeletal proteins maintaining cellular anatomical integrity. These results are summarized in Figure 4. They are expected, as these classes and functions correspond to some of the most abundant proteins in animal cells and thus the ones with the highest probability of detection in such a metaproteomic study.

Insect peptides were particularly abundant, with species such as *Drosophila*, *Manduca,* and *Bombyx*. These organisms hold significant value for studying environmental adaptation, given their global distribution and their status as model organisms in health and environmental pollution assessment, including pesticides, drugs, and heavy metals.[66]

All *Amphibia* peptide occurrences originated from *Xenopus,* which is a genus with populations established on multiple continents and is now considered as one of the major invasive amphibians in the world.[67] Numerous studies using eDNA methodologies and qPCR have identified this amphibian in collected freshwater samples and aquatic environments.[67] In this work, it was identified in all 12 samples. Notably, *Xenopus laevis* is a known model organism, hence its presence in the Swiss-Prot database.

The 48 rodent peptides were attributed to the genera *Mus* (32), *Rattus* (12), and *Cricetulus* (4).

Additionally, 18 peptides were associated with *Caenorhabditis elegans* which has widely been regarded as a suitable ecological indicator for monitoring water and soil quality and which could represent a good marker for the presence of *Nematoda* species.[68]

Fungi. In this data set, Fungi emerges as the second most prevalent kingdom within Eukaryota, comprising 231 peptides and accounting for up to 33% of all peptides in week 4. Fungal spores and mycelium fragments, which persist airborne, are discernible components in aerobiological studies and constitute a major focus of many studies in the field due to their potential allergenicity and pathogenicity.[12,14−16] Notably, they are detectable in environmental DNA samples through standard "metabarcoding" methodologies.[14,35]

The functional analysis of fungal proteins differed slightly from the other kingdoms of this data set, as translational proteins were much more represented among the detected proteins. Additionally, proteins with structural molecule activity were also more abundant than what was observed for other kingdoms (see Figure 4).

The peptides identified in the bioaerosols from this study (Table S2) indicate the presence of two prominent fungal families. The Ascomycota phylum is by far the most represented, accounting for 191 out of the 231 peptides, with an overrepresentation of the Saccharomycetes, Sordariomycetes, and Eurotiomycetes classes. Basidiomycota is the second major phylum in this data set, with the identification of 37 additional peptides, mostly corresponding to the Agaricomycetes class. This is in line with the findings of Fröhlich-Nowoisky et al., who also reported these phyla as predominant within the Fungi kingdom in ambient aerosols.[14]

Interestingly, about a quarter of the total fungal content was attributed to *Saccharomyces*, a significant microbial species in human history.[69] Other ubiquitous airborne fungi were identified in this data set, consistent with their known environmental occurrence, including *Penicillium* subspecies,[70] along with fungi indicative of environmental health in urban soils (such as the three edible mushrooms *Agaricus*,[71] *Coprinus*,[72] and *Schizophyllum*[73]), aboveground habitats (*Davidiella*[74]), and aquatic environments (*Malassezia*[75]).

Additionally, filamentous fungi, such as *Aspergillus fumigatus*,[16,76] *Thermomyces*,[77] *Thermothelomyces*,[78] and *Paxillus*,[79] were also characterized by several peptides in this study. These fungi specifically inhabit plant hotspots, functioning as reservoirs and participating in organic matter decomposition through specialized enzymes. Other fungi known to colonize various host plants have also been identified, such as *Neurospora*[80] and *Ustilago*.[81] Noteworthy is the identification of peptides attributed to the *Botryotinia* genus, a model for molecular studies on necrotrophic fungi, notorious as an airborne plant pathogen affecting over 200 crop hosts worldwide.[82]

Given that many fungi produce mycotoxins and their mycelium fragments and spores are potential allergens, monitoring their presence is crucial. In this study, several peptides were assigned to fungi that are recognized as agents of human (*Arthroderma*,[83] *Coccidioides*,[84] and *Encephalitozoon*[85]) and animal (*Emericella*[86]) diseases. Furthermore, the two *Candida* species (*Candida albicans* and *Candida glabrata*) detected in our ambient aerosol samples are common aerobiological agents of candidiasis, ranging from superficial conditions like oral thrush to life-threatening invasive diseases.[87] Although the related proteins from this data set were mostly related to "housekeeping" genes and not to pathogenic functions, their detection indicates the potential presence of the previously cited pathogenic taxa.

On the other hand, Table S4 features a list of eight known allergenic proteins detected in this study that confirms possible

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 1. Twenty-Three Proteins Associated to Plant Allergens Have Been Identified, Categorized into Four Superfamilies: Cupin, Prolamin, Polcalcin, and Plant Defense System Superfamilies[a]

type of allergens (superfamily)	accession #	taxonomy	common name	peptides
		Cupins		
Vicilins	VCLC_PEA	Pisum sativum	garden pea	[115−121] NSFNLER
				[146−155] VLDLAIPVNR
				[216−226] QQSQEENVIVK
				[246−257] SVSSESEPFNLR
				[415−427] ELAFPGSAQEVDR
Legumins	LEGA2_PEA	Pisum sativum	garden pea	[436−448] ALTVPQNYAVAAK
	LEGB2_VICFA	Vicia faba	broad bean	[175−181] ADLYNPR
				[201−207] LSAEYVR
		Prolamins		
a-amylase/trypsin inhibitors	IAAB_HORVU	Hordeum vulgare	barley	[46−54] DYVEQQACR
				[92−107] SRPDQSGLMELPGCPR
				[108−115] EVQMDFVR
	IAA1_WHEAT	Triticum aestivum	wheat	[26−39] LQCNGSQVPEAVLR
	IAAC2_WHEAT	Triticum aestivum	wheat	[77−86] ELYDASQHCR
				[97−106] TSDPNSGVLK
	IAAC3_WHEAT	Triticum aestivum	wheat	[45−60] DYVLQQTCGTFTPGSK
				[101−115] YFIALPVPSQPVDPR
				[116−132] SGNVGESGLIDLPGCPR
cereal prolamins	GLUA1_ORYSJ	Oryza sativa	rice	[51−58] LQAFEPIR
	GLU2_MAIZE	Zea mays	maize	[153−160] QQCCQQLR
	GLT0_WHEAT	Triticum aestivum	wheat	[34−44] ELQESSLEACR
				[45−54] QVVDQQLAGR
				[625−637] AQQPATQLPTVCR
2S albumins	CONB7_LUPAN	Lupinus angustifolius	blue lupine	[552−564] ELTFPGSAQDVER
nsLTPs (nonspecific lipid-transfer proteins)	NLTP1_WHEAT	Triticum aestivum	wheat	[80−90] GIHNLNEDNAR
	HPSE_SOYBN	Glycine max	soybean	[70−79] SYPSNATCPR
		Polcalcins		
seed storage proteins (globulins)	POLC7_PHLPR	Phleum pratense	common timothy	[20−30] ISLSELTDALR
				[31−42] TLGSTSADEVQR
	POLC7_CYNDA	Cynodon dactylon	Bermuda grass	[22−32] ISLAELTDALR
				[33−44] TLGSTSADEVQR
	POLC3_CHEAL	Chenopodium album	fat hen	[28−38] ISSSELGDALK
seed storage proteins (avenins)	AVLA1_WHEAT	Triticum aestivum	wheat	[77−91] QQCCQPLAQISEQAR
		Plant Defense System		
Prs (pathogenesis-related proteins)	PR4B_TOBAC	Nicotiana tabacum	common tobacco	[92−104] VTNTGTGTQATVR
	BEV1A_BETPN	Betula pendula	European white birch	[147−160] AVESYLLAHSDAYN
	MPA5A_LOLPR	Lolium perenne	perennial ryegrass	[59−68] LLEDVNAGFK
	MPAG3_DACGL	Dactylis glomerata	cock's-foot grass	[76−87] NVFDEVIPTAFK
	PER36_ARATH	Arabidopsis thaliana	mouse-ear cress	[320−334] MGNISPLTGTDGEIR
proteases	MUCIN_MUCPR	Mucuna pruriens	velvet bean	[226−241] AVANQPVSVAIEGGGR

[a]The table presents the accession numbers in the SwissProt database, along with the scientific taxonomy, the common name of the respective plant as well as the peptides identified, including their relative position on each protein.

day-to-day human exposure to those allergens. They were retrieved from the Allergome database[88] and are considered as significant fungal allergens, originating from *Aspergillus*,[16,76] *Penicillium*,[15,89] *Cladosporium*,[15,17] and *Alternaria*[15,17,90] organisms. Daily airborne *Alternaria* spp. and *Cladosporium* spp. spore concentration forecasts may be used as advisory for allergy sufferers and asthmatics,[17] and the previously mentioned detection of *A. fumigatus* is particularly critical considering its wide adverse effects on humans, from mild allergies to lethal pathologies, its growing resistance to antifungal products, and its sharpened effects with concomitant atmospheric pollution.[15,16,76]

Given that the collection filters were obtained from a botanical institute's rooftop, the elevated prevalence of fungal peptides aligns with fungi's association with neighboring soil and atmospheric plant environments. This shows the relevance of implementing metaproteomic workflows to detect the presence of plant and human pathogens among fungi and follow their geographical and seasonal trends. The variability and presence of taxa in aerobiological samples can be heavily influenced by changing climatic conditions, with fungal growth occurring whenever water is available, regardless of the relative humidity.

Viridiplantae. The third highest detection rate was attributed to Viridiplantae peptides (239 peptides, from 21% to 32% of the total peptidic content over the weeks), which aligns with the location of the filters next to the botanical garden in Strasbourg.

The Viridiplantae subdata set includes peptides from cultivated crops, with 83 peptides identified for cereals and 35 for vegetables. Additionally, peptides from annual and perennial

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX

grasses account for 81 entries, followed by green algae (21 peptides), trees (13), and ornamental flowers (6).

Regarding cultivated crops, proteotypic peptides allowed the unambiguous identification of several cereals (including rice, wheat, barley, and maize) and vegetables (including pea, carrot, tomato, beet, and bean). The detection of abundant annual and perennial grasses was expected, given their propensity to grow on rooftops, exemplified by the well-known thale cress *Arabidopsis*, a weed also found along roadsides. Other grass species identified include *Phleum* (timothy grass), *Chenopodium*, and succulents and invasive species from the bluegrass family (*Dactylis* and *Lolium*).

Peptides from green algae were also identified. Species such as *Chlamydomonas* and *Mougeotia*, the common components of freshwater aquatic habitats worldwide, were among those detected.

Peptides from trees were of particular interest due to their potential allergenic properties. Thirteen peptides were identified from widely cultivated trees such as conifers (*Pinus*) and birch (*Betula*), which are responsible for many allergies worldwide.[31,91] The major pollen allergen Bet v 1 was identified in week 1, when birch pollen alert was the most acute (see Table S3), thus confirming metaproteomic results as consistent with other sources of information on aerobiology. The remaining peptides were attributed to ornamental flowers, including bignonias (*Pandorea*) and the buttercup family Ranunculaceae (*Nigella*).

In addition to the examination of the taxonomical distribution of peptides within the Viridiplantae subset, the molecular function of the identified proteins was also investigated. Most of plant-derived proteins in this study were binding or catalytic activity proteins and were involved in cellular and metabolic processes (see Figure 4). Apart from the basic and abundant proteins such as Rubisco, housekeeping proteins, ribosomal proteins, metabolism-related proteins, and nonspecific lipid-transfer proteins, several specific enzymes were identified. Among these, three were implicated in cell wall regulation under pathogen attack (a hydrolase/glycosidase, a xylosidase/arabinosidase, and a pectin acetylesterase). Additionally, a peroxidase from *Arabidopsis* was identified and already described as playing a role in lignin degradation.[92]

Aside from aeroallergens, prominent allergens are also derived from ingested seed sources. Considering the localization of our study, the detection of allergenic seed storage proteins on PM_{10} filters over the 12 week data collection period was expected. These proteins are composed of two primary classes: globulins and avenins.[93] The latter, exemplified in this data set by wheat avenin, was already observed in developing grains and shown to be associated with celiac disease.[93,94] Inside the globulin class, our analysis revealed one pea vicilin, two pea legumins, three rice and maize glutelins, and two wheat glutenins and one lupin conglutin. Prior studies have highlighted the allergenic potential of these globulins, particularly in humans.[95−97] The species associated with these seed storage proteins hold a significant presence in the food market, warranting increased monitoring for allergenic responses.

Furthermore, three polcalcins (*Phleum*,[98] *Cynodon*,[99] and *Chenopodium*[100]) were identified. They are known for their involvement in type I allergy, affecting more than 25% of the world's population. This allergy manifests through the binding of IgE antibodies in sensitized pollen-allergic patients, leading to allergic rhinitis, conjunctivitis, dermatitis, and asthma. Three group-5 allergens corresponding to *Betula* (white birch),[101] *Lolium* (ryegrass),[102] and *Dactylis* (Cock's foot grass)[103] were also detected. This class represents the most potent grass pollen allergens recognized by more than 80% of grass pollen-allergic patients.[102] Furthermore, other allergens extracted from the PM_{10} filters included a hydrophobic seed protein from soybean (associated with asthma in individuals allergic to soybean dust), a mucunain from velvet bean (causing itching upon contact with pods), and five alpha-amylase inhibitors from wheat and barley (implicated in baker's asthma). All these allergenic plant proteins are reported with their corresponding identified peptides in Table 1.

Bacteria. A smaller proportion of peptides were assigned to the Bacteria superkingdom (122 peptides). The majority of these peptides corresponded to nonpathogenic bacteria (88), with 57 of them associated with species typically found in freshwater ecosystems, predominantly belonging to the phylum *Pseudomonadota* (e.g., *Erythrobacter*[104] and *Magnetococcus*[105]), as well as in contaminated aquatic habitats. Additionally, 8 peptides out of 85 were linked to species previously isolated from various types of soils and sediments, primarily within the *Bacillota* phylum. The 23 remaining bacterial peptides were associated with proteins observed in symbiosis with plants (such as *Bradyrhizobium*, *Azorhizobium*, and *Allorhizobium*) and insects.

In addition to these nonpathogenic species, 34 identified peptides were matched with pathogenic bacteria known to be implicated in common health issues through food poisoning (15 peptides),[106] transmission by insects or animals (11 peptides), or via water and soil (8 peptides). Among the bacterial pathogens affecting humans, several matched species capable of entering food processing environments, such as those from the *Pseudomonata* (*Escherichia*, *Pseudomonas*, and *Rickettsia*) and *Bacillota* (*Staphylococcus*, *Listeria*) orders. Other identified pathogens can affect humans through transmission by insects (e.g., *Orientia* via mite larvae[107] and *Borreliella* via ticks[108]), while some bacterial pathogens have been described as causing zoonotic infections in humans (*Anaplasma*[109] d an*Pasteurella*[110]). The remaining bacterial peptides identified in this data set and known to affect human and animal health are commonly observed in water, wastewater, and soil, such as *Mycobacterium*,[111] *Francisella*,[112] and *Bacteroides*.[113]

To note, the very low overlap between proteins identified in this metaproteomic study and proteins with reported GO in Panther precludes a comprehensive functional evaluation of bacterial proteins from this study (see Figure 4). However, considering the observed taxa, these findings support the relevance of metaproteomics to detect the presence of genera that are known airborne human pathogens of environmental origin, although future works are necessary to refine the nature of bacterial proteins released in the airborne compartment of the environment.

Archae. In addition to Metazoa, Fungi, Viridiplantae, and Bacteria, the remaining identified peptides were assigned to the Archae superkingdom (11 peptides out of 992). All identified Archae species were previously isolated in aquatic environments like wastewater and sometimes even extreme environmental conditions (like *Pyrococcus*[114] and *Methanoculleus*[115]). The identified proteins were predominantly involved in primary metabolic processes.

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX

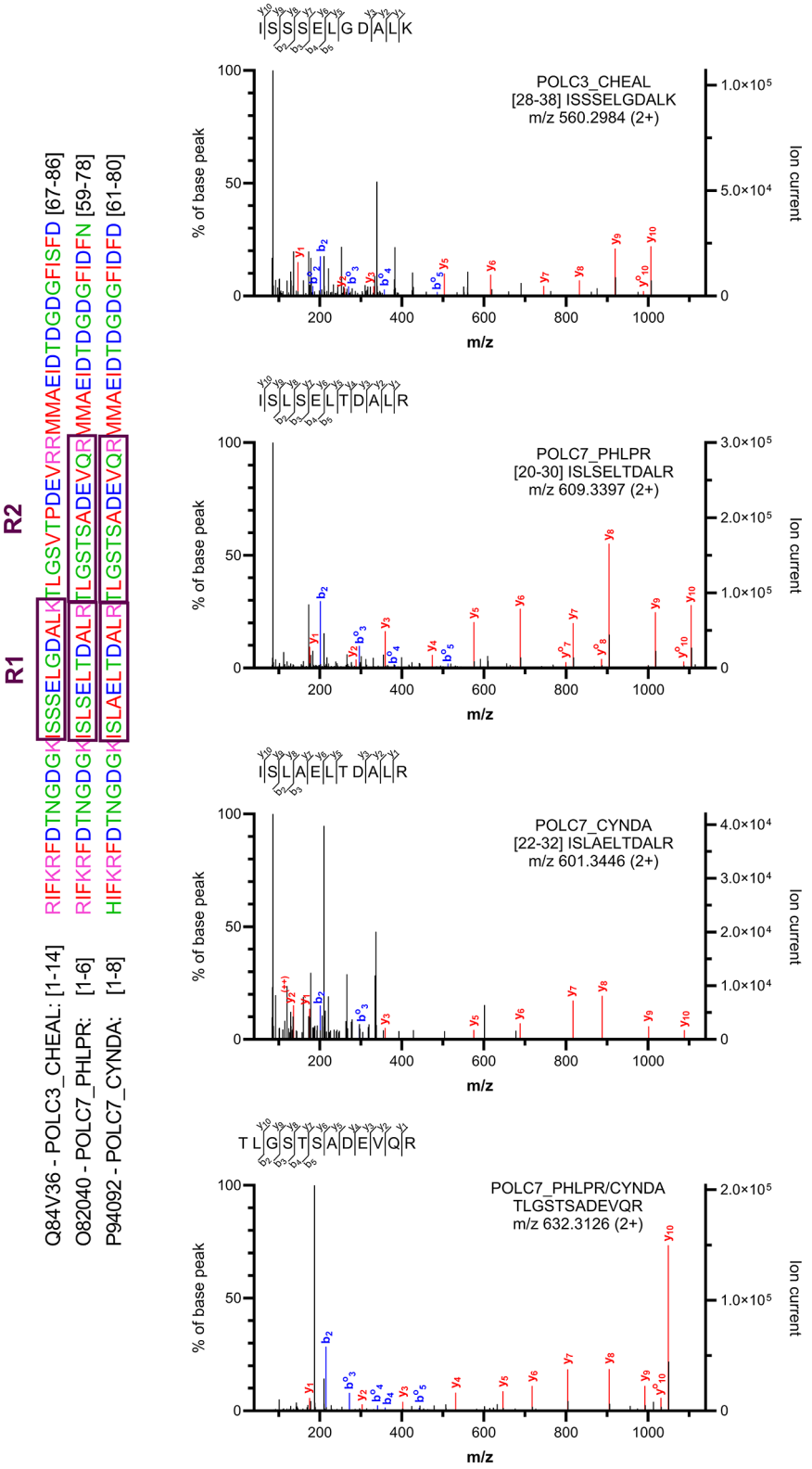


Figure 5. MS/MS spectra from polcalcin-specific peptides. The polcalcin superfamily, represented by the three proteins identified in this study, exhibited distinctive characteristics. However, the tryptic peptides from region 2 (R2) did not facilitate the unambiguous identification of the taxonomy for polcalcin 7. Conversely, the three proteotypic peptides from region 1 (R1) enabled differentiation between polcalcin 3 and 7, as well as distinction between the Cynodon and the Phleum genera for polcalcin 7.

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX

Perspectives on Metaproteomic Potential to Reveal the Composition of Bioaerosols

This work demonstrated the applicability of a metaproteomic approach to decipher the complex biological composition of ambient aerosols and is among the very first studies to do so.[39,48,49] Thanks to this approach, with a single untargeted extraction and LC−MS/MS analysis method, peptides were identified across a very wide range of taxonomies. As described in the previous section, peptide hits in the environmental samples of this work were consistent with previous findings and with the localization of the sampling site close to an urban botanical garden.

Among the previous works on aerosol metaproteomics, only Liu et al. really described ambient aerosols,[48] while Piovesana et al. and Meyer et al. rather studied specific work environments.[39,49] An average of 223 ± 104 peptides were identified in each weekly sample of the hereby-presented study. This corresponds to 46 ± 21 nonhuman proteins identified with at least two peptides across the different samples, which can be compared with the 33 proteins detected by Liu et al. in their ambient aerosol samples, also based on at least two peptides.[48] It is to be noted in this comparison that no size fractionation of the collected aerosols was performed here and that the air sampling volume was smaller (about 3 times less), reinforcing the comparative performance of the hereby-presented analytical method. A deeper evaluation of the identities of proteins detected by Liu et al. shows that in their study, there were more proteins attributed to the kingdom of plants (20/33, 61%) than in ours (from 20% to 47% of proteins attributed to Viridiplantae species, 28% on average).

Naturally, a few challenges remain for a wider use of environmental proteomics in the field of atmospheric science. First and foremost, proteomic workflows rely on nanoLC and HRMS instruments that are expensive and not available in many environmental chemistry laboratories and on experienced scientists to perform these analyses that are not common for environmental chemists. Therefore, interdisciplinary collaborations are greatly encouraged around this topic to associate expertise on environmental sampling and proteomic analyses. Nevertheless, the low throughput of bottom-up proteomic analyses remains inevitable. The sample preparation procedure presented in this work, including a 4 h extraction, freezing steps for an efficient protein precipitation, and an overnight trypsin incubation among others, has a total duration of about 2 days with not more than 10−15 samples processed simultaneously. Then, the LC−MS run time was in the frame of 2−3 h per sample in this work, and long database searches are to be anticipated depending on the search parameters applied (generally several hours per sample).

A second major pitfall of atmospheric metaproteomics resides in the low quantities of proteins collected in airborne PM, resulting in very low sequence coverages due to the abundance of peptides, thus causing difficulties in the identification process. FDR based on decoy databases would remain the gold standard to ensure sufficient quality of identification, but its ideal threshold value (from 1% to 5%) can be discussed. Minimum numbers of (unique) peptides are even more debatable, considering the considerable loss of identifications that we observed between using only one peptide (111 ± 40 proteins per sample) or at least two peptides (46 ± 21 proteins per sample). In this data set, all validated proteins contain at least one proteotypic peptide, with possible additional peptides shared among several taxonomies. Combining a stringent FDR with a tolerance toward proteins identified through one peptide only could enable an extended characterization of the proteome of bioaerosols, yet with a duty to keep a critical eye on these identifications.

This could be ensured by performing manual curation of peptide-spectrum matches, which was performed here and which we highly recommend for all the proteotypic peptides to unambiguously validate the presence of homologous proteins assigned to different taxonomies in the sample. Figure 5 illustrates this aspect, focusing on three different proteins from the EF-hand calcium-binding allergen family. The sequence alignment (Figure 5A) between polcalcins Che a 3 (Chenopodium), Phl p 7 (Phleum), and Cyn d 7 (Cynodon) reveals conserved regions. However, four peptides have been sequenced over two regions: the peptide covering the R2 region clearly validates the p7 polcalcin family, but only the R1 region discriminates between the Phleum and the Cynodon taxonomies. Remarkably, the R1 region exhibits a tryptic peptide composed of 11 amino acids, and the MS/MS fragmentation spectra (Figure 5B) confirm the presence of the three specific sequences, thus validating the presence of the three allergens. The difficulty in this example is amplified by the size of polcalcins. Such small proteins may generate only a limited number of detectable tryptic peptides. Food allergens, for instance, are proteins with a lower molecular weight compared to the average molecular weight of the whole protein content. A more complex case study is illustrated in Figure S3, showcasing different isoforms of the ATI family (amylase and trypsin inhibitors), a group of proteins present in the seeds of all cereals. Manual inspection of the MS/MS spectra, following the automatic validation of the data set at FDR < 1%, clearly confirms the detection of three wheat isoforms and one barley isoform.

Additionally, the choice of the database and search settings is critical in metaproteomic studies. Due to the lack of studies in atmospheric proteomics, a targeted database for this matrix is currently not available. In this study and in line with previous ones,[48,49] the choice was made to perform the search against the most generic database, SwissProt. This strategy enables to screen the global proteome but lacks from sensitivity and specificity.[46] Furthermore, many species are still not sequenced in SwissProt. Therefore, taxonomic annotation can be hampered by these missing species and homologous protein sequences across different species. The construction of more suitable databases over time will be key to the extension of metaproteomics for the characterization of bioaerosols. Similarly, functional annotation of environmental proteins is also currently hampered by the shortage of GO information in publicly available tools, such as the PANTHER classification system (144 genomes for the latest release).

Lastly, metaproteomic studies face a lack of a quantitative readout due to previously discussed challenges, particularly the small amount of material. First, the number of peptides per protein is often lower compared with the common proteomic studies, making it troublesome to adhere to the recommendation of protein quantification based on a minimum of three peptides. Second, as environmental samples are typically composed of proteins assigned to a wide variety of taxonomies, there is an increased probability of detecting shared peptides, which must be excluded from the overall protein quantification. However, once proteotypic peptides have been characterized by nanoLC−MS/MS and assigned to allergenic proteins, a new arsenal of mass-spectrometry-based techniques, such as MRM

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX

assays or DIA approaches, will be applicable to design assays for the precise quantification of these allergens in ambient air.

CONCLUSIONS

This work describes a successful analytical workflow for the analysis of proteins in PM_{10} samples. The optimization of the sample preparation procedure revealed that UTCT was an efficient buffer for the extraction of proteins from GFFs. This was followed by in-solution digestion and a common bottom-up nanoLC-HRMS/MS strategy.

A total of 1,087 peptides were detected in this work, shared among animals, plants, fungi, bacteria, and archaea kingdoms. Our dataset revealed the presence of several allergenic proteins, mostly from plants and fungi. Furthermore, species of interest due to their known pathogenic or allergenic potency were identified via some of their proteins in this study, even if these proteins were not directly linked to pathogenic activity. This emphasizes the potential of metaproteomics to comprehensively investigate the composition of bioaerosols and also the further steps that need to be taken in terms of methodology, technical advances, and bioinformatic data processing to identify more proteins of interest. This approach might help to retrieve spatial and temporal information on the airborne occurrence of species of particular interest at a taxonomic level and to better understand the release of specific proteins (e.g., allergens) at a functional level.

For a broader use of metaproteomics in atmospheric chemistry, standardized methods must be set, including guidelines for high-quality reporting of peptides and proteins, and appropriate protein databases and data processing pipelines need to be specifically developed for such environmental samples.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.4c00936.

Schematic description of the extraction workflow; taxonomic distribution in the data set at the protein level; and MS/MS spectra for the distinction of isoforms (PDF)

Details of identified proteins; details of identified peptides; meteorological parameters during the study; and list of detected fungal proteins with allergenic properties (XLSX)

AUTHOR INFORMATION

Corresponding Author

Mathieu Galmiche − *Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS), Université de Strasbourg − CNRS, UMR 7140, 67000 Strasbourg, France; Institut de Chimie et Procédés pour l'Énergie, l'Environnement et la Santé (ICPEES) − CAMATES, Université de Strasbourg − CNRS, UMR 7515, 67000 Strasbourg, France; School of Pharmaceutical Sciences, University of Geneva, 1211 Geneva, Switzerland; Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, 1211 Geneva, Switzerland;* orcid.org/0000-0001-5840-4786; Email: mathieu.galmiche@unige.ch

Authors

Lauriane Kuhn − *Plateforme Protéomique Strasbourg-Esplanade, Institut de Biologie Moléculaire et Cellulaire, FR 1589 CNRS, 67084 Strasbourg, France;* orcid.org/0000-0002-2637-1024

Maurice Millet − *Institut de Chimie et Procédés pour l'Énergie, l'Environnement et la Santé (ICPEES) − CAMATES, Université de Strasbourg − CNRS, UMR 7515, 67000 Strasbourg, France*

Yannis-Nicolas François − *Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS), Université de Strasbourg − CNRS, UMR 7140, 67000 Strasbourg, France;* orcid.org/0000-0002-8776-071X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.4c00936

Author Contributions

#M.G. and L.K. contributed equally and can both claim first-authorship. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

This work was supported by a Ph.D. grant from the French Ministry of Higher Education, Research and Innovation and the University of Strasbourg. The mass spectrometry instrumentation was funded by the University of Strasbourg IdEx program "Equipement mi-lourd" (2015) to "Plateforme Protéomique Strasbourg-Esplanade".

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors warmly thank Dr. Olivier Delhomme for his support and his advices regarding the implementation of the aerosol sampling device. The authors thank Philippe Hammann and Johana Chicher from the Strasbourg—Esplanade proteomic facility for their help in the MS analyses.

ABBREVIATIONS

BCA, bicinchoninic acid; BSA, bovine serum albumin; FDR, false discovery rate; GFF, glass fiber filter; LC−MS/MS, liquid chromatography tandem mass spectrometry; PBAP, primary biological aerosol particles; PCR, polymerase chain reaction; PM, airborne particulate matter; SDS, sodium dodecyl sulfate; SD, sodium dodecanoate; UTCT, urea−thiourea−CHAPS−tris

REFERENCES

(1) Jaenicke, R. Abundance of Cellular Material and Proteins in the Atmosphere. *Science* **2005**, *308* (5718), 73.

(2) Després, V. R.; Huffman, J. A.; 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 B* **2022**, *64* (1), 15598.

(3) Morawska, L.; Salthammer, T. Fundamentals of Indoor Particles and Settled Dust. In *Indoor Environment*; Morawska, L., Salthammer, T., Eds.; Wiley, 2003; pp 1−46.

(4) Prass, M.; Andreae, M. O.; De Araùjo, A. C.; Artaxo, P.; Ditas, F.; Elbert, W.; Förster, J.-D.; Franco, M. A.; Hrabe De Angelis, I.; Kesselmeier, J.; Klimach, T.; Kremper, L. A.; Thines, E.; Walter, D.; Weber, J.; Weber, B.; Fuchs, B. M.; Pöschl, U.; Pöhlker, C. Bioaerosols in the Amazon Rain Forest: Temporal Variations and Vertical Profiles

of Eukarya, Bacteria, and Archaea. *Biogeosciences* **2021**, *18* (17), 4873−4887.
(5) Yao, M. Bioaerosol: A Bridge and Opportunity for Many Scientific Research Fields. *J. Aerosol Sci.* **2018**, *115*, 108−112.
(6) Peccia, J.; Hospodsky, D.; Bibby, K. New Directions: A Revolution in DNA Sequencing Now Allows for the Meaningful Integration of Biology with Aerosol Science. *Atmos. Environ.* **2011**, *45* (10), 1896−1897.
(7) Triadó-Margarit, X.; Veillette, M.; Duchaine, C.; Talbot, M.; Amato, F.; Minguillón, M. C.; Martins, V.; De Miguel, E.; Casamayor, E. O.; Moreno, T. Bioaerosols in the Barcelona Subway System. *Indoor Air* **2017**, *27* (3), 564−575.
(8) Triadó-Margarit, X.; Cáliz, J.; Casamayor, E. O. A Long-Term Atmospheric Baseline for Intercontinental Exchange of Airborne Pathogens. *Environ. Int.* **2022**, *158*, 106916.
(9) Yoo, K.; Lee, T. K.; Choi, E. J.; Yang, J.; Shukla, S. K.; Hwang, S.; Park, J. Molecular Approaches for the Detection and Monitoring of Microbial Communities in Bioaerosols: A Review. *J. Environ. Sci.* **2017**, *51*, 234−247.
(10) Reinmuth-Selzle, K.; Kampf, C. J.; Lucas, K.; Lang-Yona, N.; Fröhlich-Nowoisky, J.; Shiraiwa, M.; Lakey, P. S. J.; Lai, S.; Liu, F.; Kunert, A. T.; Ziegler, K.; Shen, F.; Sgarbanti, R.; Weber, B.; Bellinghausen, I.; Saloga, J.; Weller, M. G.; Duschl, A.; Schuppan, D.; Pöschl, U. Air Pollution and Climate Change Effects on Allergies in the Anthropocene: Abundance, Interaction, and Modification of Allergens and Adjuvants. *Environ. Sci. Technol.* **2017**, *51* (8), 4119−4141.
(11) Lee, B. Minimum Sizes of Respiratory Particles Carrying SARS-CoV-2 and the Possibility of Aerosol Generation. *IJERPH* **2020**, *17* (19), 6960.
(12) Chen, X.; Kumari, D.; Achal, V. A Review on Airborne Microbes: The Characteristics of Sources, Pathogenicity and Geography. *Atmosphere* **2020**, *11* (9), 919.
(13) Triadó-Margarit, X.; Caliz, J.; Reche, I.; Casamayor, E. O. High Similarity in Bacterial Bioaerosol Compositions between the Free Troposphere and Atmospheric Depositions Collected at High-Elevation Mountains. *Atmos. Environ.* **2019**, *203*, 79−86.
(14) 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.
(15) Pomés, A.; Chapman, M. D.; Wünschmann, S. Indoor Allergens and Allergic Respiratory Disease. *Curr. Allergy Asthma Rep.* **2016**, *16* (6), 43.
(16) Gómez Londoño, L. F.; Brewer, M. T. Detection of Azole-Resistant Aspergillus Fumigatus in the Environment from Air, Plant Debris, Compost, and Soil. *PLoS One* **2023**, *18* (3), No. e0282499.
(17) Anees-Hill, S.; Douglas, P.; Pashley, C. H.; Hansell, A.; Marczylo, E. L. A Systematic Review of Outdoor Airborne Fungal Spore Seasonality across Europe and the Implications for Health. *Sci. Total Environ.* **2022**, *818*, 151716.
(18) Vacher, G.; Niculita-Hirzel, H.; Roger, T. Immune Responses to Airborne Fungi and Non-Invasive Airway Diseases. *Semin. Immunopathol.* **2015**, *37* (2), 83−96.
(19) Martorano, L.; Erwin, E. A. Aeroallergen Exposure and Spread in the Modern Era. *J. Allergy Clin. Immunol.* **2018**, *6* (6), 1835−1842.
(20) Matsui, E. C.; Simons, E.; Rand, C.; Butz, A.; Buckley, T. J.; Breysse, P.; Eggleston, P. A. Airborne Mouse Allergen in the Homes of Inner-City Children with Asthma. *J. Allergy Clin. Immunol.* **2005**, *115* (2), 358−363.
(21) Dallongeville, A.; Le Cann, P.; Zmirou-Navier, D.; Chevrier, C.; Costet, N.; Annesi-Maesano, I.; Blanchard, O. Concentration and Determinants of Molds and Allergens in Indoor Air and House Dust of French Dwellings. *Sci. Total Environ.* **2015**, *536*, 964−972.
(22) D'Ovidio, M. C.; Di Renzi, S.; Capone, P.; Pelliccioni, A. Pollen and Fungal Spores Evaluation in Relation to Occupants and Microclimate in Indoor Workplaces. *Sustainability* **2021**, *13* (6), 3154.
(23) Banchi, E.; Ametrano, C. G.; Stanković, D.; Verardo, P.; Moretti, O.; Gabrielli, F.; Lazzarin, S.; Borney, M. F.; Tassan, F.; Tretiach, M.; Pallavicini, A.; Muggia, L. DNA Metabarcoding Uncovers Fungal Diversity of Mixed Airborne Samples in Italy. *PLoS One* **2018**, *13* (3), No. e0194489.
(24) Staton, S. J. R.; Woodward, A.; Castillo, J. A.; Swing, K.; Hayes, M. A. Ground Level Environmental Protein Concentrations in Various Ecuadorian Environments: Potential Uses of Aerosolized Protein for Ecological Research. *Ecol. Indic.* **2015**, *48*, 389−395.
(25) Song, T.; Wang, S.; Zhang, Y.; Song, J.; Liu, F.; Fu, P.; Shiraiwa, M.; Xie, Z.; Yue, D.; Zhong, L.; Zheng, J.; Lai, S. Proteins and Amino Acids in Fine Particulate Matter in Rural Guangzhou, Southern China: Seasonal Cycles, Sources, and Atmospheric Processes. *Environ. Sci. Technol.* **2017**, *51* (12), 6773−6781.
(26) Miguel, A. G.; Cass, G. R.; Glovsky, M. M.; Weiss, J. Allergens in Paved Road Dust and Airborne Particles. *Environ. Sci. Technol.* **1999**, *33* (23), 4159−4168.
(27) Reinmuth-Selzle, K.; Tchipilov, T.; Backes, A. T.; Tscheuschner, G.; Tang, K.; Ziegler, K.; Lucas, K.; Pöschl, U.; Fröhlich-Nowoisky, J.; Weller, M. G. Determination of the Protein Content of Complex Samples by Aromatic Amino Acid Analysis, Liquid Chromatography-UV Absorbance, and Colorimetry. *Anal. Bioanal. Chem.* **2022**, *414* (15), 4457−4470.
(28) Samaké, A.; Jaffrezo, J.-L.; Favez, O.; Weber, S.; Jacob, V.; Albinet, A.; Riffault, V.; Perdrix, E.; Waked, A.; Golly, B.; Salameh, D.; Chevrier, F.; Oliveira, D. M.; Bonnaire, N.; Besombes, J.-L.; Martins, J. M. F.; Conil, S.; Guillaud, G.; Mesbah, B.; Rocq, B.; Robic, P.-Y.; Hulin, A.; Le Meur, S.; Descheemaecker, M.; Chretien, E.; Marchand, N.; Uzu, G. Polyols and Glucose Particulate Species as Tracers of Primary Biogenic Organic Aerosols at 28 French Sites. *Atmos. Chem. Phys.* **2019**, *19* (5), 3357−3374.
(29) Schneider, J.; Freutel, F.; Zorn, S. R.; Chen, Q.; Farmer, D. K.; Jimenez, J. L.; Martin, S. T.; Artaxo, P.; Wiedensohler, A.; Borrmann, S. Mass-Spectrometric Identification of Primary Biological Particle Markers and Application to Pristine Submicron Aerosol Measurements in Amazonia. *Atmos. Chem. Phys.* **2011**, *11* (22), 11415−11429.
(30) Yi, Y.; Liu, X.; Fu, M.; Li, Z.; Zhou, R.; Wang, Y.; Wei, B.; Meng, J.; Hou, Z.; Yan, L. Molecular Characteristics and Formation Mechanisms of Biogenic Secondary Organic Aerosols in the Summer Atmosphere at Mt. Tai on the North China Plain. *Aerosol Air Qual. Res.* **2019**, *9* (12), 2671−2682.
(31) Buters, J. T. M.; Thibaudon, M.; Smith, M.; Kennedy, R.; Rantio-Lehtimäki, A.; Albertini, R.; Reese, G.; Weber, B.; Galan, C.; Brandao, R.; Antunes, C. M.; Jäger, S.; Berger, U.; Celenk, S.; Grewling, Ł.; Jackowiak, B.; Sauliene, I.; Weichenmeier, I.; Pusch, G.; Sarioglu, H.; Ueffing, M.; Behrendt, H.; Prank, M.; Sofiev, M.; Cecchi, L. Release of Bet v 1 from Birch Pollen from 5 European Countries. Results from the HIALINE Study. *Atmos. Environ.* **2012**, *55*, 496−505.
(32) Celenk, S. Detection of Reactive Allergens in Long-Distance Transported Pollen Grains: Evidence from Ambrosia. *Atmos. Environ.* **2019**, *209*, 212−219.
(33) Kraaijeveld, K.; De Weger, L. A.; Ventayol García, M.; Buermans, H.; Frank, J.; Hiemstra, P. S.; Den Dunnen, J. T. Efficient and Sensitive Identification and Quantification of Airborne Pollen Using Next-Generation DNA Sequencing. *Mol. Ecol. Resour.* **2015**, *15* (1), 8−16.
(34) Lim, J.-H.; Nam, S. H.; Kim, J.; Kim, N. H.; Park, G.-S.; Maeng, J.-S.; Yook, S.-J. High-Volume Sampler for Size-Selective Sampling of Bioaerosols Including Viruses. *Atmos. Environ.* **2021**, *265*, 118720.
(35) Tordoni, E.; Ametrano, C. G.; Banchi, E.; Ongaro, S.; Pallavicini, A.; Bacaro, G.; Muggia, L. Integrated eDNA Metabarcoding and Morphological Analyses Assess Spatio-Temporal Patterns of Airborne Fungal Spores. *Ecol. Indic.* **2021**, *121*, 107032.
(36) Serrano-Silva, N.; Calderón-Ezquerro, M. C. Metagenomic Survey of Bacterial Diversity in the Atmosphere of Mexico City Using Different Sampling Methods. *Environ. Pollut.* **2018**, *235*, 20−29.
(37) Yang, T.; Wang, X.; Ng, H. Y.; Huang, S.; Zheng, X.; Bi, X. Airborne Antibiotic Resistome from Sludge Dewatering Systems: Mobility, Pathogen Accessibility, Cross-Media Migration Propensity, Impacting Factors, and Risks. *Water Res.* **2024**, *267*, 122552.
(38) Srikrishna, D. Pentagon Found Daily, Metagenomic Detection of Novel Bioaerosol Threats to Be Cost-Prohibitive: Can Virtualization and AI Make It Cost-Effective? *Health Secur.* **2024**, *22* (2), 108−129.

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX

(39) Meyer, S.; Hüttig, N.; Zenk, M.; Jäckel, U.; Pöther, D. Bioaerosols in Swine Confinement Buildings: A Metaproteomic View. *Environ. Microbiol. Rep.* **2023**, *15* (6), 684−697.
(40) Wilmes, P.; Bond, P. L. The Application of Two-Dimensional Polyacrylamide Gel Electrophoresis and Downstream Analyses to a Mixed Community of Prokaryotic Microorganisms. *Environ. Microbiol.* **2004**, *6* (9), 911−920.
(41) Püttker, S.; Kohrs, F.; Benndorf, D.; Heyer, R.; Rapp, E.; Reichl, U. Metaproteomics of Activated Sludge from a Wastewater Treatment Plant - A Pilot Study. *Proteomics* **2015**, *15* (20), 3596−3601.
(42) Li, S.; Hu, S.; Shi, S.; Ren, L.; Yan, W.; Zhao, H. Microbial Diversity and Metaproteomic Analysis of Activated Sludge Responses to Naphthalene and Anthracene Exposure. *RSC Adv.* **2019**, *9* (40), 22841−22852.
(43) Sánchez-Jiménez, E.; Abian, J.; Ginebreda, A.; Barceló, D.; Carrascal, M. Shotgun Proteomics to Characterize Wastewater Proteins. *MethodsX* **2023**, *11*, 102403.
(44) Borsetto, C.; Raguideau, S.; Travis, E.; Kim, D.-W.; Lee, D.-H.; Bottrill, A.; Stark, R.; Song, L.; Cha, C.-J.; Pearson, J.; Quince, C.; Singer, A. C.; Wellington, E. M. H. Impact of Sulfamethoxazole on a Riverine Microbiome. *Water Res.* **2021**, *201*, 117382.
(45) Bridoux, M. C.; Neibauer, J.; Ingalls, A. E.; Nunn, B. L.; Keil, R. G. Suspended Marine Particulate Proteins in Coastal and Oligotrophic Waters. *J. Mar. Syst.* **2015**, *143*, 39−48.
(46) Saito, M. A.; Bertrand, E. M.; Duffy, M. E.; Gaylord, D. A.; Held, N. A.; Hervey, W. J.; Hettich, R. L.; Jagtap, P. D.; Janech, M. G.; Kinkade, D. B.; Leary, D. H.; McIlvin, M. R.; Moore, E. K.; Morris, R. M.; Neely, B. A.; Nunn, B. L.; Saunders, J. K.; Shepherd, A. I.; Symmonds, N. I.; Walsh, D. A. Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing. *J. Proteome Res.* **2019**, *18* (4), 1461−1476.
(47) Bona, E.; Massa, N.; Novello, G.; Boatti, L.; Cesaro, P.; Todeschini, V.; Magnelli, V.; Manfredi, M.; Marengo, E.; Mignone, F.; Berta, G.; Lingua, G.; Gamalero, E. Metaproteomic Characterization of Vitis Vinifera Rhizosphere. *FEMS Microbiol. Ecol.* **2019**, *95*, fiy204.
(48) Liu, F.; Lai, S.; Reinmuth-Selzle, K.; Scheel, J. F.; Fröhlich-Nowoisky, J.; Després, V. R.; Hoffmann, T.; Pöschl, U.; Kampf, C. J. Metaproteomic Analysis of Atmospheric Aerosol Samples. *Anal. Bioanal. Chem.* **2016**, *408* (23), 6337−6348.
(49) Piovesana, S.; Capriotti, A. L.; Foglia, P.; Montone, C. M.; La Barbera, G.; Zenezini Chiozzi, R.; Laganà, A.; Cavaliere, C. Development of an Analytical Method for the Metaproteomic Investigation of Bioaerosol from Work Environments. *Proteomics* **2019**, *19* (23), 1900152.
(50) Infoclimat-Archives. Infoclimat. https://www.infoclimat.fr/fr/cartes/observations-meteo/archives/temperature_min/21/avril/2021/0h/france_ne.html, December 2022.
(51) Réseau National de Surveillance Aérobiologique (R.N.S.A.). Carte du Risque d'Allergie aux Pollens. https://www.pollens.fr/, July 2021.
(52) Bradford, M. M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **1976**, *72*, 248−254.
(53) Smith, P. K.; Krohn, R. I.; Hermanson, G. T.; Mallia, A. K.; Gartner, F. H.; Provenzano, M. D.; Fujimoto, E. K.; Goeke, N. M.; Olson, B. J.; Klenk, D. C. Measurement of Protein Using Bicinchoninic Acid. *Anal. Biochem.* **1985**, *150* (1), 76−85.
(54) Krohn, R. I. The Colorimetric Detection and Quantitation of Total Protein. In *Current Protocols in Cell Biology*; John Wiley & Sons, Inc., 2002; pp A3H1−A3H28.
(55) Perez-Riverol, Y.; Bai, J.; Bandla, C.; García-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D. J.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; Walzer, M.; Wang, S.; Brazma, A.; Vizcaíno, J. A. The PRIDE Database Resources in 2022: A Hub for Mass Spectrometry-Based Proteomics Evidences. *Nucleic Acids Res.* **2022**, *50* (D1), D543−D552.
(56) Galmiche, M.; Kuhn, L.; Millet, M.; Francois, Y. N. *Metaproteomics as a Powerful Tool for an Extensive Characterization of Ambient Bioaerosols*, 2024.
(57) Galmiche, M.; Kuhn, L.; Millet, M.; Francois, Y. N. *Metaproteomics as a Powerful Tool for an Extensive Characterization of Ambient Bioaerosols (Dataset #2)*, 2024.
(58) Bouyssié, D.; Hesse, A.-M.; Mouton-Barbosa, E.; Rompais, M.; Macron, C.; Carapito, C.; Gonzalez De Peredo, A.; Couté, Y.; Dupierris, V.; Burel, A.; Menetrey, J.-P.; Kalaitzakis, A.; Poisat, J.; Romdhani, A.; Burlet-Schiltz, O.; Cianférani, S.; Garin, J.; Bruley, C. Proline: An Efficient and User-Friendly Software Suite for Large-Scale Proteomics. *Bioinformatics* **2020**, *36* (10), 3148−3155.
(59) Jagtap, P.; Goslinga, J.; Kooren, J. A.; McGowan, T.; Wroblewski, M. S.; Seymour, S. L.; Griffin, T. J. A Two-step Database Search Method Improves Sensitivity in Peptide Sequence Matches for Metaproteomics and Proteogenomics Studies. *Proteomics* **2013**, *13* (8), 1352−1357.
(60) Jouffret, V.; Miotello, G.; Culotta, K.; Ayrault, S.; Pible, O.; Armengaud, J. Increasing the Power of Interpretation for Soil Metaproteomics Data. *Microbiome* **2021**, *9* (1), 195.
(61) Mi, H.; Muruganujan, A.; Casagrande, J. T.; Thomas, P. D. Large-Scale Gene Function Analysis with the PANTHER Classification System. *Nat. Protoc.* **2013**, *8* (8), 1551−1566.
(62) Rabilloud, T.; Lelong, C. Two-Dimensional Gel Electrophoresis in Proteomics: A Tutorial. *J. Proteonomics* **2011**, *74* (10), 1829−1841.
(63) Nakayasu, E. S.; Gritsenko, M.; Piehowski, P. D.; Gao, Y.; Orton, D. J.; Schepmoes, A. A.; Fillmore, T. L.; Frohnert, B. I.; Rewers, M.; Krischer, J. P.; Ansong, C.; Suchy-Dicey, A. M.; Evans-Molina, C.; Qian, W.-J.; Webb-Robertson, B.-J. M.; Metz, T. O. Tutorial: Best Practices and Considerations for Mass-Spectrometry-Based Protein Biomarker Discovery and Validation. *Nat. Protoc.* **2021**, *16* (8), 3737−3760.
(64) Van Den Bossche, T.; Kunath, B. J.; Schallert, K.; Schäpe, S. S.; Abraham, P. E.; Armengaud, J.; Arntzen, M. Ø.; Bassignani, A.; Benndorf, D.; Fuchs, S.; Giannone, R. J.; Griffin, T. J.; Hagen, L. H.; Halder, R.; Henry, C.; Hettich, R. L.; Heyer, R.; Jagtap, P.; Jehmlich, N.; Jensen, M.; Juste, C.; Kleiner, M.; Langella, O.; Lehmann, T.; Leith, E.; May, P.; Mesuere, B.; Miotello, G.; Peters, S. L.; Pible, O.; Queiros, P. T.; Reichl, U.; Renard, B. Y.; Schiebenhoefer, H.; Sczyrba, A.; Tanca, A.; Trappe, K.; Trezzi, J.-P.; Uzzau, S.; Verschaffelt, P.; Von Bergen, M.; Wilmes, P.; Wolf, M.; Martens, L.; Muth, T. Critical Assessment of MetaProteome Investigation (CAMPI): A Multi-Laboratory Comparison of Established Workflows. *Nat. Commun.* **2021**, *12* (1), 7305.
(65) Petersson Sjögren, M.; Alsved, M.; Šantl-Temkiv, T.; Bjerring Kristensen, T.; Löndahl, J. Measurement Report: Atmospheric Fluorescent Bioaerosol Concentrations Measured during 18 Months in a Coniferous Forest in the South of Sweden. *Atmos. Chem. Phys.* **2023**, *23* (9), 4977−4992.
(66) Bogaerts-Márquez, M.; Guirao-Rico, S.; Gautier, M.; González, J. Temperature, Rainfall and Wind Variables Underlie Environmental Adaptation in Natural Populations of Drosophila Melanogaster. *Mol. Ecol.* **2021**, *30* (4), 938−954.
(67) Secondi, J.; Dejean, T.; Valentini, A.; Audebaud, B.; Miaud, C. Detection of a Global Aquatic Invasive Amphibian, Xenopus Laevis, Using Environmental DNA. *Amphibia-Reptilia* **2016**, *37* (1), 131−136.
(68) Queirós, L.; Pereira, J. L.; Gonçalves, F. J. M.; Pacheco, M.; Aschner, M.; Pereira, P. Caenorhabditis Elegans as a Tool for Environmental Risk Assessment: Emerging and Promising Applications for a Nobelized Worm. *Crit. Rev. Toxicol.* **2019**, *49* (5), 411−429.
(69) Goddard, M. R.; Greig, D. Saccharomyces Cerevisiae: A Nomadic Yeast with No Niche? *FEMS Yeast Res.* **2015**, *15* (3), fov009.
(70) Browne, A. G. P.; Fisher, M. C.; Henk, D. A. Species-Specific PCR to Describe Local-Scale Distributions of Four Cryptic Species in the Penicillium Chrysogenum Complex. *Fungal Ecol.* **2013**, *6* (5), 419−429.
(71) García, M. A.; Alonso, J.; Fernández, M. I.; Melgar, M. J. Lead Content in Edible Wild Mushrooms in Northwest Spain as Indicator of Environmental Contamination. *Arch. Environ. Contam. Toxicol.* **1998**, *34* (4), 330−335.
(72) Falandysz, J. Mercury Bio-Extraction by Fungus Coprinus Comatus: A Possible Bioindicator and Mycoremediator of Polluted Soils? *Environ. Sci. Pollut. Res. Int.* **2016**, *23* (8), 7444−7451.

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX

(73) Liu, X.; Huang, X.; Chu, C.; Xu, H.; Wang, L.; Xue, Y.; Arifeen Muhammad, Z. U.; Inagaki, F.; Liu, C.; Liu, C. Genome, Genetic Evolution, and Environmental Adaptation Mechanisms of Schizophyllum Commune in Deep Subseafloor Coal-Bearing Sediments. *iScience* **2022**, *25* (6), 104417.
(74) Durand, A.; Maillard, F.; Foulon, J.; Gweon, H. S.; Valot, B.; Chalot, M. Environmental Metabarcoding Reveals Contrasting Belowground and Aboveground Fungal Communities from Poplar at a Hg Phytomanagement Site. *Microb. Ecol.* **2017**, *74* (4), 795−809.
(75) Amend, A. From Dandruff to Deep-Sea Vents: Malassezia-like Fungi Are Ecologically Hyper-Diverse. *PLoS Pathog.* **2014**, *10* (8), No. e1004277.
(76) Lang-Yona, N.; Shuster-Meiseles, T.; Mazar, Y.; Yarden, O.; Rudich, Y. Impact of Urban Air Pollution on the Allergenicity of Aspergillus Fumigatus Conidia: Outdoor Exposure Study Supported by Laboratory Experiments. *Sci. Total Environ.* **2016**, *541*, 365−371.
(77) Singh, S.; Madlala, A. M.; Prior, B. A. Thermomyces Lanuginosus: Properties of Strains and Their Hemicellulases. *FEMS Microbiol. Rev.* **2003**, *27* (1), 3−16.
(78) Contato, A. G.; Borelli, T. C.; Buckeridge, M. S.; Rogers, J.; Hartson, S.; Prade, R. A.; Polizeli, M. D. L. T. D. M. Secretome Analysis of Thermothelomyces Thermophilus LMBC 162 Cultivated with Tamarindus Indica Seeds Reveals CAZymes for Degradation of Lignocellulosic Biomass. *JoF* **2024**, *10* (2), 121.
(79) Rineau, F.; Roth, D.; Shah, F.; Smits, M.; Johansson, T.; Canbäck, B.; Olsen, P. B.; Persson, P.; Grell, M. N.; Lindquist, E.; Grigoriev, I. V.; Lange, L.; Tunlid, A. The Ectomycorrhizal Fungus Paxillus Involutus Converts Organic Matter in Plant Litter Using a Trimmed Brown-Rot Mechanism Involving Fenton Chemistry. *Environ. Microbiol.* **2012**, *14* (6), 1477−1487.
(80) Kuo, H.-C.; Hui, S.; Choi, J.; Asiegbu, F. O.; Valkonen, J. P. T.; Lee, Y.-H. Secret Lifestyles of Neurospora Crassa. *Sci. Rep.* **2014**, *4*, 5135.
(81) Ahmed, E.; Holmström, S. J. M. Siderophores in Environmental Research: Roles and Applications. *Microb. Biotechnol.* **2014**, *7* (3), 196−208.
(82) Williamson, B.; Tudzynski, B.; Tudzynski, P.; van Kan, J. A. L. Botrytis Cinerea: The Cause of Grey Mould Disease. *Mol. Plant Pathol.* **2007**, *8* (5), 561−580.
(83) Arammehr, A.; Dehghan, P.; Chadeganipour, M.; Katoueezadeh, M.; Shadzi, S. Detection of Dermatophytes from Dermatophytosis-Suspected Cases in Iran, Evaluation of Polymerase Chain Reaction-Sequencing Method. *Adv. Biomed. Res.* **2020**, *9*, 56.
(84) Lauer, A.; Etyemezian, V.; Nikolich, G.; Kloock, C.; Arzate, A. F.; Sadiq Batcha, F.; Kaur, M.; Garcia, E.; Mander, J.; Kayes Passaglia, A. Valley Fever: Environmental Risk Factors and Exposure Pathways Deduced from Field Measurements in California. *Int. J. Environ. Res. Publ. Health* **2020**, *17* (15), 5285.
(85) Carhan, A.; Ozkan, O.; Ozkaya, E. The First Identification of Encephalitozoon Cuniculi Infection in an Animal Care Worker in Turkey. *Iran. J. Parasitol.* **2015**, *10* (2), 280−285.
(86) Kosuge, J.; Goto, Y.; Shinjo, T.; Tohru, A.; Takatori, K. Detection of Emericella nidulans from bedding materials in horse breeding environment and its significance as a causative agent of guttural pouch mycosis in horses. *Nihon Ishinkin Gakkai Zasshi* **2000**, *41* (4), 251−256.
(87) Savastano, C.; de Oliveira Silva, E.; Gonçalves, L. L.; Nery, J. M.; Silva, N. C.; Dias, A. L. T. Candida Glabrata among Candida Spp. from Environmental Health Practitioners of a Brazilian Hospital. *Braz. J. Microbiol.* **2016**, *47* (2), 367−372.
(88) Allergome. The Platform for Allergen Knowledge; Allergome. http://allergome.org/, May 2023.
(89) Fallah, P. Exposure Potential of Other Common Penicillium Species in Indoor Environments. *J. Allergy Clin. Immunol.* **2004**, *113* (2), S223.
(90) Apangu, G. P.; Frisk, C. A.; Petch, G. M.; Muggia, L.; Pallavicini, A.; Hanson, M.; Skjøth, C. A. Environmental DNA Reveals Diversity and Abundance of Alternaria Species in Neighbouring Heterogeneous Landscapes in Worcester, UK. *Aerobiologia* **2022**, *38* (4), 457−481.
(91) Gusenkov, S.; Stutz, H. Top-down and Bottom-up Characterization of Nitrated Birch Pollen Allergen Bet v 1a with CZE Hyphenated to an Orbitrap Mass Spectrometer. *Electrophoresis* **2018**, *39* (9−10), 1190−1200.
(92) Kamimura, N.; Sakamoto, S.; Mitsuda, N.; Masai, E.; Kajita, S. Advances in Microbial Lignin Degradation and Its Applications. *Curr. Opin. Biotechnol.* **2019**, *56*, 179−186.
(93) Anderson, O. D. The Spectrum of Major Seed Storage Genes and Proteins in Oats (Avena Sativa). *PLoS One* **2014**, *9* (7), No. e83569.
(94) Ma, F.; Li, M.; Li, T.; Liu, W.; Liu, Y.; Li, Y.; Hu, W.; Zheng, Q.; Wang, Y.; Li, K.; Chang, J.; Chen, M.; Yang, G.; Wang, Y.; He, G. Overexpression of Avenin-like b Proteins in Bread Wheat (Triticum Aestivum L.) Improves Dough Mixing Properties by Their Incorporation into Glutenin Polymers. *PLoS One* **2013**, *8* (7), No. e66758.
(95) Taylor, S. L.; Marsh, J. T.; Koppelman, S. J.; Kabourek, J. L.; Johnson, P. E.; Baumert, J. L. A Perspective on Pea Allergy and Pea Allergens. *Trends Food Sci. Technol.* **2021**, *116*, 186−198.
(96) Jin, Y.; Acharya, H. G.; Acharya, D.; Jorgensen, R.; Gao, H.; Secord, J.; Ng, P. K. W.; Gangur, V. Advances in Molecular Mechanisms of Wheat Allergenicity in Animal Models: A Comprehensive Review. *Molecules* **2019**, *24* (6), 1142.
(97) Aguilera-Insunza, R.; Iturriaga, C.; Mariñanco, A.; Venegas, L.; Aravena, G.; Perez-Mateluna, G.; Baptista-Dias, N.; Borzutzky, A.; Wandersleben, T. High Prevalence of Lupin Allergy among Patients with Peanut Allergy. *Ann. Allergy Asthma Immunol.* **2023**, *130* (2), 225−232.
(98) Niederberger, V.; Hayek, B.; Vrtala, S.; Laffer, S.; Twardosz, A.; Vangelista, L.; Sperr, W. R.; Valent, P.; Rumpold, H.; Kraft, D.; Ehrenberger, K.; Valenta, R.; Spitzauer, S. Calcium-Dependent Immunoglobulin E Recognition of the Apo- and Calcium-Bound Form of a Cross-Reactive Two EF-Hand Timothy Grass Pollen Allergen, Phl p 7. *FASEB J.* **1999**, *13* (8), 843−856.
(99) Matthiesen, F.; Schumacher, M. J.; Løwenstein, H. Characterization of the Major Allergen of Cynodon Dactylon (Bermuda Grass) Pollen, Cyn d I. *J. Allergy Clin. Immunol.* **1991**, *88* (5), 763−774.
(100) Verdino, P.; Barderas, R.; Villalba, M.; Westritschnig, K.; Valenta, R.; Rodriguez, R.; Keller, W. Three-Dimensional Structure of the Cross-Reactive Pollen Allergen Che a 3: Visualizing Cross-Reactivity on the Molecular Surfaces of Weed, Grass, and Tree Pollen Allergens. *J. Immunol.* **2008**, *180* (4), 2313−2321.
(101) Breiteneder, H.; Pettenburger, K.; Bito, A.; Valenta, R.; Kraft, D.; Rumpold, H.; Scheiner, O.; Breitenbach, M. The Gene Coding for the Major Birch Pollen Allergen Betv1, Is Highly Homologous to a Pea Disease Resistance Response Gene. *EMBO J.* **1989**, *8* (7), 1935−1938.
(102) Swoboda, I.; De Weerd, N.; Bhalla, P. L.; Niederberger, V.; Sperr, W. R.; Valent, P.; Kahlert, H.; Fiebig, H.; Verdino, P.; Keller, W.; Ebner, C.; Spitzauer, S.; Valenta, R.; Singh, M. B. Mutants of the Major Ryegrass Pollen Allergen, Lol p 5, with Reduced IgE-Binding Capacity: Candidates for Grass Pollen-Specific Immunotherapy. *Eur. J. Immunol.* **2002**, *32* (1), 270−280.
(103) Calam, D. H.; Davidson, J.; Ford, A. W. Investigations of the Allergens of Cocksfoot Grass (Dactylis Glomerata) Pollen. *J. Chromatogr.* **1983**, *266*, 293−300.
(104) Fiebig, A.; Varesio, L. M.; Alejandro Navarreto, X.; Crosson, S. Regulation of the Erythrobacter Litoralis DSM 8509 General Stress Response by Visible Light. *Mol. Microbiol.* **2019**, *112* (2), 442−460.
(105) Bazylinski, D. A.; Williams, T. J.; Lefèvre, C. T.; Berg, R. J.; Zhang, C. L.; Bowser, S. S.; Dean, A. J.; Beveridge, T. J. Magnetococcus Marinus Gen. Nov., Sp. Nov., a Marine, Magnetotactic Bacterium That Represents a Novel Lineage (Magnetococcaceae Fam. Nov., Magnetococcales Ord. Nov.) at the Base of the Alphaproteobacteria. *Int. J. Syst. Evol. Microbiol.* **2013**, *63* (Pt 3), 801−808.
(106) Ge, H.; Wang, Y.; Zhao, X. Research on the Drug Resistance Mechanism of Foodborne Pathogens. *Microb. Pathog.* **2022**, *162*, 105306.
(107) Salje, J. Orientia Tsutsugamushi: A Neglected but Fascinating Obligate Intracellular Bacterial Pathogen. *PLoS Pathog.* **2017**, *13* (12), No. e1006657.

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX

(108) Genné, D.; Rossel, M.; Sarr, A.; Battilotti, F.; Rais, O.; Rego, R. O. M.; Voordouw, M. J. Competition between Strains of Borrelia Afzelii in the Host Tissues and Consequences for Transmission to Ticks. *ISME J.* **2021**, *15* (8), 2390−2400.
(109) Stuen, S.; Granquist, E. G.; Silaghi, C. Anaplasma Phagocytophilum–a Widespread Multi-Host Pathogen with Highly Adaptive Strategies. *Front. Cell. Infect. Microbiol.* **2013**, *3*, 31.
(110) Wilkie, I. W.; Harper, M.; Boyce, J. D.; Adler, B. Pasteurella Multocida: Diseases and Pathogenesis. *Curr. Top. Microbiol. Immunol.* **2012**, *361*, 1−22.
(111) Tateishi, Y.; Ozeki, Y.; Nishiyama, A.; Miki, M.; Maekura, R.; Kida, H.; Matsumoto, S. Virulence of Mycobacterium Intracellulare Clinical Strains in a Mouse Model of Lung Infection - Role of Neutrophilic Inflammation in Disease Severity. *BMC Microbiol.* **2023**, *23* (1), 94.
(112) Celli, J.; Zahrt, T. C. Mechanisms of Francisella Tularensis Intracellular Pathogenesis. *Cold Spring Harb. Perspect. Med.* **2013**, *3* (4), a010314.
(113) Patrick, S. A Tale of Two Habitats: Bacteroides Fragilis, a Lethal Pathogen and Resident in the Human Gastrointestinal Microbiome. *Microbiology* **2022**, *168* (4), 001156.
(114) Gunbin, K. V.; Afonnikov, D. A.; Kolchanov, N. A. Molecular Evolution of the Hyperthermophilic Archaea of the Pyrococcus Genus: Analysis of Adaptation to Different Environmental Conditions. *BMC Genom.* **2009**, *10*, 639.
(115) Anderson, I. J.; Sieprawska-Lupa, M.; Lapidus, A.; Nolan, M.; Copeland, A.; Glavina Del Rio, T.; Tice, H.; Dalin, E.; Barry, K.; Saunders, E.; Han, C.; Brettin, T.; Detter, J. C.; Bruce, D.; Mikhailova, N.; Pitluck, S.; Hauser, L.; Land, M.; Lucas, S.; Richardson, P.; Whitman, W. B.; Kyrpides, N. C. Complete Genome Sequence of Methanoculleus Marisnigri Romesser et al. 1981 Type Strain JR1. *Stand. Genomic Sci.* **2009**, *1* (2), 189−196.

https://doi.org/10.1021/acs.jproteome.4c00936
J. Proteome Res. XXXX, XXX, XXX−XXX