nature portfolio

https://doi.org/10.1038/s41579-021-00621-9

Supplementary information

Mass spectrometry-based metabolomics in microbiome investigations

In the format provided by the authors and unedited

Supplementary Information

NRMICRO-16-134S1

Title: Mass spectrometry-based metabolomics in microbiome investigations

Authors: Anelize Bauermeister (1,3), Helena Mannochio-Russo (2,3), Letícia V. Costa-Lotufo (1), Alan K. Jarmusch (3, *), Pieter C. Dorrestein (3, 4, 5, *).

Address:

1) Institute of Biomedical Science, Universidade de São Paulo, São Paulo, SP, Brazil;
2) Department of Biochemistry and Organic Chemistry, Institute of Chemistry, São Paulo State University, Araraquara, SP, Brazil;
3) Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, CA, USA.
4) Department of Pediatrics, University of California, San Diego, CA, USA.
5) Center for Microbiome Innovation, University of California, San Diego, CA, USA.

***** To whom correspondence should be addressed pdorrestein@health.ucsd.edu and ajarmusch@health.ucsd.edu

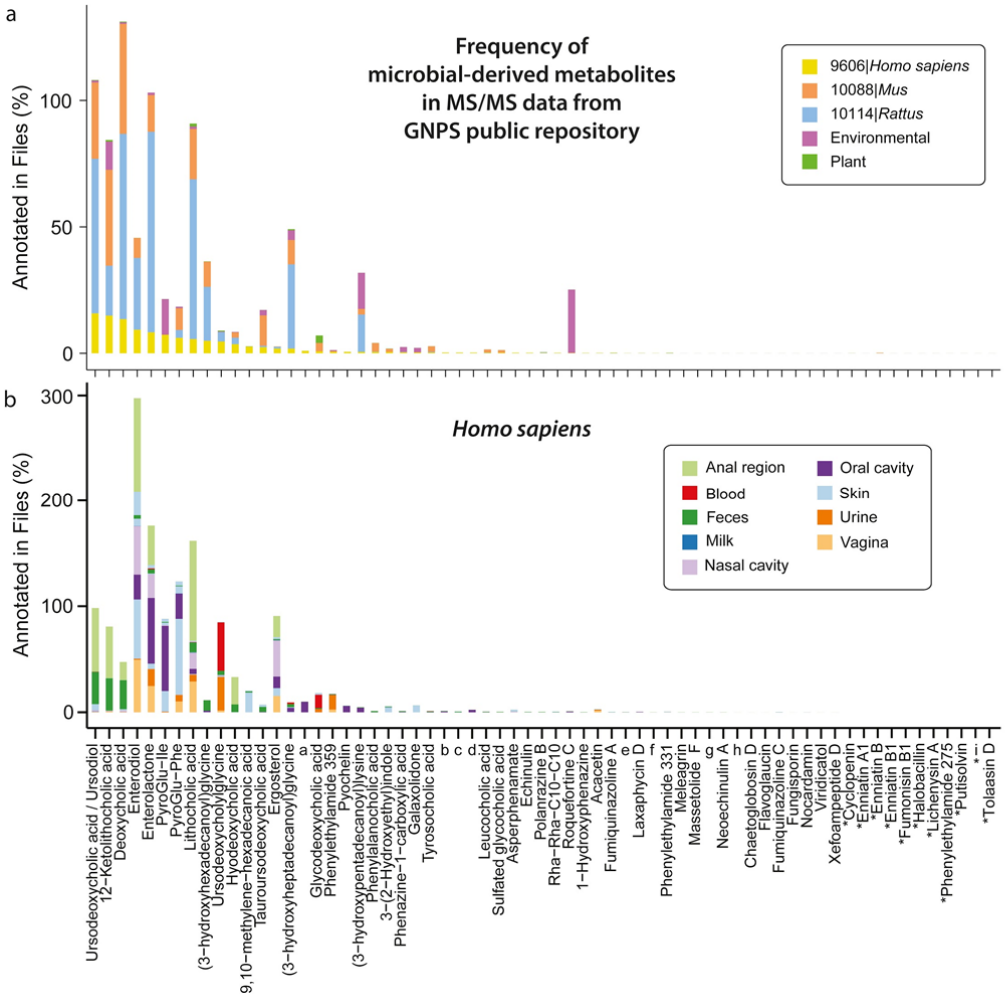
Supplementary Box 1. The emerging utility of public mass spectrometry data repositories

What to do when annotations are proposed with the strategies, discussed in the main text, the context, and meaning derived from context, is not readily available. For instance, “Is this mass spectrum in my human data set potentially microbial or host-derived?” This is, despite all the scientific advances in the field of MS, still very difficult to address at this time. This question may be, at least in part, answered with proper reference databases that are connected with search engines as it did for the sequencing community. BLAST is one such sequence search tool. BLAST provided context to sequencing data by searching a query sequence against a sequence repository[1], and reporting which sequences are most similar as well as the connection to the relevant data. Inspired by BLAST, MASST[2] is an open, web-enabled tool to query MS/MS spectra against public datasets available in GNPS/MassIVE, including imported data that were originally deposited in other data repositories, such as MetaboLights[3] and Metabolomics Workbench[4]. The reported results display the files and data sets in which the query MS/MS spectrum was observed, providing context. The utility of MASST and similar tools will increase as the data in the public domain grows. One limitation is the lack of agreement to make metabolomics data public, unlike in the genomics and proteomics fields, but the trend in increasing public data submissions is encouraging and suggests that the metabolomics field is going to transform into one where community knowledge is readily accessible. A strategic countermeasure to the apprehensiveness of public data sharing is via the development of useful tools to analyze public data and thereby the depositors gain, scientifically, from making data public[5–7].

One method of re-using public data is to perform a meta-analysis, *i.e.* analyze the results of multiple studies, for which there are numerous options. More recently, the

Reanalysis of Data User interface (ReDU) provides the ability to re-analyze raw public data, specifically MS/MS data, rather than reported results[8]. Striving to meet the FAIR (findable, accessible, interoperable, and reusable) principles,[9] ReDU includes a systematic way of entering metadata based on existing ontologies supplemented by a controlled vocabulary. Files in public data repositories can be discovered using ReDU-compliant metadata and re-analysis can be performed. For example, the microbe-associated annotations observed in humans, rats, mice, plants, and environmental files were tallied, Fig. Box 1a. ReDU allows one to gain insight into the frequency known microbial molecules are detected via spectral matching. Fig. Box 1a displays the chemical annotation versus the percentage of files in which the chemical was observed as a stacked bar plot. The chemical annotations displayed are not comprehensive, but illustrate microbe-modified dietary chemicals (*e.g.* enterolactone) as well as microbe-modified human metabolites (*e.g.* deoxycholic acid). Fig Box 1b displays the distribution of the same molecules in human samples based on the biofluid or tissue in which certain chemicals are observed predominantly in one sample type (*e.g.* 12-Ketodeoxycholic acid), whereas others are common in many sample types (*e.g.* Enterodiol).

The Figure below shows the results of querying public MS/MS data for microbe-associated chemicals via ReDU[8], in which it is presented in a) Stacked bar plots displaying the percentage of files (normalized by sample type) in which microbe-associated chemicals were observed for humans (yellow), mouse (orange), rat (light blue), plant (light green) and environmental (purple) samples. b) Stacked bar plots displaying the percentage of files (normalized by tissue or biofluid) in human samples. [a] 2−(undec−1−en−1−yl)quinolin−4−ol (Series 2 HAQ C11:1), [b] 4−hydroxy−2−octylquinoline 1−oxide (Series 4 HAQ C8), [c] mixed-MS/MS: 2−(non−1−en−1−yl)quinolin−4−ol (Series 2 HAQ C9:1) and 2−nonylquinolin−4−ol (Series 1 HAQ C9 aka HNQ), [d] 2−(non−1−en−1−yl)quinolin−4−ol (Series 2 HAQ C9:1), [e] mixed-MS/MS: 2−(hept−1−en−1−yl)quinolin−4−ol (Series 2 HAQ C7:1) and 2−heptylquinolin−4−ol (Series 1 HAQ C7), [f] Methyl (3beta, 5beta, 8alpha, 9beta, 10alpha, 13alpha)−3−acetoxy−4,4,8,12,16−pentamethyl−15,17,19−trioxoandrost−11−ene−14−carboxylate, [g] mixed-MS/MS: 2−(undec−1−en−1−yl)quinolin−4−ol (Series 2 HAQ C11:1) and 2−undecylquinolin−4−ol (Series 1 HAQ C11), [h] 2−nonylquinoline−3,4−diol (Series 3 HAQ C9), [i] Putisolvin Derivative.(*) Metabolites not detected in *Homo sapiens*. All annotations are Level 2 or 3[10].

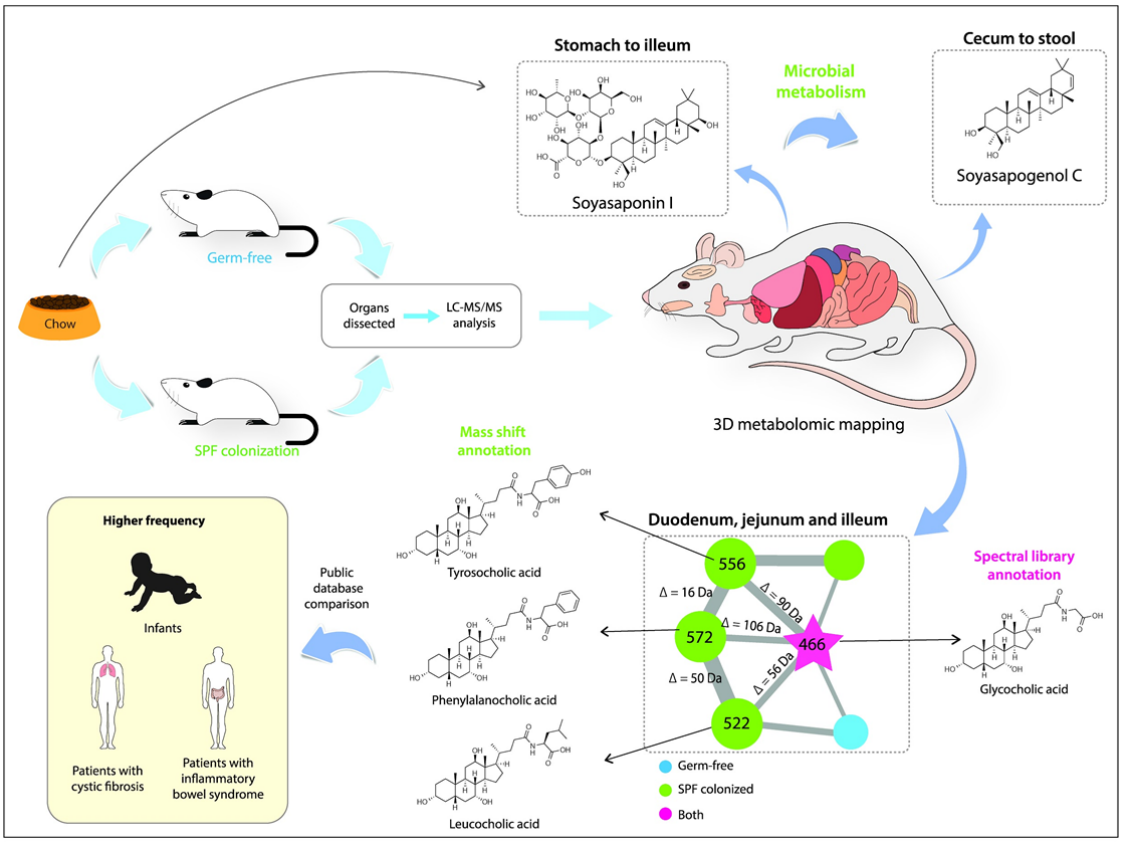


Supplementary Box 2. Animal models and their contribution to host-microbiota studies

Animal models allow investigations of microbiome interactions at high levels of experimental control, which is hard to do in human studies[11]. Many animal models have been used in microbiome investigations, such as the bobtail squid (*Euprymna scolopes*), the fruit fly (*Drosophila melanogaster*), the zebrafish (*Danio rerio*), among others[12–14]. Particularly, the mice (*Mus musculus*) is one of the most widely used animal models to understand and establish the roles of microbiota in mammals[11]. The possibility of using gnotobiotic (germ-free) animals allows the evaluation of the effects by colonization with specific microbes, or microbial communities in which it is possible to examine the effect of the human microbiota, generating a human-like condition[15]. The transplant of human microbiota into germ-free mice can provide new information about how the microbiome impacts depression[16], schizophrenia[17], autism[18], among others.

Recently, we performed a metabolomic approach in combination with 3D molecular cartography to reveal the effect of microbiota on the host in a regiospecific manner. All 29 organs investigated, even ones distal from the gut such as the brain, were significantly altered upon colonization of germ-free mice with a specific pathogen-free microbial

community[19]. It revealed that the microbiome is largely catabolic and is involved in food processing but also makes specific molecules that regulate host physiology. Soyasaponins, dietary molecules from soy, were detected throughout the entire gastrointestinal tract of the germ-free mice, while in the colonized mice (specific pathogen-free, SPF), these compounds were detected only from stomach to ileum, and its microbial-derived soyasapogenol was detected from cecum to stool. This shows the direct role of microbiota in the removal of the sugars from the soyasaponins as part of the digestion process (see figure below). In addition, molecular networking revealed the production of three new conjugated bile acids (microbial-derived), only detected in mice that were colonized in the upper intestine. It is therefore important to understand microbiome location with respect to the molecular environment and host physiology that they influence. These new bile acids are the most potent natural FXR antagonists found to date. This means that microbes have a sting effect on host physiology. A MASST search revealed that these findings were translational to humans and these new conjugated bile acids were also detected in public data from humans. A higher frequency of these compounds was observed in infants, and pancreatic insufficient cystic fibrosis patients and in the dysbiotic state of Crohn's patients. The above observations are a clear illustration of how model animals can be used to understand the roles of microbiota on the entire host.



References

1. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
2. Wang, M. *et al.* MASST: A Web-based Basic Mass Spectrometry Search Tool for Molecules to Search Public Data. doi:10.1101/591016.
3. Kale, N. S. *et al.* MetaboLights: An Open-Access Database Repository for Metabolomics Data. *Curr. Protoc. Bioinformatics* **53**, 14.13.1–14.13.18 (2016).
4. Sud, M. *et al.* Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* **44**, D463–70 (2016).
5. Misra, B. B. & Mohapatra, S. Tools and resources for metabolomics research community: A 2017-2018 update. *Electrophoresis* **40**, 227–246 (2019).
6. Shaffer, M., Armstrong, A. J. S., Phelan, V. V., Reisdorph, N. & Lozupone, C. A. Microbiome and metabolome data integration provides insight into health and disease. *Transl. Res.* **189**, 51–64 (2017).
7. Chong, J. & Xia, J. Computational Approaches for Integrative Analysis of the Metabolome and Microbiome. *Metabolites* **7**, (2017).
8. Jarmusch, A. K. *et al.* ReDU: a framework to find and reanalyze public mass spectrometry data. *Nat. Methods* **17**, 901–904 (2020).
9. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).
10. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211–221 (2007).
11. Douglas, A. E. Simple animal models for microbiome research. *Nat. Rev. Microbiol.* (2019) doi:10.1038/s41579-019-0242-1.

12. Bauermeister, A. *et al.* Tunicates: A model organism to investigate the effects of associated-microbiota on the production of pharmaceuticals. *Drug Discovery Today: Disease Models* (2019) doi:10.1016/j.ddmod.2019.08.008.
13. Wang, X., Zhang, X., Zhang, Z., Lang, H. & Zheng, H. Honey bee as a model organism to study gut microbiota and diseases. *Drug Discovery Today: Disease Models* (2019) doi:10.1016/j.ddmod.2019.08.010.
14. Vlasova, A. N., Rajashekara, G. & Saif, L. J. Interactions between human microbiome, diet, enteric viruses and immune system: Novel insights from gnotobiotic pig research. *Drug Discovery Today: Disease Models* (2019) doi:10.1016/j.ddmod.2019.08.006.
15. Martín, R., Bermúdez-Humarán, L. G. & Langella, P. Gnotobiotic Rodents: An In Vivo Model for the Study of Microbe–Microbe Interactions. *Frontiers in Microbiology* vol. 7 (2016).
16. Zheng, P. *et al.* Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. *Molecular Psychiatry* vol. 21 786–796 (2016).
17. Zheng, P. *et al.* The gut microbiome from patients with schizophrenia modulates the glutamate-glutamine-GABA cycle and schizophrenia-relevant behaviors in mice. *Science Advances* vol. 5 eaau8317 (2019).
18. Sharon, G. *et al.* Human Gut Microbiota from Autism Spectrum Disorder Promote Behavioral Symptoms in Mice. *Cell* **177**, 1600–1618.e17 (2019).
19. Quinn, R. A. *et al.* Global chemical effects of the microbiome include new bile-acid conjugations. *Nature* **579**, 123–129 (2020).