Bridging the Gap Between Analytical and Microbial Sciences in Microbiome Research

Robert A. Quinn[1], Kehau A. Hagiwara[2], Ken Liu[3], Maryam Goudarzi[4], Ewy Mathe[5], Wimal Pathmasiri6, Lloyd W. Sumner[7] and Thomas O. Metz[8]

[1]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI
[2]Chemical Sciences Division, National Institute of Standards and Technology, Hollings Marine Laboratory, Charleston, SC
[3]Clinical Biomarkers Laboratory, Emory University, Atlanta, GA
[4]Cleveland Clinic Lerner Research Institute, Cleveland, OH
[5]National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD
[6]Nutrition Research Institute, University of North Carolina at Chapel Hill, 500 Laureate Way, Kannapolis, NC
[7]Department of Biochemistry, University of Missouri Life Sciences Center, Columbia, MO
[8]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA

Abstract: Metabolites produced by the microbiome influence human, animal, and environmental health, but the diversity and functional roles of these compounds have only begun to be elucidated. Comprehensively, profiling, identifying, and quantifying these molecules are significant challenges, as it requires expertise in analytical methods, such as mass spectrometry and nuclear magnetic resonance spectroscopy, skills that not many traditional microbiologists or microbial ecologists possess. This creates a gap between microbiome scientists that want to contextualize the role of microbial metabolites in systems biology, and the skills required to generate and interpret complex metabolomics datasets. To bridge this gap, microbiome scientists must engage analytical scientists to best understand the underlying chemical and physical principles of the data. Conversely, analytical scientists must engage with microbiome scientists to better understand the biological questions being asked with metabolomics data and to best communicate its intricacies. This scientific dialogue is most beneficial if it begins at a project's inception and is maintained throughout the analysis steps. There is also a co-evolving need for cheminformatic information exchange to bridge the gap between the instrumental data and biological interpretation. Simple raw or minimally processed data dumps are insufficient for most microbiome scientists to interpret. Collaborative data translation must occur to ensure that biologists appropriately interpret the complex spectral information, which is often rife with misinterpretation and ambiguous annotation. This two-way engagement must occur at the level of basic scientists, group leaders, and institutional core centers that generate both the samples and the analytical data. Better communication across the chemistry/biology disciplines will further enable the understanding of the 'dark matter' within microbiomes that maintain healthy humans and healthy environments.

Main Text

The chemical language of microbes is almost infinitely diverse. By this very nature, we poorly understand the molecules that make up the text of their metabolic narratives. Many tools are available to decipher the microbe-microbe and microbe-host chemical interactions, with some of the most powerful being mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy. These highly advanced and technical instruments are the *Rosetta Stones* of microbial chemical ecology, because they enable translation of these chemical languages. Many impactful discoveries have been made with these tools demonstrating how microbial

chemistry promotes health, disease, immunity and metabolism of xenobiotics in both human and environmental systems (1–5). Newly evolving technologies, such as metabolomics, enable more comprehensive assessments of the chemicals within a biological system. The aim of metabolomics is the large-scale quantitative and qualitative characterization of the small molecules present in a biological sample, which represent the functional outputs of the microbiome, its host and/or its environment. However, challenges exist in interpreting the complex and highly technical data that metabolomics analyses provide. These challenges are becoming more pronounced as the capacity to generate in-depth MS and NMR data on microbial metabolomes grows. While it is becoming routine to capture this chemical information with high degrees of accuracy, interpretation of its biological meaning, particularly relating to microbiomes, requires extensive insight and some degree of validation of the chemical species identified. Chemical ambiguity can be problematic in microbiome science, as it is the actual chemical structure that determines the function of a microbial metabolite. For example, chemical changes a subtle as unique epimers of bile acids induced by the microbiome can have dramatic effects on host immunity (2). Identification of these microbiome-dependent or microbiome-altering metabolites and their biological activities continues to be an active area of research, since the identities of many microbial metabolites have not yet been elucidated (6–8) and the biological context begins with metabolite identification. Thus, there is dire need for an in-depth dialogue between analytical scientists who generate the metabolomics data and the microbiome scientists who aim to interpret it.

The concept of metabolite identification and quantification exists on a spectrum of certainty.

An important concept in analytical chemistry and metabolomics is that there is a spectrum of certainty regarding the annotation and quantification of all metabolites, depending on the methods and analytical approaches used. Biologists are cautioned to not assume that a molecule identified from even the best metabolomics informatics pipelines is a known compound without some further validation. Confidence in chemical identification is dependent on the analytical platform, quality of the data and access to chemical standards. Analytical platforms (NMR or MS) employ various mechanisms for distinguishing one chemical from another (selectivity) and have differing ability to detect lower abundance chemicals (sensitivity). Depending on the metabolomics platform, metabolite identification can be highly certain to merely a marginal association, and biological interpretation will depend on this degree of analytical (un)certainty (Fig. 1). Traditionally, unambiguous chemical identification requires matching an authentic standard with experimentally-derived spectra and associated retention times, drift times, or chemical shifts (9). The use of in-house libraries derived from analyses of authentic reference standards (acquired under identical analytical conditions as study samples) is the preferred approach for high confidence identifications. However, authentic microbial standards are often not available for recently discovered chemicals or metabolic pathways, which limits the ability to validate the identity of these molecules and their associated functions. In some cases, such as in the absence of authentic standards, there is an additional need for the synthesis or isolation and co-characterization of metabolites to confirm the identity. These validation approaches are vital for reproducible microbiome research and dissemination of metabolite data from microbiomes across laboratories.

However, extensive validation of a chemical's structure with labor intensive analytical rigor may not always be required depending upon the goals of a microbiome scientist. Microbiome scientists can obtain valuable information about a biological system without the need to identify, for example, the stereochemistry of a particular chemical group. Instead, annotation at the molecular family level can still be valuable, as one can assess overall chemical shifts from a host or environmental perturbations. Microbiome scientists can further calculate diversity measures (both alpha- and beta-diversity) using metabolomics data, whether

the metabolites measured are annotated or not, and these metrics can reveal important biological phenomena, such as resistance and resilience of a microbiome system. Recent advances in MS data analyses have furthered the biological information that can be mined from spectral data without knowledge of a metabolite annotation, such as molecular networking (10) and capture of chemical mass shifts between related molecules (11). Herein lies the need for dialogue between microbiome scientists and analytical scientists to learn from one another about what can be harnessed from metabolomics data and how to best interpret that information.

Similarly, quantification of metabolites in microbiome studies can be done from the level of precise concentrations to relative changes in abundance. This too must be interpreted appropriately. It is important for the microbiome scientists to know that accurately quantifying a compound comes at the cost of measuring only a few compounds at a time. The thousands of metabolites measured in an untargeted metabolomics experiment will only be quantified in relative abundance across samples. However, this too is a concept microbiome scientists are quite familiar with, creating a common place for dialogue and sharing of strategies for the analysis of multi-variate and relative data. Much of the data analyses approaches commonly used in omics studies were developed by ecologists and fine-tuned by microbial ecologists (12, 13). Thus, it is important for microbiome scientists to communicate to the more analytically inclined that many of these approaches can be applied to metabolomics datasets with the potential to enrich their interpretability (14). Another important concept in metabolomics is its untargeted or targeted nature. For example, a metabolite or a panel of metabolites identified as important in a discovery study can then be rigorously quantified in a confirmatory or replicating study, reducing the cost by narrowing down the number of targeted metabolite analyses. Parallels exist in microbiome science as well, such as qPCR-based targeting of specific genes for rigorous quantification compared to more exploratory metagenomics methods for characterizing a microbiome's genetic complement.

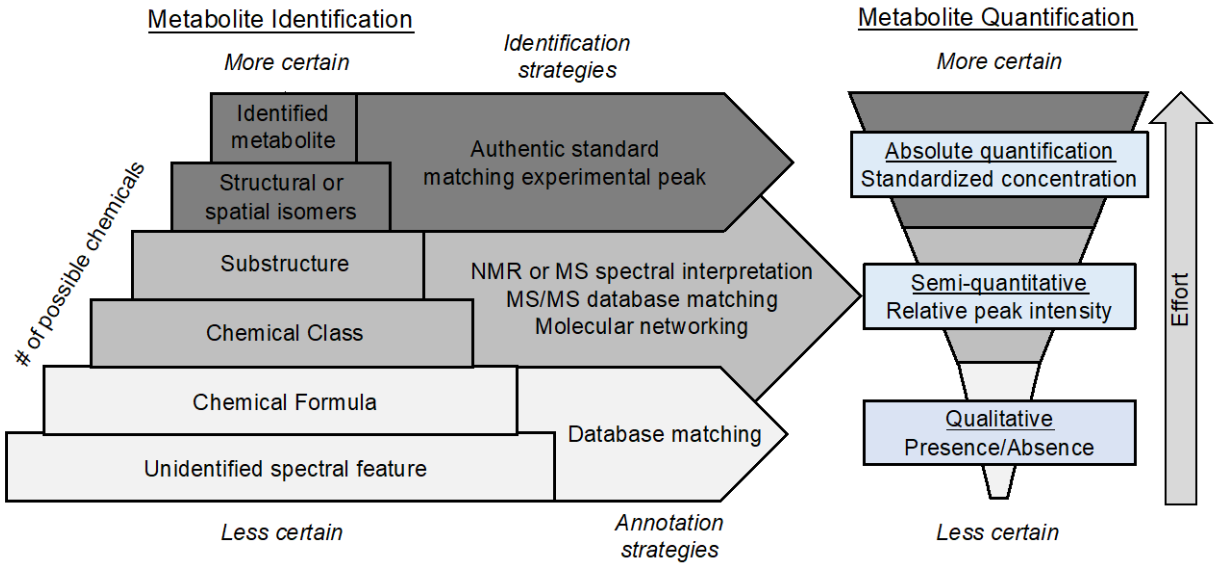

Fig. 1. The spectrum of (un)certainty in identifying and quantifying microbial metabolites.

Bridging the gap

Analytical scientists, many of which work in, or are in charge of, institutional core centers, are highly encouraged to transparently discuss the challenges and intricacies of metabolomics data for microbiome samples at the initial stages of a project's discussion and

maintain a continuous dialogue through its completion. This includes explaining the spectrum of certainty described above and the per sample cost of targeted and untargeted analysis. Analytical scientists are also encouraged to learn the details of the microbiome to be studied, such as field conditions for naturally acquired samples, or growth media composition for laboratory-derived samples. Resources are needed at academic and federal institutions to provide core centers and analytical chemists time and resources to work with microbiome scientists to avoid the inevitable 'data dump' that can lead to either incorrect biological interpretation or wasted effort on data that is never fully analyzed. In turn, many microbiome scientists are well versed in analytical methods, but most do not have the academic or technical training to interpret raw metabolomics data. Thus, educating oneself in the types of metabolomics platforms and instruments used is one step towards improving the dialogue. Perhaps more importantly, there is also a need to explain to the analytical scientists a project's specific goals. Most analytical scientists are unaware of the statistical approaches that can pull out biological information from multi-omics datasets even without labor intensive accurate quantification and annotation of compounds. Diversity indices, machine learning approaches and multi-omics integration can provide biological insights at the dataset scale (15–17). If accurate quantification is required, and it often is, this must be explained to the analytical scientist so they can appropriately design an assay and provide a fair cost estimate.

The need for cheminformaticians in microbiome science.

We propose that research institutions, core centers and academic labs, support and train 'cheminformaticians' to bridge the gap between the highly technical science of metabolomics and the urgent need to understand the role of microbiomes in human and environmental health, The desired dialogue described above is intensive and time consuming, prohibitively so for many core centers and academic labs. Thus, including and training individuals with experience in interpreting the technical language of metabolomics data but with a microbiome/biological background is highly beneficial. This is akin to the early years of genomics in microbiology, when labs with bench microbiology experience began to need bioinformaticians to help interpret the genomic data that they were generating. We advocate for funding agencies, academic institutions and PIs to advertise the need for and training of 'cheminformaticians' in microbiome science to develop a workforce of analytical language translators who can bridge the gap between microbiome and analytical science. Together this will lead to new and exciting findings about our microbiome's metabolic narratives.

References

1. Haiser HJ, Gootenberg DB, Chatman K, Sirasani G, Balskus EP, Turnbaugh PJ. 2013. Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta. Science 341:295–8.
2. Hang S, Paik D, Yao L, Kim E, Trinath J, Lu J, Ha S, Nelson BN, Kelly SP, Wu L, Zheng Y, Longman RS, Rastinejad F, Devlin AS, Krout MR, Fischbach MA, Littman DR, Huh JR. 2019. Bile acid metabolites control TH17 and Treg cell differentiation. Nature 576:143–148.
3. Byndloss MX, Olsan EE, Rivera-Chávez F, Tiffany CR, Cevallos SA, Lokken KL, Torres TP, Byndloss AJ, Faber F, Gao Y, Litvak Y, Lopez CA, Xu G, Napoli E, Giulivi C, Tsolis RM, Revzin A, Lebrilla CB, Bäumler AJ. 2017. Microbiota-activated PPAR-γ signaling inhibits dysbiotic Enterobacteriaceae expansion. Science (80-) 357:570–575.
4. Korenblum E, Dong Y, Szymanski J, Panda S, Jozwiak A, Massalha H, Meir S,

Rogachev I, Aharoni A. 2020. Rhizosphere microbiome mediates systemic root metabolite exudation by root-to-root signaling. Proc Natl Acad Sci 117:3874 LP – 3883.
5. Vizcaino MI, Crawford JM. 2015. The colibactin warhead crosslinks DNA. Nat Chem 7:411–7.
6. Da Silva RR, Dorrestein PC, Quinn RA. 2015. Illuminating the dark matter in metabolomics. Proc Natl Acad Sci U S A 112.
7. Peisl BYL, Schymanski EL, Wilmes P. 2018. Dark matter in host-microbiome metabolomics: Tackling the unknowns–A review. Anal Chim Acta 1037:13–27.
8. Quinn RA, Melnik A V., Vrbanac A, Fu T, Patras KA, Christy MP, Bodai Z, Belda-Ferre P, Tripathi A, Chung LK, Downes M, Welch RD, Quinn M, Humphrey G, Panitchpakdi M, Weldon KC, Aksenov A, Silva R da, Avila-Pacheco J, Clish C, Bae S, Mallick H, Franzosa EA, Lloyd-Price J, Bussell R, Thron T, Nelson AT, Wang M, Leszczynski E, Vargas F, Gauglitz JM, Meehan MJ, Gentry E, Arthur TD, Komor AC, Poulsen O, Boland BS, Chang JT, Sandborn WJ, Lim M, Garg N, Lumeng JC, Xavier RJ, Kazmierczak BI, Jain R, Egan M, Rhee KE, Ferguson D, Raffatellu M, Vlamakis H, Haddad GG, Siegel D, Huttenhower C, Mazmanian SK, Evans RM, Nizet V, Knight R, Dorrestein PC. 2020. Global chemical effects of the microbiome include new bile-acid conjugations. Nat 2020 579:123–129.
9. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW-M, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR. 2007. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 3:211–221.
10. Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, van der Voort M, Pogliano K, Gross H, Raaijmakers JM, Moore BS, Laskin J, Bandeira N, Dorrestein PC. 2012. Mass spectral molecular networking of living microbial colonies. Proc Natl Acad Sci U S A 109:1743–52.
11. Hartmann AC, Petras D, Quinn RA, Protsyuk I, Archer FI, Ransome E, Williams GJ, Bailey BA, Vermeij MJA, Alexandrov T, Dorrestein PC, Rohwer FL. 2017. Meta-mass shift chemical profiling of metabolomes from coral reefs. Proc Natl Acad Sci U S A 114.
12. Bray J, Curtis J. 1957. An Ordination of Upland Forest Communities of Southern Wisconsin. Ecol Monogr 27:325–349.
13. Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71:8228–35.
14. Morton JT, Aksenov AA, Nothias LF, Foulds JR, Quinn RA, Badri MH, Swenson TL, Van Goethem MW, Northen TR, Vazquez-Baeza Y, Wang M, Bokulich NA, Watters A, Song SJ, Bonneau R, Dorrestein PC, Knight R. 2019. Learning representations of microbe–metabolite interactions. Nat Methods 1–9.
15. Wilmanski T, Rappaport N, Earls JC, Magis AT, Manor O, Lovejoy J, Omenn GS, Hood L, Gibbons SM, Price ND. 2019. Blood metabolome predicts gut microbiome α-diversity in humans. Nat Biotechnol 37:1217–1228.
16. Raghuvanshi R, Vasco K, Vázquez-Baeza Y, Jiang L, Morton JT, Li D, Gonzalez A, DeRight Goldasich L, Humphrey G, Ackermann G, Swafford AD, Conrad D,

Knight R, Dorrestein PC, Quinn RA. 2020. High-Resolution Longitudinal Dynamics of the Cystic Fibrosis Sputum Microbiome and Metabolome through Antibiotic Therapy. mSystems 5:e00292-20.
17. Zierer J, Jackson MA, Kastenmüller G, Mangino M, Long T, Telenti A, Mohney RP, Small KS, Bell JT, Steves CJ, Valdes AM, Spector TD, Menni C. 2018. The fecal metabolome as a functional readout of the gut microbiome. Nat Genet 50:790–795.