

**DEVELOPMENT OF AN ANALYTICAL METHOD FOR THE METAPROTEOMIC INVESTIGATION OF
BIOAEROSOL FROM WORK ENVIRONMENTS**

Susy Piovesana, Anna Laura Capriotti*, Patrizia Foglia, Carmela Maria Montone, Giorgia La Barbera,
Riccardo Zenezini Chiozzi, Aldo Laganà, Chiara Cavaliere

Department of Chemistry, Sapienza Università di Roma, Piazzale Aldo Moro 5, 00185 Rome, Italy

susy.piovesana@uniroma1.it; patrizia.foglia@uniroma1.it; carmelamaria.montone@uniroma1.it;
giorgia.labarbera@uniroma1.it; riccardo.zenezini@uniroma1.it; aldo.lagana@uniroma1.it
chiara.cavaliere@uniroma1.it;

***Corresponding author:** Anna Laura Capriotti, Department of Chemistry, Sapienza Università di
Roma, Piazzale Aldo Moro 5, 00185 Rome, Italy.

E-mail: annalaura.capriotti@uniroma1.it

Telephone: (+39) 06 4991 3679

Fax (+39) 06 4906 31

Received: 04 17, 2019; Revised: 06 21, 2019; Accepted: 07 12, 2019

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/pmic.201900152](https://doi.org/10.1002/pmic.201900152).

This article is protected by copyright. All rights reserved.

Abstract

The metaproteomic analysis of air particulate matter provides valuable information about the properties of bioaerosols in the atmosphere and their influence on climate and public health. In this work, a new method for the extraction and analysis of proteins in airborne particulate matter from quartz microfiber filters was developed. Different protein extraction procedures were tested to select the best extraction protocol based on protein recovery. The optimized method was tested for the extraction of proteins from spores of ubiquitous bacteria species and used for the metaproteomic characterization of filters from three work environments. In particular, ambient aerosol samples were collected in a composting plant, in a wastewater treatment plant and in an agricultural holding. A total of 179, 15, 205 and 444 proteins were identified in the composting plant, the wastewater treatment plant and the agricultural holding, respectively. In agreement with the major categories of primary biological aerosol particles, **all identified proteins originated primarily from fungi, bacteria and plants**. The paper is the first metaproteomic study applied to bioaerosol samples collected in occupationally relevant environmental sites and, even though not aimed at monitoring the risk exposure of workers, it provides information on the possible exposure in the working environmental sites.

Keywords: bioaerosol; metaproteomics; extraction methods; composting plant; wastewater treatment plant; agricultural holding; high resolution mass spectrometry

Statement of Significance

This manuscript describes the metaproteomic analysis of aerosol samples collected in work environments. This is a novel use of aerosol samples and it is needed, as there is no really comprehensive way of analysing aerosol samples from a metaproteomic point of view. This paper could help to advance methods for metaproteomic analysis of bioaerosols, specifically by comparing protein extraction protocols and pairing the best performing extraction protocol with a gel-free protein separation procedure applied for the first time for analysis of bioaerosol samples. The obtained data showed that bioaerosol was essentially made of fungi, bacteria and plant proteins, many of which could be associated to possible aerosolisation and could be a major health concern for workers on site and to the populations residing in neighbouring area.

Introduction

Primary biological aerosols (PBAs), generally called bioaerosols, are a subset of atmospheric particles, which are directly released from the biosphere into the atmosphere. They comprise living and dead organisms (e.g., algae, archaea, bacteria), dispersal units (e.g., fungal spores and plant pollen), and various fragments or excretions. They have a direct effect on climate changes and they can induce genotoxic effects, cardiovascular effects, infectious, allergenic, or toxic effects on living organisms, impacting public health and agriculture on local, regional, and global scale.^[1]

PBAs have a dimension ranging between 0.3 and 100 μm and they contain a large variety of different biomolecules, such as lipids, nucleic acids, amino acids and glycosylates. Recently, it has been found that proteins compose a much larger portion of the atmospheric particle budget than the trace amount previously assumed, with potentially up to 25 % of atmospheric particles having a biological origin.^[2]

Concerning atmospheric proteins, it has been recognized that outdoor environment is composed of a “protein soup” containing bacteria, fungi, spores, house dust mites, multiple pollens, animal dander, molds, and fragments of animals, insects and plants which are implicated in severe lung diseases, causing respiratory disorders, asthma and chronic obstructive pulmonary disease to exposed individuals.^[3,4] Due to the difficulty of specifically tracing these microorganisms in air, there is lack of knowledge about the dispersal of airborne microorganisms even though it is well recognized that an exposition to a complex mixture of toxins and allergens can lead to respiratory tract infections. In recent years, molecular tools have been used to develop new tracers which should help in risk assessments.^[5] PBAs were extensively studied over the past two decades in terms of microbial diversity of bioaerosol originated from composting plants (CPs),^[5] wastewater treatment plants (WWTPs)^[6] and from intensive farming.^[7] Most of the studies have been carried out by cell culture techniques and quantitative polymerase chain reaction targeting DNA; while cultivation-based techniques systematically underestimate the diversity of bioaerosols, culture-independent studies do not underestimate PBAs concentration and can reveal a higher biodiversity via amplified ribosomal DNA analysis; furthermore, next-generation sequencing techniques are sensitive and robust methods to assess microbial composition not only from composting facilities and WWTPs but also from environments, such as soil and water.^[5] Despite the introduction of powerful DNA sequencing techniques, such as shotgun or amplicon (typically the 16S rRNA gene) sequencing and whole metagenome shotgun sequencing, which allow comprehensive analysis of global gene expression profiles by isolating single bacterial strains, little is known about global protein expression of bioaerosol from compost, wastewater and livestock. Metaproteomic study is nowadays fundamental to understand functional information and gain knowledge on the bacterial and fungal communities present in these specific environmental sites, since protein profiles can furnish more direct information than functional genes without isolating single bacterial strains.^[8,9] To date, most research examining protein PBAs are based on protein quantification by traditional assays, e.g. bicinchoninic acid (BCA) assay,^[10] nano-orange,^[11] and Bradford assay. The only metaproteomic

paper already published^[12] was focused on ambient aerosol samples collected on the roof of the Max Planck Institute for Chemistry and exploited a gel-based, time-consuming proteomic approach which provided a very small number of protein identifications. Shotgun gel-free separation is nowadays the best choice to analyze highly complex protein mixtures as it provides, compared to gel-based approaches, a considerable reduction to the complexity of the sample without any loss of information.^[13] The focus of the current study was the development of a comprehensive metaproteomic platform for the determination of protein present in bioaerosol from work environments, e.g. CP, WWTP and agricultural holding (AH). In particular, five different extraction methods were compared to maximize the protein extraction recovery. The developed method was based on shotgun proteomics: proteins were digested by trypsin, separated by nano-high performance liquid chromatography (nanoHPLC) and analyzed by high resolution tandem mass spectrometry. Then, different bioinformatics tools, such as Proteome Discoverer and Unipept, were employed both to identify proteins and peptides and to describe the biodiversity in the analyzed samples. The final method was tested for protein extraction from spores of ubiquitous bacteria species and used for the metaproteomic characterization of real filter samples. This study provided additional and complementary information compared to culture dependent and independent techniques about microbial, fungal and plant diversity of bioaerosol and its taxonomic composition. To the best of our knowledge, this work is the first metaproteomic study performed on bioaerosol in work environments and it allows to characterize PBA emission as well as assess exposure risks for workers and public health.

2. MATERIALS AND METHOD

2.1 Materials

All chemicals of the highest grade available were purchased from Sigma-Aldrich now Merck (St. Louis, MO, USA) unless otherwise stated. Mass Spec Grade trypsin was provided by Promega (Madison, WI, USA). Ultra-pure water was prepared by arium 611 VF system from Sartorius (Göttingen, Germany). HPLC-MS grade water and acetonitrile (ACN) were provided by VWR International (Milan, Italy).

2.2 PM₁₀ sample collection

Bioaerosol samples were collected at three different working environmental sites: the CP ACEA Ambiente U.L.4, (Località Pian del Vantaggio, Orvieto), the WWTP Roma-Est ACEA-Ato2 (Rome) and the AH Maccarese S.p.A. (Fiumicino, Rome). In particular, 7 filters were collected at the CP (subsequently named F1_CP, F2_CP, F3_CP, F4_CP, F5_CP, F6_CP, F7_CP), 2 filters were collected at the WWTP (subsequently named F1_WWTP, F2_WWTP) and 5 samples were collected at the AH; in particular, in the AH two filters were collected in proximity of the cow barn (these filters were named F1_AH_S1, F2_AH_S1) and three filters were collected in blending plant (these filters were named F1_AH_S2, F2_AH_S2, F3_AH_S2). For each filter three technical replicates (nanoHPLC-MS/MS runs) were performed.

Daily PM₁₀ samples were collected using a high volume air sampler (Tisch Environ, Inc.). The flow rate of the Tisch sampler was 1.13 m³ min⁻¹, and PM₁₀ samples were collected on quartz microfiber filters (QMA, 20.3 cm × 25.4 cm, Whatman), which were heat treated (600 °C) to reduce trace organics and sterilized under UV lamp for 4 h prior to use. The loaded samples were stored in decontaminated aluminum foil bags at -80 °C. To detect possible contaminations from the samplers

and sample handling, blank samples were taken as well. Blank sample filters were mounted in the sampler like for regular sampling, but the pump was turned on only for up to 30 s.

Table S1 in the Supporting Information summarizes all the analyzed filters, with related detailed sampling information and number of identified proteins.

2.3 Protein extraction

Five different extraction protocols were compared to select the best protocol based on the protein recovery from quartz filters. The five extraction protocols were evaluated by spiking on blank filters five standard proteins, i.e. bovine serum albumin (BSA), bovine cytochrome C (CYT-C), equine apomyoglobin (ApoMb), bovine apo-transferrin (ApoTF) and human immunoglobulin G (IgG), at two different amounts (50 and 100 μg) for each protein, and comparing protein recovery obtained by the BCA assay. All spiking experiments were performed in triplicate in order to assess experimental variability.

Filters were cut into small pieces, dispersed with 3 g of washed sand in a ceramic mortar and blended by a pestle with the aid of liquid nitrogen. For all tested protocols, 15 mL of a buffer consisting of 50 mmol L^{-1} Tris-HCl (pH 8.8) with 0.1% protease inhibitors was added.

In protocol A, the extraction buffer was prepared as previously described, with some modifications,^[14] and was made up of 15 mmol L^{-1} KCl, 2% (w/v) sodium dodecyl sulfate (SDS), 10 mmol L^{-1} ethylenediaminetetraacetic acid (EDTA) and 20 mmol L^{-1} dithiothreitol (DTT). Protocol B was based on an optimization of protein extraction from air filter samples developed in a previous work^[12]. The buffer used in protocol B contained 192 mmol L^{-1} glycine and 0.1 % SDS in 25 mmol L^{-1} Tris-HCl. Protocol C used the same reagents as protocol A without EDTA. In protocol D and E, 1% (w/v) sodium deoxycholate (SDC)^[15] and 0.5% (w/v) sodium dodecanoate (SD)^[16] were solubilized in Tris-HCl (pH 8.5) instead of SDS, respectively. All the samples coming from the 5 protocols were incubated on ice for 1 h with intermittent vortexing (1 min) every 15 min. After this step, samples

were placed in an ultrasonic bath for 1 h; the entire cycle was repeated twice and finally the insoluble material was removed by centrifugation at room temperature for 20 min at $20,000 \times g$.

The supernatants coming from experiments A, B and C were transferred into new centrifuge tubes and subjected to protein precipitation to remove SDS, which is incompatible with the subsequent proteomic workflow. Proteins were precipitated by addition of 4 volumes of cold acetone with 10% (w/v) trichloroacetic acid. Samples were vortexed and incubated overnight at $-20\text{ }^{\circ}\text{C}$. The obtained pellets were collected by centrifugation ($10,000 \times g$ for 15 min at $4\text{ }^{\circ}\text{C}$), washed three times with ice-cold acetone, air-dried, and dissolved in 0.5 mL of 8 mol L^{-1} urea in 50 mmol L^{-1} Tris-HCl (pH 8.8). Protocol D and E did not require detergent removal. All proteins samples were quantified by the BCA assay using BSA as standard and stored at $-80\text{ }^{\circ}\text{C}$ until digestion.

The optimized extraction method (protocol D) was tested on filters spiked with 3 mg of spores of ubiquitous bacteria species (*Bacillus subtilis*, equivalent to $1\text{E}8$ spores) and finally used for the metaproteomics characterization of filters from the three working environments (Table S1).

2.4 Protein digestion and purification

One-hundred μg of the spores of *Bacillus subtilis* and 1 mg of proteins extracted from real samples were reduced, alkylated and digested with trypsin as previously described.^[17] Briefly, samples were first diluted 4 times with 50 mmol L^{-1} Tris-HCl (pH 8.5), then sequencing grade-modified trypsin was added (1:20, enzyme/protein ratio) and the samples were incubated overnight at $37\text{ }^{\circ}\text{C}$.

Enzymatic digestion was quenched by trifluoroacetic acid (TFA) to reach pH 2 and SDC precipitate was removed by centrifugation at $14,000 \times g$ for 5 min. Finally, the peptides were desalted on C18 cartridges (Bond Elut 1 mg C18, Agilent Technologies, Palo Alto, CA, USA) as previously described.^[18] Eluates were dried in a Speed-Vac SC 250 Express (Thermo S 164 avant, Holbrook, NY, USA), re-constituted with $500\text{ }\mu\text{L}$ 0.1 % HCOOH and stored at $-80\text{ }^{\circ}\text{C}$ until analysis.

2.5 NanoHPLC-tandem mass spectrometry analysis

Samples were analyzed by nanoHPLC coupled to tandem mass spectrometry (MS/MS). The analysis was performed on a Dionex Ultimate 3000 (Dionex Corporation Sunnyvale, CA, U.S.A.) directly connected to a LTQ-Orbitrap XL mass spectrometer (Thermo Scientific, Bremen, Germany) by a nanoelectrospray ion source. Peptide mixtures were enriched on a 300 μm ID \times 5 mm Acclaim PepMap 100 C18 (5 μm particle size, 100 \AA pore size) precolumn (Dionex Corporation Sunnyvale, CA, U.S.A.), employing a premixed mobile phase made up of $\text{H}_2\text{O}/\text{ACN}$, 98:2 (v/v) containing 0.1% (v/v) TFA, at a flow-rate of 10 $\mu\text{L min}^{-1}$. Then, peptide mixtures were separated by reversed phase chromatography using a 25 cm long fused silica nanocolumn (75 μm ID), in-house packed with Acclaim-C18 2.2 μm microparticles. The gradient was optimized to detect the largest set of peptides using $\text{H}_2\text{O}/\text{HCOOH}$ (99.9:0.1, v/v) as mobile phase A and ACN/HCOOH (99.9:0.1, v/v) as mobile phase B. Phase B was maintained for 5 min at 2%, then linearly increased to 7% within 10 min; afterwards, phase B was increased to 15% in 85 min, to 21% in 45 min, to 30% in 30 min and then to 50% in 10 min. Finally, phase B was increased to 80% in 15 min and kept constant for 20 min for column washing, then lowered to 2% and maintained for 45 min for column equilibration. The nanoelectrospray ion source was operated in positive ionization mode, with spray and capillary voltage set at 2.90 kV and 42 V, respectively, and capillary temperature set at 180 $^\circ\text{C}$. Full MS spectra were acquired in profile mode in the m/z range 350–1800 in the Orbitrap with resolution set at 60,000 (Full Width Half Maximum at m/z 400). Data-dependent MS/MS acquisition of the five most intense monoisotopic peaks in the spectra was performed by collision-induced dissociation at normalized collision energy of 35%, with isolation window of 2 m/z , in the LTQ. Rejection of +1, and unassigned charge states was enabled. Ion trap and Orbitrap maximum ion injection times were set to 1000 and 200 ms, respectively. Automatic gain control was used to prevent overfilling of the ion traps and was set to 5×10^5 .

Database search and peptide identification

The acquired raw MS/MS data files from Xcalibur software (version 2.2 SP1.48, Thermo Fisher Scientific) were searched by Proteome Discoverer software (version 1.3, Thermo Scientific) and the Mascot (v.2.3.2, Matrix Science) search engine, as previously described.^[19] Spectra were searched by decoy strategy in the SwissProt database, with no taxonomy (all entries, 557275 entries). Single peptide identifications were checked and contaminants removed. The monoisotopic mass tolerance for precursor ions and product ions were set to 10 ppm and 0.8 Da, respectively. The relaxed false discovery rate (FDR) was 0.05 while the strict FDR was 0.01.

We decided to accept protein identifications also with a single unique peptide and low coverage, after a manual spectrum check, to provide a more comprehensive protein profile description.

A gene ontology (GO) analysis was performed using STRAP.^[20] We also used Unipept web application^[21] to infer the lowest common ancestor taxon for peptides detected in each sample.

RESULTS AND DISCUSSION

3.1 Method comparison for protein extraction from bioaerosol

Protein extraction is a crucial challenge, mostly in samples where it is not possible to provide a universal and simple sample preparation method. For this reason, in this work five different extraction protocols were compared to select the best extraction procedure for an effective metaproteomic analysis for both free proteins, proteins within cells and proteins bound to the aerosol material. Initially, methods were compared based on the effectiveness in the recovery of free proteins, which was determined by spiking 5 different standard proteins on blank filters; BSA, CYT-C, ApoMb, ApoTF and IgG were chosen as representative proteins to cover a wide range of molecular weights and their recoveries were evaluated for each extraction procedure. Figure 1

displays the average free protein recovery (%), with standard deviation, at two spiking levels (50 μg or 100 μg) for the different tested methods (A-E).

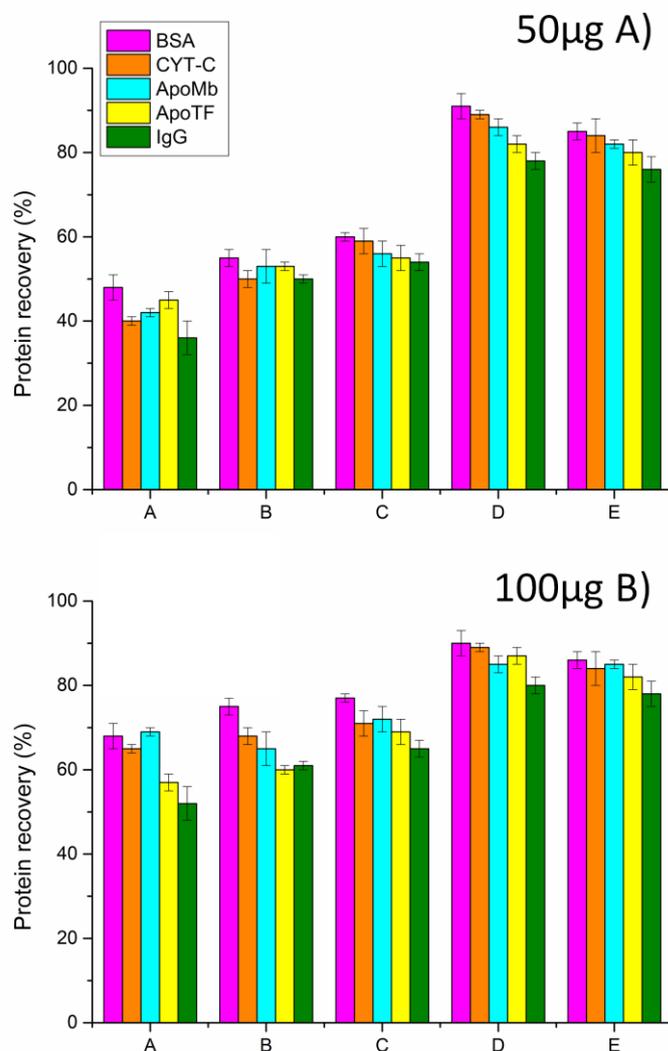


Figure 1. Average protein recovery (%), with standard deviation, from blank air filter spiked with 50 μg (A) or 100 μg (B) of BSA, CYT-C, ApoMb, ApoTF, IgG, and extracted by the different tested methods (A-E).

Protocols A, B and C all exploited buffers containing SDS, which can denature secondary and non-disulfide-linked tertiary structures of proteins and, therefore, facilitates the solubilization of otherwise water-insoluble proteins as well as water-soluble proteins.^[22] Protocols A and C only differed for the use of EDTA to chelate metals, such as heavy metals, that could occur in air samples due to their ubiquitous and persistent nature. Protocol B was tested because it was the best protocol employed in a previous metaproteomic work^[12]; protocol B exploited a Tris/Gly/SDS buffer, which is suitable for the extraction of both water-soluble and water-insoluble proteins while minimizing other potential non-covalent interactions, such as the ones between proteins and ambient aerosol components (e.g., soot, dust) or between proteins and the filter material. All three protocols required a protein precipitation step to remove detergents incompatible with the downstream protein quantification, digestion and MS analysis. The recovery of the five tested standard proteins showed no meaningful difference for protocols A, B and C, and it was 36-60% for 50 µg experiments (Figure 1A) and 60-77% for 100 µg experiments (Figure 1B). Recoveries were significantly lower for filters spiked at 50 µg, a result which could be attributed to an incomplete precipitation or resolubilization of the proteins. In fact, sample losses become crippling with precipitation procedures from mass-limited samples, as they make it challenging to obtain sufficient amounts of proteins to generate high-quality MS data. To overcome this challenge, a micro-scale technique, i.e. filter aided sample preparation (FASP), was applied to protocols A, B and C, as it is particularly suited for low sample sizes (100 µg or lower) and still enables an excellent proteome coverage.^[23] We tested two FASP devices with different molecular weight cutoff (10 kDa and 30 kDa), however both clogged due to the presence of “carbonaceous aerosols” that remain in solution after centrifugation; therefore, no data were obtained by the FASP method variations and no further modifications were tested on protocols A, B and C.

Compared to protocols A, B or C, protocols D and E resulted in higher protein recovery for all the 5 standard proteins at both spiking levels, since they exploited detergents, such as SDC (D) and SD (E), which are compatible with in-solution protein digestion and the subsequent nanoHPLC-MS/MS

analysis. Such detergents can be removed by simple acidification and centrifugation,^[16] thus avoiding loss of proteins and highlighting that even if the protein amount decreases, the extraction efficiency remains exactly the same with a 85-90% protein recovery (Figure 1A and B). Thus, the application of lysis buffers containing components that do not require removal by protein precipitation resulted suitable for analysis of low abundance protein samples. Moreover, protocols D and E were faster, easier, more robust and less expensive than the common protocols employed in proteomic studies using SDS. The subsequent choice to carry out all the metaproteomic experiments on real environmental samples using protocol D rather than protocol E was based on the ease of use of SDC, as it is more soluble than SD in Tris-HCl buffer. Moreover, considering that atmospheric aerosol particles contain a series of chemical components which may greatly affect the recovery of proteins in filter samples, a mixture of the 5 proteins at the two concentration levels, namely 50 and 100 µg, was also spiked on a filter sampling in WWTP. Data are shown in Supplementary Material Figure S1. Also in this case the best results were obtained employing protocols D with a protein recovery percentage higher than 90% for both spiked concentration. The protein recovery could become an indicator to show the influence of matrix effects in real aerosol samples.

Before analyzing the samples collected at the different work environmental sites, further evaluation experiments were performed on protocol D, to simulate a more realistic condition and evaluate the extraction efficiency of proteins within cells; protocol D was applied to blank filter samples spiked with the spores from a ubiquitous bacterium, i.e. *Bacillus subtilis*. The disruption of the cell wall is a critical phase of the extraction procedure for proteins collected on air filters, as PBAs are essentially made of fungal spores, bacteria, and other cells and cellular fragments. Compared to the previous paper already published on bioaerosol metaproteomics,^[12] where the authors presented an analytical method primarily aimed at the extraction of proteins that are easily released from PBAs, the extraction method tested in this work on *Bacillus subtilis* did not only use a detergent and surfactant, but also included a mechanical lysis step for the disruption of microbial cell walls and spores present in PBAs.^[22,23] By this approach, the characteristic proteins of *Bacillus subtilis* were

identified by bioinformatics analysis, thus confirming the efficacy of the selected extraction method (data and chromatograms are reported in Supporting Information Table S2).

At this stage the method was finally applied to the analysis and characterization of filters from different work environmental sites collected in the Lazio (Italy) and influenced by urban and rural boundary layer air masses, i.e. a WWTP, a CP and an AH. However, in the samples coming from the WWTP and the CP, the abundance of carbonaceous material required further modification of protocol D. In fact, the interactions between proteins and particles collected from outdoor sources or used as model environmental sources (ultrafine carbon black) have already been reported in past literature.^[24] For this reason, we did not only analyze the purified supernatants, but also the pellets obtained after digestion quenching. The dispersed carbon material could, indeed, adsorb peptides, produced by protein tryptic digestion, or proteins, directly after cell lysis and filter disruption. Therefore, the carbon pellet was dispersed in a 0.1 mol L⁻¹ NH₃ aqueous solution, to elute bound peptides, and then the eluate extracted with ethyl acetate to remove SDC. However, in most cases this carbon washing step provided only a limited number of additional identifications, but the procedure proved significant in samples with large amounts of carbon material, in particular for samples F4_CP and F5_CP.

3.2 Protein identification in ambient aerosol samples from environmental working sites

In this work, a total of 179, 15, 205 and 444 proteins were identified in the bioaerosols collected at the CP, the WWTP, the AH cow stable and the AH feed blending plant, respectively (Table S1, for the complete list of identified proteins, peptides and their taxonomic classification see Supporting Information Tables S3, S4 and S5). A closer look to individual filter identifications showed a large variability in the number of proteins identified for each filter. In particular, for the CP filters (Supporting Information Table S2 and Table S4) variations were due to bioaerosol emission rates and dispersion influenced by many factors, including compost temperature, sorting, shredding and

turning of the piles, meteorological conditions (e.g. temperature, humidity, wind and weather), and the composition of the source organic material.^[25] Few proteins were identified from the two filters from the WWTP (only 15 proteins and 17 peptides); such small numbers were probably due to the fact that aerosol components, such as carbonaceous material, may hamper protein sample preparation. Moreover, the presence of humic compounds interfering with protein separation, the high complexity of the bacterial community and the lack of sufficient genomic sequences for protein identification can also affect protein identification.

For a better understanding of the biological significance, the gene ontology (GO) analysis was used to get additional insight and it was obtained by STRAP^[20]. The results reported for biological processes, molecular function, and cellular components are reported in the Supporting Information Figure S2 for the CP samples (F1-F7), in Figure S3 for the WWTP samples (F1, F2) and in Figure S4 for the AH samples (F1_S1, F2_S1 and F1_S2, F2_S2, F3_S2). For all samples, metabolic and cellular processes were the most represented biological processes, while regulation and localization with catalytic activity, binding and structure molecule activity were the principal molecular functions (Figure S2, S3 and S4 A and B). Regarding the cellular component ontology, the identified proteins belonged to cytoplasm, cytoskeleton, macromolecular complex, nucleus, mitochondria, ribosome, and other intracellular compartment (Figure S2, S3 and S4 C). From these data it appeared that SDC extraction was effective not only for hydrophilic proteins (i.e. those belonging to cytosol and cytoplasm) but also for proteins for which a lysis of cell membrane was required (i.e. membrane-bound organelles, nucleus, ribosomal and mitochondrial proteins).

3.3 Metaproteomic characterization of ambient aerosol samples

A deep and complete characterization of bioaerosol metaproteome is crucial to highlight many aspects related to the abundance and properties of bioaerosol from working environments and their influence on workers and health. Exposures to bioaerosols in the environment are associated with a

wide range of health effects, including contagious and infectious diseases, respiratory symptoms, acute toxic effects, allergies and cancer^[26]. Even if we are constantly exposed to bioaerosol in every moment of the day, both in outdoor or indoor environments, some of these environments could pose a higher risk for the health due to particular proteins and microorganisms contained onto bioaerosol. The major microbial constituents of bioaerosols are fungi and bacteria, while their products constitute allergens and pathogens. An allergen is any substance (antigen), most often eaten or inhaled, that causes an unusual immune response and triggers an allergic reaction. The most common symptoms of allergens are usually runny nose, stuffy nose, scratchy throat, itchy eyes and sneezing. Many *Bacillus* species transmitted as bioaerosol are known to be potent allergens causing respiratory tract discomfort.^[27,28] In our samples different proteins coming from different *Bacillus*, such as *Bacillus subtilis*, *Lactobacillus casei*, *Lactobacillus helveticus* were identified (Table S3, S4 and S5). Along with allergens, some pathogen microorganisms were identified in our samples, such as *Staphylococcus*, which commonly resides on the skin, nasal passage, and axillae causing various diseases due to its enterotoxins and antigens, and *Bacillus subtilis*, which can form endospores to ensure longer survival in the environment.^[29] Research into several industrial activities, e.g., wastewater treatment^[30] and composting,^[31] has highlighted the high exposure levels to biological agents. Given the above, hierarchical taxonomic profiles were therefore produced by the means of the Unipept web application, which is able to support biodiversity analysis of large and complex metaproteome samples using tryptic peptide information obtained from shotgun MS/MS experiments and from the Proteome Discoverer software.^[21]

3.3.1 Bacterial and fungal diversity of the filters collected from the composting plant

It has been reported that severe lung diseases and allergies may occur as a result of exposure to organic dust, bacteria, actinomycetes and fungi from municipal waste composting sites^[32]. To search

for possible species responsible for these adverse effects, a treeview displaying the complete results of the identified taxonomic hierarchy from superkingdoms to species was obtained (Figure 2 and Supporting Information Table S3). Of 206 peptides uploaded in Unipept software, 122 were with highest sequence identity to eukaryotic peptides, 83 to bacterial peptides and just one peptides to *Archaea* domain. Within the eukaryotic subset, the majority had the best match to sequences from members of *Fungi* (98) followed by *Viridiplantae* (21) and *Dictyosteliida* (3); within the bacterial subset, the highest sequence identity was for sequences from *Proteobacteria* (30), *Actinobacteria* (13), *Firmicutes* (8), *Cyanobacteria* (2), *Bacteroidetes* (1) and *Elusimicrobia* (1). One peptide from *Euryarchaeota*, a phylum of *Archaea*, was detected (Figure 2 and Supporting Information Table S3). The results agreed with the most representative domains for the CP environment, namely fungi and bacteria; fungi are important especially during the curing stage, whereas bacteria are predominant during the earlier thermophilic stage.^[33]

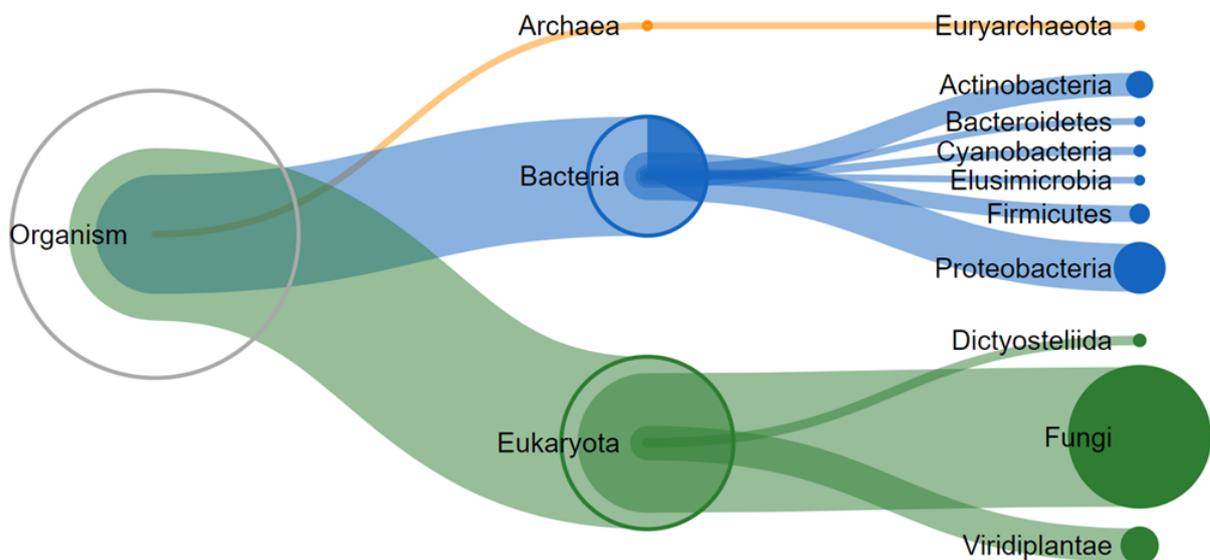


Figure 2. Treeview showing the results of the biodiversity analysis of filters from composting plant (CP).

Concerning fungi, the samples collected were dominated by Ascomycota (*Aspergillus fumigatus*, *Candida*, *Histoplasma capsulatum*, *Hypocreales*, *Hypocreomycetidae*, *Kluyveromyces marxianus*, *Lachancea kluyveri*, *Parengyodontium album*, and *Saccharomyces* species) and Basidiomycota (*Cryptococcus neoformans*). It is well known from the literature that Ascomycota and Basidiomycota can include the most common fungal allergens, such as *Alternaria*, *Cladosporium*, *Penicillium*, and *Aspergillus*. In our sample, two proteins from *Aspergillus flavus* and *Aspergillus clavatus* were identified; moreover, concerning *Aspergillus fumigatus*, it can cause infections in humans due to the presence in its cell walls of mycotoxins or β -(1-3)-glucans.^[34] Yeast have also been listed as species causing allergic diseases, such as chronic urticarial and respiratory allergic diseases. In our sample 46 proteins were identified as *Saccharomyces cerevisiae* species and 4 proteins as *Candida albicans* and *Candida glabrata*. For example, *Candida albicans* may attack the human body, causing severe infection and even death, particularly for immunosuppressed persons. Enolase, phosphoglycerate kinase, and aldolase proteins are well known allergen proteins and they were found associated to different species in our samples, for example *Saccharomyces cerevisiae*, *Gluconobacter oxydans*, *Cutibacterium acnes*, *Mycobacterium smegmatis*, *Yersinia enterocolitica*, *Salinispora tropica*, *Corynebacterium urealyticum* and other fungal and microbial species (Table S3).^[35] We identified also a protein from *Saccharopolyspora* species which is usually implicated in allergic reaction such as alveolitis or bronchial asthma^[36].

Regarding bacterial diversity, we found different thermophilic species and gram-negative bacteria, such as 7 proteins belonging to *Bacillus*, 1 protein from *Streptococcus*, 1 protein from *Staphylococcus*, 11 proteins from *E. Coli* and 6 proteins from *Thermobifida fusca*, which are usually monitored in composting aerosol since they have some pathogenic effect on the human health. *Thermobifida fusca*, for example, is a typical thermophilic bacterium found in heated organic materials, such as compost heaps, rotting hay, manure piles or mushroom growth medium, and produces spores that can be allergenic and cause a condition called farmers lung.^[37]

Our data agreed with the data reported in a recent review article^[5] indicating the dominant bacteria and fungi identified in aerosols from composting facilities by cultivation based techniques. The most identified phyla were the same reported in this work for bacteria and fungi, but the methodology described here provides some advantages over metaproteomic investigations based on cultivation methods. In fact, culture dependent methods are limited due to the difficulty of accessing every genotype from the fungal or bacterial community and the low culturability of bacteria and fungi occurring in bioaerosol. Our data also agreed with a previous metaproteomic work carried out directly on urban solid waste in a large-scale aerobic CP; in the paper the authors highlighted that some bacteria, such as *Bacillales*, *Actinobacteria*, and *Saccharomyces* increased significantly compared to their abundance in the composting process^[38]. Our work represents a step forward in this field since the information related to bioaerosol is more important compared to information on urban solid waste, as bioaerosol can travel up to 1 km from the point of emission, leading to risks of infection not only for workers in the composting facility but also for nearby residents.

Finally, we demonstrated the presence of amoebae in CP aerosol; in particular we identified 3 proteins belonging to *Dictyostelium discoideum*, a species of soil-living amoeba belonging to the phylum *Amoebozoa*; until now just one work carried out by Conza et al. demonstrated the presence of free-living amoebae in composting bioaerosol.^[25]

3.3.2 Bacterial and fungal diversity of the filters collected from the wastewater treatment plant

The WWTP is another important sampling site for the monitoring of bioaerosol, as the wastewater treatment process can generate bioaerosol which may contain the pathogenic microorganisms present in wastewater. Airborne microorganisms, usually present in this type of bioaerosol, belong to mesophilic heterotrophic bacteria, coliforms and enterococci. Pre-treatment and primary settlers are the stages with the highest concentration of bioaerosol. Many works showed that a gradual

decrease of bioaerosol emissions was observed during the advanced wastewater treatment, starting from the pre-treatment to the final tertiary treatment.^[6,39]

In this metaproteomic work, WWTP bioaerosol was collected on two filters (F1-F2) in proximity of the oxidation ditches, a tertiary treatment process. The 15 proteins and 17 peptides identified in these samples were ascribed to three domains (Figure 3 and in Supporting Information Table S4); in particular, 9 peptide sequences were specific at eukaryote level (5 fungal species, i.e. *Agaricomycetes*, *Aspergillus fumigatus*, *Candida albicans*, *Hypocreomycetidae*, *Saccharomyces*) and 3 sequences belonged to *Viridiplantae* kingdom (*Brassicaceae*, *Streptophytina*); 6 peptide sequences were specific at bacteria level (*Bacteria*, *Candidatus Protochlamydia amoebophila*, *Gammaproteobacteria*, *Lawsonia intracellularis*, *Nitrosomonas Rhodobacteraceae*); finally, 3 peptide sequences were specific at *Archea* level and attributed to a protein related to the pathogenic microorganism *Archaeoglobus fulgidus*, a sulphur-metabolizing organism (identified in F1); a previous study provided evidence of inflammatory potential of some species of *Archea* in mice^[40], and possible impact on respiratory health. Therefore, the monitoring of *Archea* species in bioaerosol is fundamental in order to evaluate the exposure risks.

In our samples no proteins related to pathogenic heterotrophic bacteria were identified. In F1 we identified 5 proteins from coliforms (4 proteins from *E. Coli*, 1 protein from *Enterobacter asburiae*) and 1 protein from yeast (*Saccharomyces cerevisiae*). Our results agreed with the previous works showing that the production of bioaerosol decreases with microbial and fungal diversity during tertiary treatment. This could be due to Gram negative bacteria generally having a low degree of survival during aerosolization and the desiccation processes^[39].

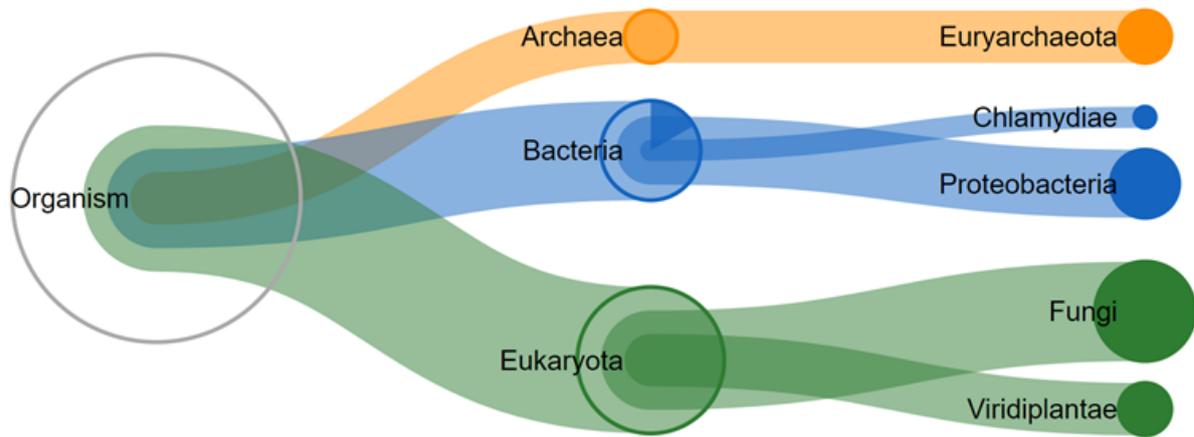


Figure 3. Treeview showing the results of the biodiversity analysis of filters from waste water treatment plant (WWTP).

3.3.3 Bacterial and fungal diversity of filters collected from the agricultural holding

For the AH bioaerosol investigation, a total of 5 quartz filters were collected and analysed: two filters were collected in proximity of the cow stable (F1-F2_S1) and three filters were collected in the blending plant (F1-F3_S2). A total of 205 proteins and 348 peptides and 444 proteins and 999 peptides were identified in filters from the cow stable and the blending plant, respectively (SI, Table S5). In samples collected in the cow stable, 95% of peptide sequences belonged to *Eukaryota* superkingdom (334 peptides) and only 5% to bacteria domain (18 peptides). *Fungi* (157), *Kinetoplastida* (1), *Metazoa* (114) and *Viridiplantae* (62) were the four dominant eukaryotic phyla (Figure 4, Supporting Information Table S4). Concerning samples collected in the blending plant, the microorganism diversity followed a trend similar to the one of the cow stable site, with a larger number of sequences, 952, assigned to *Eukaryota* superkingdom (562 *Fungi*, 317 *Viridiplantae*, 73 *Metazoa*); only 46 peptide sequence belonged to bacterial domain and 1 sequence belonged to viruses (*alfalfa mosaic virus* species, who can lead to necrosis on crops and belonged to category 2 of risk, Figure 4 and Supporting Information Table S5).

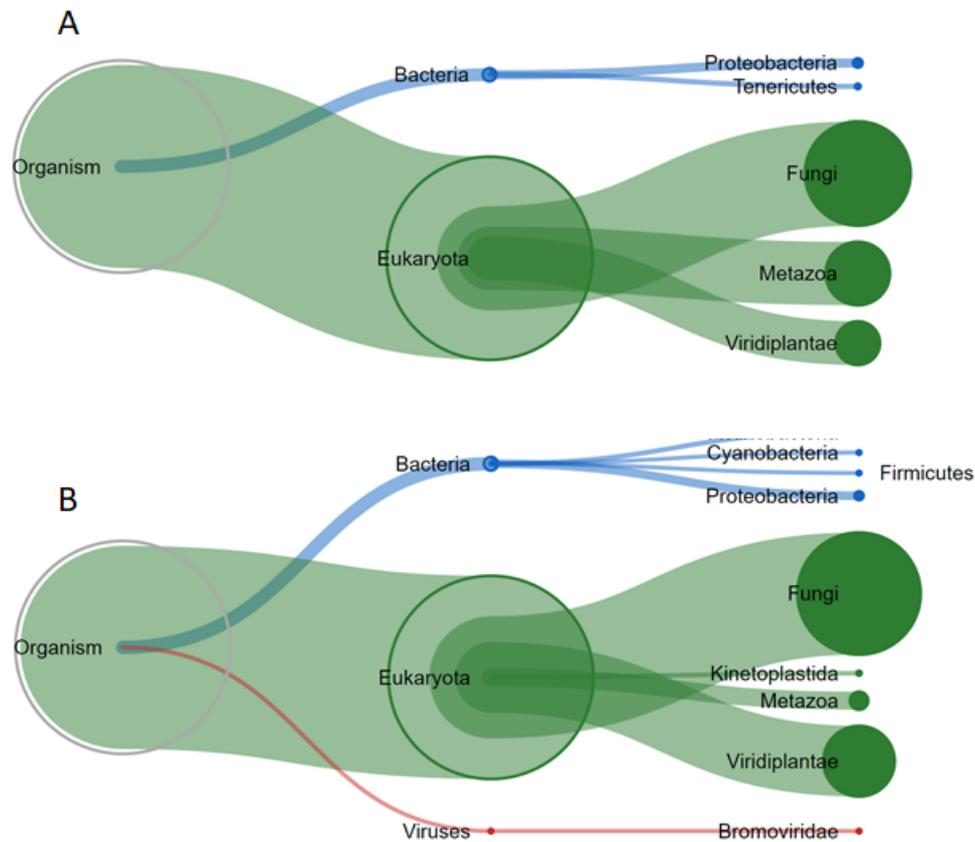


Figure 4. Tree view showing the results of the biodiversity analysis of filters from cow farm (A) and blending plant (B).

In both sampling sites the most representative category was *Fungi*, which are usually ubiquitous in farms, due to straw being an important source of fungal aerosol. Fungal biodiversity is essential in occupational exposure, and was previously evaluated mainly by culture-based methods or next generation sequencing.^[41] In comparison to culture methodologies, the developed metaproteomic analysis provided a more comprehensive view of bioaerosol fungal diversity. For instance, 6 fungal classes were identified as dominant in a recent work by Mbareche et co-workers,^[41] namely *Eurotiomycetes*, *Dothideomycetes*, *Wallemiomycetes*, *Agaricomycetes*, *Sordariomycetes* and *Tremellomycetes*. In this work, 13 classes were identified in the cow stable samples (*Agaricomycetes*, *Ascomycota*, *Cladosporium herbarum*, *Dikarya*, *Eurotiales*, *Hypocreomycetidae*, *Pezizomycotina*,

Pleosporineae, Puccinia, Saccharomyces, Saccharomycetaceae, Schizosaccharomyces pombe, Ustilago maydis) and 12 classes in the blending plant samples (*Eurotiomycetes, Dothideomycetes, Wallemiomycetes, Agaricomycetes, Sordariomycetes, Tremellomycetes, Agaricomycetes, Ascomycota, Boletales, Dikarya, Saccharomyces, Saccharomycetaceae*).

Two proteins from *Alternaria alternata*, one of the allergens implicated in case of respiratory arrest in sensitive children and young adults, were identified in filters from the blending plant site. Different allergenic proteins were also identified, in particular 7 enolases, 2 phosphoglycerate kinases and 6 aldolases.

The analysed filters have shown a broad spectrum of fungi, with many of them with pathogenic characteristics, thus highlighting the necessity of an adequate monitoring of bioaerosol in order to minimize exposure risks. Animal confinement also affected the fungal composition of bioaerosols. Proteins from *Viridiplantae* were also identified in large amounts, especially in the three filters collected where handling, storage and blending of animal feed was carried out (S2). Some of these proteins were grass pollen allergens from the genera *Glycine max* (soybean), for example we identified immunoglobulin E binding proteins^[42]. We also identified a soybean Kunitz trypsin inhibitor protein, another important food allergen for soy allergic patients.^[42]

Many proteins related to *Saccaromyces cerevisiae* species were identified, showing that this is commonly used as a supplement in feed for ruminants, as well as *Glycine max* and *Heliantus annus*.

The human exposure to harmful fungi may be higher during handling of feed.

This study demonstrated that the diversity of microorganisms in bioaerosols is so high that the development of high-throughput metaproteomic methodologies is fundamental to deeply characterize the bioaerosol protein profile and to better assess the occupational exposure.

Understanding the microbial and fungal characteristics of bioaerosols will help developing effective measures to control emission.

4. Conclusion

In this work an analytical methodology was developed in order to extract proteins from bioaerosol samples. We compared 5 different extraction protocols by evaluating the protein recovery for each of them, to cope with the challenge of trace protein amounts typical of bioaerosol samples. The developed method was applied to the analysis of filter samples from three working environment sites, i.e. CP, WWTP and AH (which included cow farm and blending plant). The metaproteomic analysis presented here allowed a global profiling of proteins in bioaerosol and demonstrated that microorganism diversity is so high that the development of a high-throughput metaproteomic methodology is fundamental to deeply characterize the protein profile and better assess the occupational exposure. Such data on the microbial and fungal characteristics of bioaerosol will help develop effective measures to control the nature of bioaerosol emission. The results confirmed that close to the CP, the WWTP and the blending plant in the HA, a large number of proteins from bacteria, actinomycetes and fungi species may be aerosolised and turned into a major health concern for workers and for the populations residing in neighbouring areas. Even if the risk assessment of workers is outside the scope of this work and no monitoring study was performed, the results of this investigation suggested potential risks for workers, which may require a reduction of the exposition to bioaerosols. The study also indicated how a metaproteomic investigation of bioaerosols can be a valuable tool in bioaerosol characterization. However, further studies should be focused on the improvement of data analysis in order to identify proteins with a higher coverage and sequence identity. Furthermore, to obtain more reliable results and meaningful conclusions, enough samples should be collected for each working place. The metaproteomic characterization of a larger number of aerosol samples in working environments is of importance for the future study of bioaerosol effect on human health.

ASSOCIATE CONTENT

Figure S1 in the Supporting Information provides average protein recovery (%), with standard deviation, from WWTP air filter (real sample) spiked with 100µg (green) or 50 µg (pink) of BSA, CYT-C, ApoMb, ApoTF, IgG protein mixture, and extracted by the different tested methods (A-E). Figures S2, S3 and S4 in the Supporting Information provide the gene ontology data for the three working environment sites (CP, WWTP and AH). Tables S1, S2, S3, S4 and S5 contain the information on the identified proteins and peptides for every filter sample analysed in the work.

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work was supported by INAIL within the project BRIC ID23.

We also thank Dr Francesca Buiarelli and her research group for providing us with the filter samples analysed in this work.

The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE^[43] partner repository with the dataset identifier PXD012345.

Reviewer account details: Username: reviewer47552@ebi.ac.uk; Password: PSzZwuzD

REFERENCES

- [1] J. Fröhlich-Nowoisky, C. J. Kampf, B. Weber, J. A. Huffman, C. Pöhlker, M. O. Andreae, N. Lang-Yona, S. M. Burrows, S. S. Gunthe, W. Elbert, H. Su, P. Hoor, E. Thines, T. Hoffmann, V. R. Després, U. Pöschl, *Atmos. Res.* **2016**, *182*, 346.
- [2] S. J. R. Staton, J. A. Castillo, T. J. Taylor, P. Herckes, M. A. Hayes, *Anal. Bioanal. Chem.* **2013**, *405*, 351.
- [3] P. Di Filippo, D. Pomata, C. Riccardi, F. Buiarelli, V. Gallo, A. Quaranta, *Atmos. Environ.* **2014**, *98*, 179.
- [4] S. J. R. Staton, A. Woodward, J. A. Castillo, K. Swing, M. A. Hayes, *Ecol. Indic.* **2015**, *48*, 389.
- [5] N. Wéry, *Front. Cell. Infect. Microbiol.* **2014**, *42*, DOI 10.3389/fcimb.2014.00042.
- [6] L. Pascual, S. Pérez-Luz, M. A. Yáñez, A. Santamaría, K. Gibert, M. Salgot, D. Apraiz, V. Catalán, *Aerobiologia (Bologna)*. **2003**, *19*, 261.
- [7] P. Douglas, S. Robertson, R. Gay, A. L. Hansell, T. W. Gant, *Int. J. Hyg. Environ. Health* **2018**, *221*, 134.
- [8] D.-Z. Wang, L.-F. Kong, Y.-Y. Li, Z.-X. Xie, *Int. J. Mol. Sci.* **2016**, *17*, 1275.
- [9] D. B. Sloan, G. M. Bennett, P. Engel, D. Williams, H. Ochman, in *Haroon, M.F., Skennerton, C.T., Steen, J.A., Lachner, N., Hugenholtz, P., Tyson, G.W. Microb. Metagenomics, Metatranscriptomics, Metaproteomics Elsevier.Methods Enzymol.*, **2013**, pp. 445–464.
- [10] T. Song, S. Wang, Y. Zhang, J. Song, F. Liu, P. Fu, M. Shiraiwa, Z. Xie, D. Yue, L. Zhong, J. Zheng, S. Lai, *Environ. Sci. Technol.* **2017**, *51*, 6773.
- [11] M. Y. Menetrez, K. K. Foarde, R. K. Esch, T. R. Dean, D. A. Betancourt, S. A. Moore, E. R.

- Svendsen, K. Yeatts, *Aerosol Sci. Technol.* **2007**, *41*, 884.
- [12] F. Liu, S. Lai, K. Reinmuth-Selzle, J. F. Scheel, J. Fröhlich-Nowoisky, V. R. Després, T. Hoffmann, U. Pöschl, C. J. Kampf, *Anal. Bioanal. Chem.* **2016**, *408*, 6337.
- [13] A. Fercha, A. L. Capriotti, G. Caruso, C. Cavaliere, H. Gherroucha, R. Samperi, S. Stampachiacchiere, A. Laganà, *J. Proteomics* **2013**, *91*, 486.
- [14] R. Zenezini Chiozzi, A. L. Capriotti, C. Cavaliere, G. La Barbera, S. Piovesana, A. Laganà, *J. Funct. Foods* **2016**, *27*, 262.
- [15] N. A. Kulak, G. Pichler, I. Paron, N. Nagaraj, M. Mann, *Nat. Methods* **2014**, *11*, 319.
- [16] Y. Pasing, S. Colnoe, T. Hansen, *Proteomics* **2017**, *17*, 1500.
- [17] A. L. Capriotti, G. Caruso, C. Cavaliere, S. Piovesana, R. Samperi, A. Laganà, *Anal. Chim. Acta* **2013**, *776*, 57.
- [18] A. L. Capriotti, G. Caruso, C. Cavaliere, S. Piovesana, R. Samperi, A. Laganà, *Anal. Chim. Acta* **2012**, *740*, 58.
- [19] A. L. Capriotti, C. Cavaliere, P. Foglia, S. Piovesana, R. Samperi, R. Zenezini Chiozzi, A. Laganà, *Anal. Bioanal. Chem.* **2015**, *407*, 845.
- [20] V. N. Bhatia, D. H. Perlman, C. E. Costello, M. E. McComb, *Anal. Chem.* **2009**, *81*, 9819.
- [21] B. Mesuere, B. Devreese, G. Debyser, M. Aerts, P. Vandamme, P. Dawyndt, *J. Proteome Res.* **2012**, *11*, 5773.
- [22] A. L. Capriotti, C. Cavaliere, P. Foglia, S. Piovesana, R. Samperi, S. Stampachiacchiere, A. Laganà, *Anal. Chim. Acta* **2013**, *800*, 36.
- [23] P. Feist, A. B. Hummon, *Int. J. Mol. Sci.* **2015**, *16*, 3537.

- [24] A. Kocbach, A. I. Totlandsdal, M. Låg, M. Refsnes, P. E. Schwarze, *Toxicol. Lett.* **2008**, *176*, 131.
- [25] L. Conza, S. C. Pagani, V. Gaia, *PLoS One* **2013**, *8*, DOI 10.1371/journal.pone.0068244.
- [26] J. Douwes, P. Thorne, N. Pearce, D. Heederik, *Ann. Occup. Hyg.* **2003**, DOI 10.1093/annhyg/meg032.
- [27] C. Humbal, S. Gautam, U. Trivedi, *Environ. Int.* **2018**, *118*, 189.
- [28] K. H. Kim, E. Kabir, S. A. Jahan, *J. Environ. Sci. (China)* **2018**, *67*, 23.
- [29] H. A. M. Pagalilauan, C. E. M. Paraoan, P. G. Vital, *Arch. Environ. Occup. Heal.* **2018**, *73*, 107.
- [30] S. Püttker, F. Kohrs, D. Benndorf, R. Heyer, E. Rapp, U. Reichl, *Proteomics* **2015**, *15*, 3596.
- [31] R. Dehghani, M. A. Asadi, E. Charkhloo, G. Mostafaie, M. Saffari, G. A. Mousavi, M. Pourbabaei, *J. Environ. Prot. (Irvine, Calif.)* **2012**, DOI 10.4236/jep.2012.31008.
- [32] S. Agarwal, P. Mandal, A. Srivastava, *Procedia Environ. Sci.* **2016**, *35*, 400.
- [33] P. F. Strom, *Appl. Environ. Microbiol.* **1985**, *50*, 906.
- [34] M.-W. Chang, C.-R. Lee, H.-F. Hung, K.-S. Teng, H. Huang, C.-Y. Chuang, *Int. J. Environ. Res. Public Health* **2014**, *11*, 337.
- [35] W. E. Horner, A. Helbling, J. E. Salvaggio, S. B. Lehrer, *Clin. Microbiol. Rev.* **1995**, *8*, 161.
- [36] D. J. O'Connor, S. M. Daly, J. R. Sodeau, *Waste Manag.* **2015**, *42*, 23.
- [37] K. C. Nayak, *In Silico Biol.* **2009**, *9*, 337.
- [38] D. Liu, M. Li, B. Xi, Y. Zhao, Z. Wei, C. Song, C. Zhu, *Microb. Biotechnol.* **2015**, *8*, 950.
- [39] S. Karra, E. Katsivela, *Water Res.* **2007**, *41*, 1355.
- [40] E. Bernatchez, M. J. Gold, A. Langlois, P. Blais-Lecours, M. Boucher, C. Duchaine, D. Marsolais,

K. M. McNagny, M.-R. Blanchet, *Physiol. Rep.* **2017**, *5*, e13163.

[41] H. Mbareche, M. Veillette, G. J. Bilodeau, C. Duchaine, **2018**, *1*.

[42] M. Lu, Y. Jin, R. Cerny, B. Ballmer-Weber, R. E. Goodman, *Food Chem. Toxicol.* **2018**, *216*, 207.

[43] J. A. Vizcaíno, A. Csordas, N. Del-Toro, J. A. Dienes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q.-W. Xu, R. Wang, H. Hermjakob, *Nucleic Acids Res.* **2016**, *44*, D447.