



Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities

 Manuel Kleiner^a

^aDepartment of Plant and Microbial Biology, North Carolina State University, Raleigh, North Carolina, USA

ABSTRACT Metaproteomics is the large-scale identification and quantification of proteins from microbial communities and thus provides direct insight into the phenotypes of microorganisms on the molecular level. Initially, metaproteomics was mainly used to assess the “expressed” metabolism and physiology of microbial community members. However, recently developed metaproteomic tools allow quantification of per-species biomass to determine community structure, *in situ* carbon sources of community members, and the uptake of labeled substrates by community members. In this perspective, I provide a brief overview of the questions that we can currently address, as well as new metaproteomics-based approaches that we and others are developing to address even more questions in the study of microbial communities and plant and animal microbiota. I also highlight some areas and technologies where I anticipate developments and potentially major breakthroughs in the next 5 years and beyond.

KEYWORDS LC-MS, metagenomics, metaproteomics, metatranscriptomics, microbiome, microbiota, proteomics, symbiosis

I recently taught a class entitled “Microbial Symbiosis and Microbiomes.” After discussing the myriad of approaches for studying microbial communities and host-microbe interactions, I provided an overview of a personal favorite—metaproteomics. The students seemed enthusiastic about the unique capabilities of metaproteomics, but one student raised a valid question: “If metaproteomics is such a great approach, why aren’t more scientists using it?”

Since its inception in 2004 as “the large-scale characterization of the entire protein complement of environmental microbiota at a given point in time” (1), only ~500 publications in PubMed include “metaproteom*” (compared to >10,000 for metagenomics). Since 2012, the number of publications that mention metaproteomics has grown exponentially, partly due to quantum leaps in enabling technologies that have made metaproteomics more feasible and affordable. Technological advances have mostly occurred in the realm of liquid chromatography (LC) enabling separation of highly complex peptide mixtures, high-resolution mass spectrometry (MS) enabling acquisition of large numbers of accurate mass spectra, and computational tools for data processing and analyses. While in 2004 it was standard practice to separate proteins on a 2D gel and then manually pick individual protein spots for mass spectrometric analyses (a laborious and low-throughput method), we now routinely employ LC-MS/MS approaches to identify and quantify tens of thousands of peptides and >10,000 proteins per sample. The underlying advances in liquid chromatography and high-resolution mass spectrometry have also driven the development of new metaproteomic approaches that enable researchers to address a whole array of questions (Fig. 1).


Citation Kleiner M. 2019. Metaproteomics: much more than measuring gene expression in microbial communities. *mSystems* 4:e00115-19. <https://doi.org/10.1128/mSystems.00115-19>.

Copyright © 2019 Kleiner. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to manuel_kleiner@ncsu.edu.

Conflict of Interest Disclosures: M.K. has nothing to disclose.

mSystems[®] vol. 4, no. 3, is a special issue sponsored by *Illumina*.

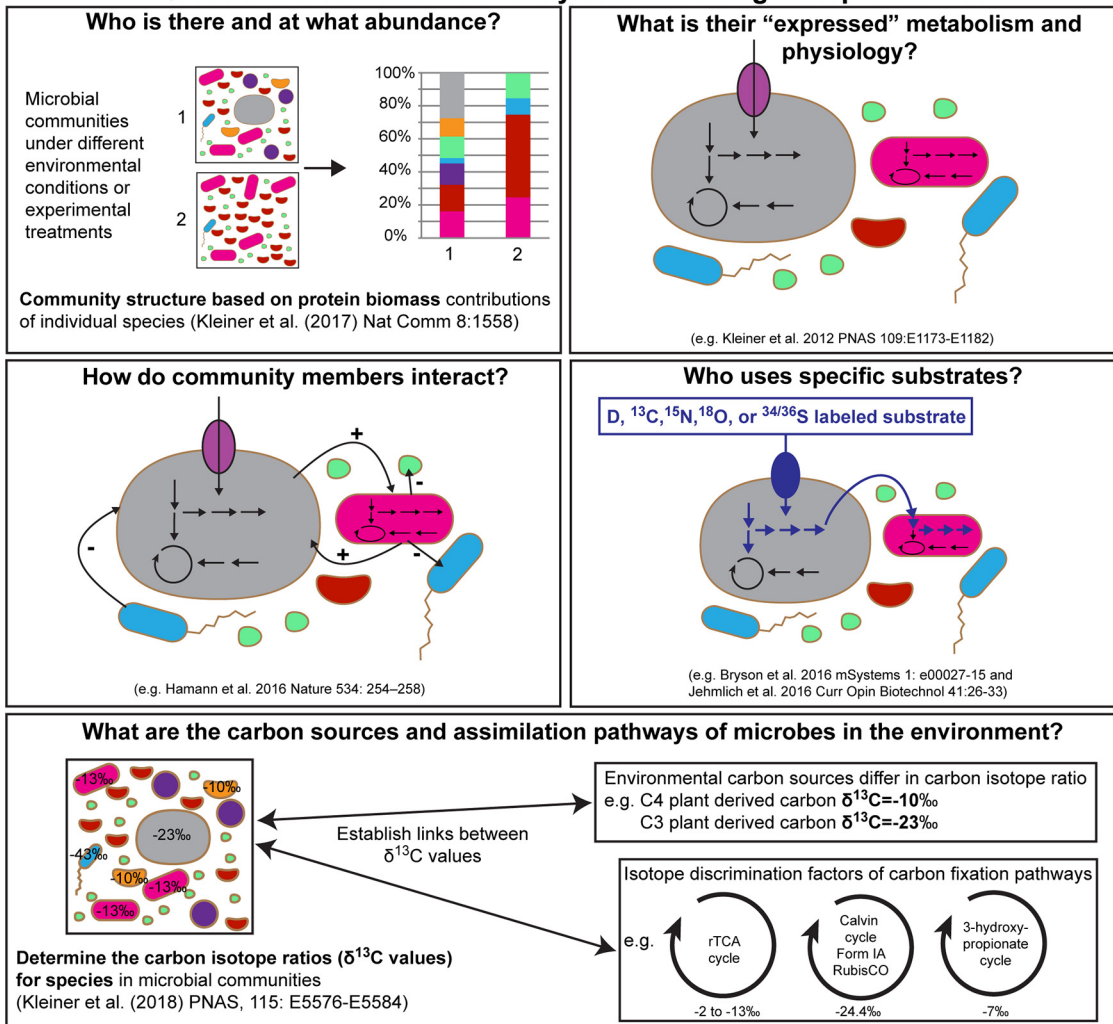
 #metaproteomics allows us to address a large diversity of questions in the study of microbial communities and microbiota

Received 15 February 2019

Accepted 22 March 2019

Published 21 May 2019

Questions that we can currently address using metaproteomics



Additional questions that we should be able to address in the future (methods development needed)

- (1) Who takes up, metabolizes and transfers isotopically labeled substrates? (advancing Protein-SIP to work with very small amounts of label and/or short labeling pulses)
- (2) What is the growth rate of individual community members?
- (3) Which proteins exhibit a specific targeted activity? (mass spectrometric detection of activity-based protein profiling (ABPP) probes)
- (4) What are the abundances of environmental viruses? (highly sensitive and quantitative metaproteomics to detect environmental viruses)
- (5) What are the ages and the roles of cell free ("relic") proteins in the environment? (differentiation of cell bound versus extracellular protein and determination of relative age)

FIG 1 Questions that can or will be addressed with metaproteomics.

QUESTIONS THAT CAN BE ADDRESSED ONLY THROUGH METAPROTEOMICS

The aim of this perspective is not to convince the reader that metaproteomics is the best or only approach to study what transpires in microbial communities. The aim is to highlight the strengths of metaproteomics, which integrates with and often relies on other approaches. For example, building a reference database through metagenomic sequencing is often a prerequisite for metaproteomics (2).

In its basic form, metaproteomics allows us to study the presence and abundances of proteins in any microbial community. With a well-curated protein sequence database

in hand, we can assign these proteins to individual species or higher taxa and understand the functional roles and interactions of individual members in the community. Since proteins convey structure and activities to cells, knowing their abundances provides a picture of cellular phenotypes on the molecular level. Pure culture studies from a diversity of organisms have shown that the abundance of a protein indicates its relevance and activity under a given condition. Applying this at the community level can lead to critical insights and significant progress. For example, in a marine worm that relies on its five bacterial symbionts for all nutrition, we did not know which environmental energy sources the symbionts were using for carbon fixation. A metaproteomic study revealed that some of the symbionts abundantly expressed carbon monoxide dehydrogenases, which suggested they were using carbon monoxide as an energy source (3). We were subsequently able to confirm this hypothesis through physiological incubation experiments and nanoscale secondary-ion mass spectrometry (NanoSIMS) measurements (4).

Differential metaproteomics can be used to identify changes in the expression of individual genes, for example, after actively manipulating a microbial community by changing substrates or if environmental conditions naturally fluctuate. We have used this approach to show that hydrogen transfer is the basis of the symbiosis between a *Breviatea* protist and its *Arcobacter* symbiont by comparing metaproteomes from both partners to those of only one partner (5).

The metaproteomic data generated to examine function can also be used to analyze community structure. Protein constitutes the largest amount of cellular material (in most organisms), and thus, total per-species protein can be quantified to assess biomass contributions of individual community members. We recently developed and validated such an approach, which uses standard metaproteomic data to analyze community structure on the basis of biomass instead of gene/genome copy counts (e.g., 16S rRNA amplicon or metagenome sequencing) (6).

If the discovery of uncharacterized proteins with specific enzyme functions is of interest, activity-based probes (ABP) can be used to enrich proteins with specific functions from environmental samples prior to metaproteomic analysis. This approach has, for example, led to the discovery of overrepresented microbial proteases in the intestinal microbiota of mice with inflammatory bowel disease (7).

The high-resolution mass spectrometry data produced by most metaproteomics approaches can also be used to analyze the isotope content of individual proteins and species. Based on this, we have developed a highly sensitive protein stable isotope fingerprinting (direct Protein-SIF) method to analyze natural ratios of carbon stable isotopes, which allows one to determine carbon sources of individual microorganisms in microbial communities, as well as carbon assimilation pathways used by autotrophic organisms (8). Others have used similar approaches to follow the incorporation of isotopically labeled substrates by individual community members using a metaproteomic approach called protein-based stable isotope probing (Protein-SIP) (9, 10). These approaches have much higher resolution and throughput than other strategies that enable isotopic analyses of environmental microorganisms such as NanoSIMS and DNA-SIP.

QUESTIONS THAT WE PLAN TO ADDRESS IN THE FUTURE

Despite metaproteomics having been around for >15 years, it is still in its infancy and there is great potential for its continued development to address otherwise intractable questions.

In my laboratory, we develop and improve metaproteomic methods to study animal- and plant-associated microbiota with a focus on understanding critical interactions between all partners, including the host. Such studies are challenging for multiple reasons, including the presence of fecal or soil-derived substances that interfere with sample preparation and the high abundances of host-associated proteins that potentially swamp out the microbial signal. In the past we have addressed these challenges by applying prefractionation methods such as density gradient centrifuga-

tions and filtrations (3, 11). However, sample preparation and analytical methods have improved so that in most cases samples need not be prefractionated and we can analyze microbial and host gene expression within the same sample. Future improvements in metaproteomic coverage are expected to come with further technological advances (see below).

Together with our collaborators, we are developing a variety of approaches that rely on metaproteomics to address novel questions. These approaches include (i) following the flow of carbon, nitrogen, oxygen, and hydrogen in microbial communities by using highly sensitive Protein-SIP to measure their incorporation into proteins. The high sensitivity will allow working with substrates where only a small fraction is labeled with heavier isotopes rather than the larger fractions that were necessary with earlier versions of this method. The approaches also include (ii) estimating growth rates of community members based on protein turnover rates, (iii) identifying uncharacterized proteins with specific enzymatic functions (as determined by binding of highly specific probes) using activity-based protein profiling with mass spectrometric detection of the probes in metaproteomics, (iv) detecting environmental viruses with higher sensitivity, and (v) differentiating cell-bound from extracellular protein (secreted and relic proteins) in environmental samples. The last approach will be critical to better understand the role of protein in a diversity of environments, as it can be expected that, similar to relic DNA (12), relic/extracellular protein may influence our measurements of environmental gene expression. In contrast to DNA, however, understanding extracellular protein is critical not only from a perspective of impacting our measurements. If it retains activity, this protein could have major impacts on biogeochemical cycling (13).

FUTURE AREAS FOR DEVELOPMENT AND POTENTIAL BREAKTHROUGHS

In my opinion, there are several developments that will drive metaproteomics advances in the next 5 years and beyond. These include the appearance of new technologies, training additional personnel on metaproteomics approaches, validating and standardizing approaches, developing custom computational tools, and establishing dedicated metaproteomics meetings.

One of the technological factors that limits metaproteomics in terms of throughput is the number of mass spectra that can be acquired by an instrument per unit of time. Currently, most instruments measure, at most, 40,000 spectra per hour, which necessitates several hours of instrument time for an in-depth community analysis. New generations of mass spectrometers with increased measurement frequency will lead to increased metaproteomic throughput. For example, a recent study demonstrated the measurement of 600,000 spectra within 1 h on an instrument that combines ion mobility and time of flight mass spectrometry (14).

Technological innovation could also come from a completely different direction. New technologies are being developed to directly sequence single protein molecules by an approach similar to the DNA sequencing technology developed by Oxford Nanopore (15). Direct sequencing of intact protein molecules could be a game changer, particularly for the difficult task of distinguishing proteins with similar sequences. Protein sequencing would eliminate our reliance on a so-called bottom-up (metaproteomics) approach in which the identification of protein fragments (peptides) is used to ultimately infer the presence of proteins. Currently, there is no adequate mass spectrometric approach to analyze intact proteins (top-down approach) from complex mixtures, and this will likely not change in the foreseeable future.

The number of research labs developing or using metaproteomics has increased steadily over the last few years, and this has led to an increased number of students and postdocs acquiring the skills necessary to generate and analyze metaproteomic data sets. The growth in the metaproteomic community has also led to an increased demand for bioinformatics tools specifically designed to work with these data. While most data can be analyzed with standard proteomic tools, the emergence of tools specifically developed for metaproteomic data, such as MetaProteomAnalyzer (16), Unipept (17), and Calis-p (8), is promising, as is the integration of metaproteomic data

with multi-omic data sets (e.g., references 18 and 19). This trend is expected to continue, and more tools will become available.

The growth of the community has also led us to recognize the need for better validation and standardization of metaproteomic approaches. Currently, there are as many approaches for metaproteomics as there are labs doing metaproteomics, and it is often unclear in publications which approaches work better for particular questions or sample types. Although a few recent articles used fully controlled samples to validate and optimize metaproteomic approaches (6, 20, 21), much more effort is needed in this realm to make metaproteomics more broadly applicable and comparable across laboratories. Fortunately, the increase in the number of labs that do metaproteomics also means that the critical mass was reached a few years ago to establish a dedicated conference, which facilitates the exchange of ideas and discussion on how to standardize and compare approaches. The most recent meeting was the 3rd International Metaproteome Symposium hosted in December 2018 in Leipzig by a team from the Centre for Environmental Research (UFZ). The next meeting will be in summer 2020 hosted by the group of Paul Wilmes in Luxembourg. I look forward to meeting many new metaproteomics enthusiasts there!

ACKNOWLEDGMENTS

I thank Nicole Webster and Jack Gilbert for the opportunity to contribute to this special issue. I am grateful to Heather Maughan and Angela Kouris for feedback on the manuscript. I sincerely apologize to all of my colleagues whose work could not be cited because of reference number limitations.

The metaproteomics work in the Kleiner Lab is supported by the NC State Chancellor's Faculty Excellence Program Cluster on Microbiomes and Complex Microbial Communities, by grants from the Plant Soil Microbial Community Consortium, and by the Foundation for Food and Agriculture Research (grant ID 593607).

REFERENCES

1. Wilmes P, Bond PL. 2004. The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ Microbiol* 6:911–920. <https://doi.org/10.1111/j.1462-2920.2004.00687.x>.
2. Tanca A, Palomba A, Fraumene C, Pagnozzi D, Manghina V, Deligios M, Muth T, Rapp E, Martens L, Addis MF, Uzzau S. 2016. The impact of sequence database choice on metaproteomic results in gut microbiota studies. *Microbiome* 4:51. <https://doi.org/10.1186/s40168-016-0196-8>.
3. Kleiner M, Wentrup C, Lott C, Teeling H, Wetzal S, Young J, Chang Y-J, Shah M, VerBerkmoes NC, Zarzycki J, Fuchs G, Markert S, Hempel K, Voigt B, Becher D, Liebeke M, Lalk M, Albrecht D, Hecker M, Schweder T, Dubilier N. 2012. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc Natl Acad Sci U S A* 109:E1173–E1182. <https://doi.org/10.1073/pnas.1121198109>.
4. Kleiner M, Wentrup C, Holler T, Lavik G, Harder J, Lott C, Littmann S, Kuypers MMM, Dubilier N. 2015. Use of carbon monoxide and hydrogen by a bacteria-animal symbiosis from seagrass sediments. *Environ Microbiol* 17:5023–5035. <https://doi.org/10.1111/1462-2920.12912>.
5. Hamann E, Gruber-Vodicka H, Kleiner M, Tegetmeyer HE, Riedel D, Littmann S, Chen J, Milucka J, Viehweger B, Becker KW, Dong X, Stairs CW, Hinrichs K-U, Brown MW, Roger AJ, Strous M. 2016. Environmental Breviatea harbour mutualistic *Arcobacter* epibionts. *Nature* 534:254–258. <https://doi.org/10.1038/nature18297>.
6. Kleiner M, Thorson E, Sharp CE, Dong X, Liu D, Li C, Strous M. 2017. Assessing species biomass contributions in microbial communities via metaproteomics. *Nat Commun* 8:1558. <https://doi.org/10.1038/s41467-017-01544-x>.
7. Mayers MD, Moon C, Stupp GS, Su AI, Wolan DW. 2017. Quantitative metaproteomics and activity-based probe enrichment reveals significant alterations in protein expression from a mouse model of inflammatory bowel disease. *J Proteome Res* 16:1014–1026. <https://doi.org/10.1021/acs.jproteome.6b00938>.
8. Kleiner M, Dong X, Hinzke T, Wippler J, Thorson E, Mayer B, Strous M. 2018. Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. *Proc Natl Acad Sci U S A* 115:E5576–E5584. <https://doi.org/10.1073/pnas.1722325115>.
9. Jehmlich N, Vogt C, Lünsmann V, Richnow HH, von Bergen M. 2016. Protein-SIP in environmental studies. *Curr Opin Biotechnol* 41:26–33. <https://doi.org/10.1016/j.copbio.2016.04.010>.
10. Bryson S, Li Z, Pett-Ridge J, Hettich RL, Mayali X, Pan C, Mueller RS. 2016. Proteomic stable isotope probing reveals taxonomically distinct patterns in amino acid assimilation by coastal marine bacterioplankton. *mSystems* 11:e00027–15. <https://doi.org/10.1128/mSystems.00027-15>.
11. Hinzke T, Kleiner M, Markert S. 2018. Centrifugation-based enrichment of bacterial cell populations for metaproteomic studies on bacteria-invertebrate symbioses, p 319–334. *In* Becher D (ed), *Microbial proteomics: methods and protocols*. Springer, New York, NY.
12. Carini P, Marsden PJ, Leff JW, Morgan EE, Strickland MS, Fierer N. 2016. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat Microbiol* 2:16242. <https://doi.org/10.1038/nmicrobiol.2016.242>.
13. Baltar F. 2018. Watch out for the “living dead”: cell-free enzymes and their fate. *Front Microbiol* 8:2438. <https://doi.org/10.3389/fmicb.2017.02438>.
14. Meier F, Brunner A-D, Koch S, Koch H, Lubeck M, Krause M, Goedecke N, Decker J, Kosinski T, Park MA, Bache N, Hoerning O, Cox J, Räther O, Mann M. 2018. Online parallel accumulation-serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol Cell Proteomics* 17:2534–2545. <https://doi.org/10.1074/mcp.TIR118.000900>.
15. Chinappi M, Ceconi F. 2018. Protein sequencing via nanopore based devices: a nanofluidics perspective. *J Phys Condens Matter* 30:204002. <https://doi.org/10.1088/1361-648X/aababe>.
16. Muth T, Kohrs F, Heyer R, Benndorf D, Rapp E, Reichl U, Martens L, Renard BY. 2018. MPA portable: a stand-alone software package for analyzing metaproteome samples on the go. *Anal Chem* 90:685–689. <https://doi.org/10.1021/acs.analchem.7b03544>.
17. Gurdeep Singh R, Tanca A, Palomba A, Van der Jeugt F, Verschaffelt P,

- Uzzau S, Martens L, Dawyndt P, Mesuere B. 2019. Unipept 4.0: functional analysis of metaproteome data. *J Proteome Res* 18:606–615. <https://doi.org/10.1021/acs.jproteome.8b00716>.
18. Heintz-Buschart A, May P, Laczny CC, Lebrun LA, Bellora C, Krishna A, Wampach L, Schneider JG, Hogan A, de Beaufort C, Wilmes P. 2016. Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nat Microbiol* 2:16180. <https://doi.org/10.1038/nmicrobiol.2016.180>.
19. Blank C, Easterly C, Gruening B, Johnson J, Kolmeder CA, Kumar P, May D, Mehta S, Mesuere B, Brown Z, Elias JE, Hervey WJ, McGowan T, Muth T, Nunn BL, Rudney J, Tanca A, Griffin TJ, Jagtap PD. 2018. Disseminating metaproteomic informatics capabilities and knowledge using the Galaxy-P framework. *Proteomes* 6:7. <https://doi.org/10.3390/proteomes6010007>.
20. Tanca A, Palomba A, Deligios M, Cubeddu T, Fraumene C, Biosia G, Pagnozzi D, Addis MF, Uzzau S. 2013. Evaluating the impact of different sequence databases on metaproteome analysis: insights from a lab-assembled microbial mixture. *PLoS One* 8:e82981. <https://doi.org/10.1371/journal.pone.0082981>.
21. Hinzke T, Kouris A, Hughes R-A, Strous M, Kleiner M. 2019. More is not always better: evaluation of 1D and 2D-LC-MS/MS methods for metaproteomics. *Front Microbiol* 10:238. <https://doi.org/10.3389/fmicb.2019.00238>.