The National Microbiome Data Collaborative community survey on microbiome research data

Pajau Vangay* (pvangay@lbl.gov), Faiza Ahmed, Anubhav, Jeffrey Baumés, Jonathan Beezley, Mark Borkum, Lisa Bramer, Shane Canon, Patrick Chain, Danielle Christianson, Yuri Corilo, Karen Davenport, Brandon Davis, Meghan Drake, William Duncan, Kjersten Fagnani, Mark Flynn, David Hays, Bin Hu, Marcel Huntemann, Julia Kelliher, Sonya Lebedeva, Po-E Li, Mary Lipton, Chien-Chi Lo, Douglas Mans, Stanton Martin, Lee Ann McCue, David Millard, Kayd Miller, Nigel Mouncey, Chris Mungall, Elais Player Jackson, Anastasiya Prymolenna, Samuel Purvine, TBK Reddy, Rachel Richardson, Migun Shakya, Montana Smith, Jagadish Chandrashe Hariharamorthy, Deepak Unni, Bruce Wilson, Donny Winsten, Elisha Wood-Charlson, Yan Xu, Emiley Elo-Fadrosh

* Lawrence Berkeley National Laboratory, Berkeley, CA; ² Kitware, Clifton Park, NY; ³ Pacific Northwest National Laboratory, Richland, WA; ⁴ Los Alamos National Laboratory, Los Alamos, NM; ⁵ Oak Ridge National Laboratory, Oak Ridge, TN; ⁶ Polyneme LLC, New York, NY

**Project:** The National Microbiome Data Collaborative is a pilot initiative launched to support microbiome data exploration and discovery through a collaborative, integrative science gateway. With a community-centered design approach, the NMDC team is building an integrated data science ecosystem that leverages existing data standards, data resources, and infrastructure within the DOE complex.

**Objectives**

- Understand how the research community works with microbiome data
- Capture diversity of the community
- Identify people who want to get more involved with the NMDC

**Methods**

- Survey was hosted by Qualtrics and made available via an anonymous, shareable, web-based URL from Dec 2020 - Jan 2021
- Survey was shared with NMDC collaborators and partner stakeholders across all research fields
- We encouraged redistribution of the survey via email and Twitter
- The survey was approved by the LBNL Human Subjects Committee (#394NR001-5OC21)

**Results**

Survey respondents (n=773) are primarily principal investigators in the environmental sciences and health sciences, who are based in academia.

The majority of respondents work with metagenomics and amplicon data to study bacteria.

And have a range of experience in generating, processing, and analyzing microbiome data.

**Metadata standards across research fields**

- 65% of respondents want to get more involved with the NMDC

**Anonymous general feedback**

"I’m really excited about the work the NMDC is doing. It is really inspiring."

"I have done several microbiome meta-analysis and less than 50% of the datasets are usable with good metadata."

"Provide most of the microbiome bioinformatics workflows freely available and if possible allocate enough free space for in cloud bioinformatic computation."

"30% of respondents had heard of the NMDC before this survey"

"65% of respondents want to get more involved with the NMDC"

**Next Steps**

- Continue analysis of the remaining survey data
- Together with the research community, develop training on metadata standards
- Use the feedback on search, access, and computational environments to prioritize directions
- Reach out to the 500+ individuals who want to get more involved

**Funding:** This work is supported by the Genomic Science Program in the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER) under contract numbers DE-AC02-05CH11231 (LBNL), DE-AC03-76SF00098 (LANL), DE-AC05-00OR22725 (ORNL), and DE-AC02-76SF00515 (PNNL).