



A Small Effort for Researchers, a Big Gain for Soil Metaproteomics

Marco Chiapello^{1*†}, Elisa Zampieri^{2†} and Antonietta Mello^{1*}

¹ Institute for Sustainable Plant Protection, National Research Council, Turin, Italy, ² Council for Agricultural Research and Economics Research Centre for Cereal and Industrial Crops (CREA-CI), Vercelli, Italy

Keywords: soil metaproteomics, environment, data sharing, PRIDE, microbial communities

OPEN ACCESS

Edited by:

Biswarup Sen,
Tianjin University, China

Reviewed by:

V. L. S. Prasad Burra,
K L University, India
Jérôme Hamelin,
Institut National de la Recherche
Agronomique (INRA), France

*Correspondence:

Marco Chiapello
marco.chiapello@ipsp.cnr.it
Antonietta Mello
antonietta.mello@ipsp.cnr.it

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 15 October 2019

Accepted: 15 January 2020

Published: 04 February 2020

Citation:

Chiapello M, Zampieri E and Mello A
(2020) A Small Effort for Researchers,
a Big Gain for Soil Metaproteomics.
Front. Microbiol. 11:88.
doi: 10.3389/fmicb.2020.00088

The soil is one of the most complex systems where species belonging to different kingdoms live together (Young and Crawford, 2004). In particular, microorganisms such as bacteria and fungi are involved in nutrient cycling and organic matter transformation (Qian and Hettich, 2017): microbial community members with their activities determine nitrogen, sulfur, and carbon fluxes in the terrestrial subsurface (Hug et al., 2016). The soil microbiota with plant growth promoting rhizobacteria (PGPR), P-solubilizing bacteria, mycorrhizal-helping bacteria (MHB), and arbuscular mycorrhizal fungi (AMF) can be moreover involved in transfer and mobilization of trace elements, allowing bioremediation of heavy metal contaminated fields (Khan, 2005) or oil-polluted Alpine soils (Margesin, 2000). Microorganisms can be considered as tools to remove pollutants in soil, water, and sediments (Abatenh et al., 2017).

In order to understand the soil ecosystems and the biological processes that characterize it, it is necessary to study its microbial community composition, but also the metabolic activities performed by microbes (Siggins et al., 2012; Mello and Zampieri, 2017). The metagenomics advent made possible to identify the microorganism communities present in the soil (Vogel et al., 2009), while metaproteomics made possible to investigate the biological functions of these communities (Qian and Hettich, 2017). When the two approaches are applied to the same target it becomes possible to link microbial community composition to ecological processes, as performed by Zampieri et al. (2016) who tried to decipher the functioning of the brulé, the particular niche where a fungus in symbiosis with forest trees drives out the other symbiotic fungi. Recently, Martinez-Alonso et al. (2019) combined different -omic techniques (16S rRNA sequencing, culturomics, and metaproteomics) in order to identify microbial species and to clarify functions of microbial populations in the glacial ecosystem.

The study of the proteins expressed in an ecosystem at a specific time is a hot topic. Considering only soil ecosystem, using “metaproteomics AND soil” as topics on Web of Science website (<http://www.webofknowledge.com/>), it was possible to find 151 papers (as on September 15th 2019). Among the soil related papers according to Web of Science classification, there are: 36 reviews, 9 book chapters, 4 meetings, 2 abstracts, 1 letter, 1 early access, 71 others, and 148 articles.

Although the metaproteomics technique is more than 10 years old, it is still challenged both by technical and computational limitations. Humic acids and other contaminants which interfere with the protein extraction, render it highly dependent on the soil type. Different extraction methods can influence the observed metaproteome (Taylor and Williams, 2010; Becher et al., 2013; Zampieri et al., 2016; Mattarozzi et al., 2017; Keiblinger and Riedel, 2018). A strategy to overcome this obstacle is to use parallel different extraction methods and pool the extracted proteins before the

subsequent analysis. The absence of complete protein databases (Wilmes and Bond, 2006; Bastida et al., 2009; Siggins et al., 2012; Becher et al., 2013; Wilmes et al., 2015; Keiblinger et al., 2016; Wang et al., 2016; Heyer et al., 2017; Callister et al., 2018; Starke et al., 2019) limits the protein identification. A promising solution is to build in-house databases based on metagenomics data previously obtained from the same environment (Zampieri et al., 2016; Mattarozzi et al., 2017). The combination of metagenomics and metaproteomics has surely received some help by the recent advent of inexpensive high-throughput sequencing (Wilmes et al., 2015); by next generation sequencing it is in fact possible to obtain in less time more reads that can allow the organism identification and can create a starting point for building a database tailored for protein identification.

Despite all the limitations, the metaproteomics is a powerful technique to study the biological functions of microbial communities, to correlate the taxonomic and functional soil composition within the environment (Heyer et al., 2017) and also to evaluate the responses of microbial communities to climate change (i.e., global warming) (Liu et al., 2017). Moreover, soil protein identification could give information about the soil biogeochemical potential and pollutant degradation and be an indicator of soil quality (Bastida et al., 2019) and regeneration.

Looking at the 148 articles found in Web of Science website, a consistent number of studies has been done all around the world, on different soil types, such as agricultural, forest, contaminated by heavy metals, desert, riparian ones. The metaproteomic analyses were principally carried out in the Northern hemisphere (with the exception of few works in the Southern part of Africa). While in the USA the studied matrix is associated mainly with water, in Asia, basically in China, it is mostly associated with crops. Considering Europe, the studied soil types are heterogeneous, characterized by prevalence of forest and arid soils. The effects of contaminants are topic of different studies: in some of them contaminants were present in a natural way (e.g., heavy metals in serpentine soils), in others they have been human-introduced (e.g., petroleum).

The ability to compare studies could shed light on many soil processes, identify new insight, discover similar communities around the world, explore and understand the soil biodiversity, but this is at the moment very complicated due to lack of standards in the field both for the soil and the proteomics data. Even though many studies have reported the soil properties, the way to report them has not been consistent across the papers and there is no standard framework that researchers can use to compare similarities and differences between the studied soils. Comparative studies enable a deeper understanding of the soil physical and chemical properties. One interesting resource for comparison of studies is the Paleontology, Geobiology and Earth Archives Research Center (PANGEA) (<http://www.pangea.unsw.edu.au/research>) that has already emphasized the process of discovery and integration of ideas in different areas such as landscape evolution. This resource has allowed to create a standard framework, to understand the range of natural variability present in biological systems,

enhancing the capacity to discriminate natural cycles from recent human perturbations.

Concerning the proteomics data, currently not many scientists release upon publication the raw data underlying their experiment and on which they have built their conclusions. On the other hand, many publishers require as mandatory the raw data deposit in the guidelines, but not always this requirement is fulfilled by the authors and checked by the editors. The main repository for proteomics data, since 2012, is the proteomeXchange (PX) consortium (Vizcaino et al., 2014). The aims of the consortium are the data submission standardization and the dissemination of the proteomics data. Today the consortium includes different repositories from different countries and institutions, with different proteomic targets. The members of the consortium are: PRIDE, PASSEL, MassIVE, jPOST, iProX, Panorama Public, and Peptide Atlas and their main targets are: universal archive, Re-analysis, focused archive, Universal archive, Re-analysis, Universal archive, Universal archive, and focused archive (Deutsch et al., 2020), respectively. The submission process requires several details, including data and metadata. First of all, the authors have to provide the raw data (mandatory) and the derived peak list (optional). Secondly, experimental and technical metadata have to be provided; they are slightly different among the diverse members of the consortium but with a sufficient information to fulfill the requirements for the PX XML format file (http://www.proteomexchange.org/docs/guidelines_px.pdf). Finally, the processed results should be provided including the peptide and the protein identifications (mandatory) and quantification results (optional at present). Currently, two types of submission are supported: complete or partial. The former allows to connect, through PX resource, the identification data to the corresponding mass spectra. The latter provides all the submitted files for download, but it is not possible to parse, integrate and visualize the identification and/or connect the processed results to the corresponding mass spectra (Deutsch et al., 2020).

Concerning the proteomics data released, there are only 17 datasets related to “metaproteomics and soil” in PRIDE

TABLE 1 | Papers and raw data published in 2019: paper number results from Web of Science, PubMed and Scopus searches using as topics the two first columns.

Term 1	Term 2	Web of science	Pubmed	Scopus	Raw data (PRIDE)
Metaproteomics	Soil	15	10	13	3
Metaproteomics	Compost	1	0	0	0
Metaproteomics	Sediments	3	3	4	3
Metaproteomics	Gut	32	28	28	7
Metaproteomics	Water	7	4	10	6
Metaproteomics	Air	2	2	2	1
Metaproteomics	Feces	2	3	12	0
Metaproteomics	Lakes	0	1	1	1

Raw data numbers result from PRIDE query using the terms in the two first columns. All searches were limited to paper from 2019. The term “sediments” include both terrestrial and aquatic ones.

TABLE 2 | Checklist of standardized soil metadata according to Schoeneberger et al. (2012).

Sample		Sampling time		Location		Element content		
Name	Replicate	Date	GPS coord	Altitude	Place name	C%	N%	pH
Soil		Land coverage			Vegetation			
Clay %— sand %— silt %	Soil classification	Contaminants	Moisture	Temperature regime	Type of land coverage	Plant common names	Plant scientific names	Vegetation coverage %
Collection methods			Storage system		Note			
Sampling method	Depth of sampling	Unit	Soil treatment (sieved...)	Storage temperature	Container type	Additional information		

(<https://www.ebi.ac.uk/pride/archive/>) (Martens et al., 2005), among them 11 are associated to a paper, while seven are deposited without any link to a published paper, meaning that only a very small proportion of the authors deposited the raw data in public repositories. The sampling sites related to the data deposited on PRIDE span from Spain to Antarctica and from Northern California to Sweden, indicating the absence of geographical bias and a great variability as spatial coordinates. The act of publishing without depositing the raw data limits other researchers from performing comparative studies. We strongly believe that this restricts the progress, discovering better insights and potential application of metaproteomics field. Publishing the raw data should be made mandatory, both for open and reproducible science, and for allowing the data reuse and exploring new insights. The lack of raw data on repositories is a problem that also concerns other fields, such as studies carried out on gut, water and so on, as shown by the outputs provided by a search of these studies on Web of Science, Pubmed, and Scopus databases and of their deposited raw data on PRIDE (Table 1).

The aim of this opinion is not only to report the low percentage of dataset related to the soil metaproteomic studies, but also to point out the still concealed potential of the technique if flanked by a proper repository of data. The possibility of comparing studies could shed light on many soil processes, but this is very complicated due to a lack of shared proteomics data. Only 6 (5%; 3 out of the 9 papers

cited on PRIDE are not found on WOS) out of 148 studies uploaded the raw data to PRIDE; moreover the remaining studies not always included in the supplementary materials the complete list of identified proteins. On the other hand, many studies provided detailed information about the soil composition, but unfortunately, not in a standardized way. The lack of shared proteomics data, and at the same time the lack of standard metadata on the soil composition, render the different comparative studies a complicated challenge. Owing on the previous considerations, we strongly advise the metaproteomics community to adopt standardized soil metadata, to publish the raw data on PRIDE and to follow the procedures pointed out by ProteomeXchange Consortium. Standardized soil metadata could follow the checklist (Table 2) we have extrapolated by the useful guide for soil describing and sampling, proposed by Schoeneberger et al. (2012).

These two simple good practices will massively increase the ability to compare studies and carry out bioinformatic analyses, using already published data. Although this opinion focuses on soil metaproteomics data, we hope it will ring a bell for scientists involved in other disciplines and ecosystems.

AUTHOR CONTRIBUTIONS

MC, EZ, and AM conceived the idea. MC and EZ designed the structure of the manuscript and drafted it. AM has critically read, corrected, and covered the costs to publish in open access.

REFERENCES

- Abatenh, E., Gizaw, B., Tsegaye, Z., and Wassie, M. (2017). Application of microorganisms in bioremediation-review. *J. Environ. Microbiol.* 1, 2–9. doi: 10.17352/ojeb.000007
- Bastida, F., Jehmlich, N., Martinez-Navarro, J., Bayona, V., Garcia, C., and Moreno, J. L. (2019). The effects of struvite and sewage sludge on plant yield and the microbial community of a semiarid Mediterranean soil. *Geoderma* 337, 1051–1057. doi: 10.1016/j.geoderma.2018.10.046
- Bastida, F., Moreno, J. L., Nicolas, C., Hernandez, T., and Garcia, C. (2009). Soil metaproteomics: a review of an emerging environmental science. Significance, methodology and perspectives. *Eur. J. Soil Sci.* 60, 845–859. doi: 10.1111/j.1365-2389.2009.01184.x
- Becher, D., Bernhardt, J., Fuchs, S., and Riedel, K. (2013). Metaproteomics to unravel major microbial players in leaf litter and soil environments: challenges and perspectives. *Proteomics* 13, 2895–2909. doi: 10.1002/pmic.201300095
- Callister, S. J., Fillmore, T. L., Nicora, C. D., Shaw, J. B., Purvine, S. O., Orton, D. J., et al. (2018). Addressing the challenge of soil metaproteome complexity by improving metaproteome depth of coverage through two-dimensional liquid chromatography. *Soil Biol. Biochem.* 125, 290–299. doi: 10.1016/j.soilbio.2018.07.018
- Deutsch, E. W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J. J., Kundu, D. J., et al. (2020). The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Res.* 48, D1145–D1152. doi: 10.1093/nar/gkz984

- Heyer, R., Schallert, K., Zoun, R., Becher, B., Saake, G., and Benndorf, D. (2017). Challenges and perspectives of metaproteomic data analysis. *J. Biotechnol.* 261, 24–36. doi: 10.1016/j.jbiotec.2017.06.1201
- Hug, L. A., Thomas, B. C., Sharon, I., Brown, C. T., Sharma, R., Hettich, R. L., et al. (2016). Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. *Environ. Microbiol.* 18, 159–173. doi: 10.1111/1462-2920.12930
- Keiblinger, K. M., Fuchs, S., Zechmeister-Boltenstern, S., and Riedel, K. (2016). Soil and leaf litter metaproteomics—a brief guideline from sampling to understanding. *FEMS Microbiol. Ecol.* 92:fiw180. doi: 10.1093/femsec/fiw180
- Keiblinger, K. M., and Riedel, K. (2018). Sample preparation for metaproteome analyses of soil and leaf litter. *Methods Mol. Biol.* 1841, 303–318. doi: 10.1007/978-1-4939-8695-8_21
- Khan, A. G. (2005). Role of soil microbes in the rhizospheres of plants growing on trace metal contaminated soils in phytoremediation. *J. Trace Elements Med. Biol.* 18, 355–364. doi: 10.1016/j.jtemb.2005.02.006
- Liu, D., Keiblinger, K. M., Schindlbacher, A., Wegner, U., Sun, H., Fuchs, S., et al. (2017). Microbial functionality as affected by experimental warming of a temperate mountain forest soil—A metaproteomics survey. *Appl. Soil Ecol.* 117, 196–202. doi: 10.1016/j.apsoil.2017.04.021
- Margesin, R. (2000). Potential of cold-adapted microorganisms for bioremediation of oil-polluted Alpine soils. *Int. Biodeterior. Biodegrad.* 46, 3–10. doi: 10.1016/S0964-8305(00)00049-4
- Martens, L., Hermjakob, H., Jones, P., Adamski, M., Taylor, C., States, D., et al. (2005). PRIDE: the proteomics identifications database. *Proteomics* 5, 3537–3545.
- Martinez-Alonso, E., Pena-Perez, S., Serrano, S., Garcia-Lopez, E., Alcazar, A., and Cid, C. (2019). Taxonomic and functional characterization of a microbial community from a volcanic englacial ecosystem in Deception Island, Antarctica. *Sci. Rep.* 9:12158. doi: 10.1038/s41598-019-47994-9
- Mattarozzi, M., Manfredi, M., Montanini, B., Gosetti, F., Sanangelantoni, A. M., Marengo, E., et al. (2017). A metaproteomic approach dissecting major bacterial functions in the rhizosphere of plants living in serpentine soil. *Anal. Bioanal. Chem.* 409, 2327–2339. doi: 10.1007/s00216-016-0175-8
- Mello, A., and Zampieri, E. (2017). Who is out there? What are they doing? Application of metagenomics and metaproteomics to reveal soil functioning. *Ital. J. Mycol.* 46, 1–7. doi: 10.6092/issn.2531-7342/6647.
- Qian, C., and Hettich, R. L. (2017). Optimized extraction method to remove humic acid interferences from soil samples prior to microbial proteome measurements. *J. Proteome Res.* 16, 2537–2546. doi: 10.1021/acs.jproteome.7b00103
- Schoeneberger, P. J., Wysocki, D. A., and Benham, E. C. (2012). *Field Book for Describing and Sampling Soils, Version 3.0 | NRCS Soils*. Lincoln: Natural Resources Conservation Service; National Soil Survey Center.
- Siggins, A., Gunnigle, E., and Abram, F. (2012). Exploring mixed microbial community functioning: recent advances in metaproteomics. *FEMS Microbiol. Ecol.* 80, 265–280. doi: 10.1111/j.1574-6941.2011.01284.x
- Starke, R., Jehmlich, N., and Bastida, F. (2019). Using proteins to study how microbes contribute to soil ecosystem services: the current state and future perspectives of soil metaproteomics. *J. Proteomics* 198, 50–58. doi: 10.1016/j.jprote.2018.11.011
- Taylor, E. B., and Williams, M. A. (2010). Microbial protein in soil: influence of extraction method and C amendment on extraction and recovery. *Microb. Ecol.* 59, 390–399. doi: 10.1007/s00248-009-9593-x
- Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Rios, D., et al. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* 32, 223–226. doi: 10.1038/nbt.2839
- Vogel, T. M., Hirsch, P. R., Simonet, P., Jansson, J. K., Tiedje, J. M., van Elsas, J. D., et al. (2009). Advantages of the metagenomic approach for soil exploration: reply from Vogel et al. *Nat. Rev. Microbiol.* 7, 756–757. doi: 10.1038/nrmicro2119-c3
- Wang, D.-Z., Kong, L.-F., Li, Y.-Y., and Xie, Z.-X. (2016). Environmental microbial community proteomics: status, challenges and perspectives. *Int. J. Mol. Sci.* 17:E1275. doi: 10.3390/ijms17081275
- Wilmes, P., and Bond, P. (2006). Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends Microbiol.* 14, 92–97. doi: 10.1016/j.tim.2005.12.006
- Wilmes, P., Heintz-Buschart, A., and Bond, P. L. (2015). A decade of metaproteomics: where we stand and what the future holds. *Proteomics* 15, 3409–3417. doi: 10.1002/pmic.201500183
- Young, I. M., and Crawford, J. W. (2004). Interactions and self-organization in the soil-microbe complex. *Science* 304, 1634–1637. doi: 10.1126/science.1097394
- Zampieri, E., Chiapello, M., Daghino, S., Bonfante, P., and Mello, A. (2016). Soil metaproteomics reveals an inter-kingdom stress response to the presence of black truffles. *Sci. Rep.* 6:25773. doi: 10.1038/srep25773

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Chiapello, Zampieri and Mello. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.