



nmdc
National Microbiome
Data Collaborative

The National Microbiome Data Collaborative community survey on microbiome research data

Pajau Vangay^{1*} (pvangay@lbl.gov), Faiza Ahmed², Anubhav³, Jeffrey Baumes², Jonathan Beezley², Mark Borkum³, Lisa Bramer³, Shane Canon¹, Patrick Chain⁴, Danielle Christianson¹, Yuri Corilo³, Karen Davenport⁴, Brandon Davis², Meghan Drake⁵, William Duncan¹, Kjersten Fagnan¹, Mark Flynn⁴, David Hays¹, Bin Hu⁴, Marcel Huntemann¹, Julia Kelliher⁴, Sonya Lebedeva¹, Po-E Li⁴, Mary Lipton³, Chien-Chi Lo⁴, Douglas Mans³, Stanton Martin⁵, Lee Ann McCue³, David Millard³, Kayd Miller¹, Nigel Mouncey¹, Chris Mungall¹, Paul Piehowski³, Elais Player Jackson⁴, Anastasiya Prymolenna³, Samuel Purvine³, TBK Reddy¹, Rachel Richardson³, Migun Shaky⁴, Montana Smith³, Jagadish Chandrabose Sundaramurthi¹, Deepak Unni¹, Bruce Wilson⁵, Donny Winston⁶, Elisha Wood-Charlson¹, Yan Xu⁴, Emiley Elloe-Fadros¹

¹ 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

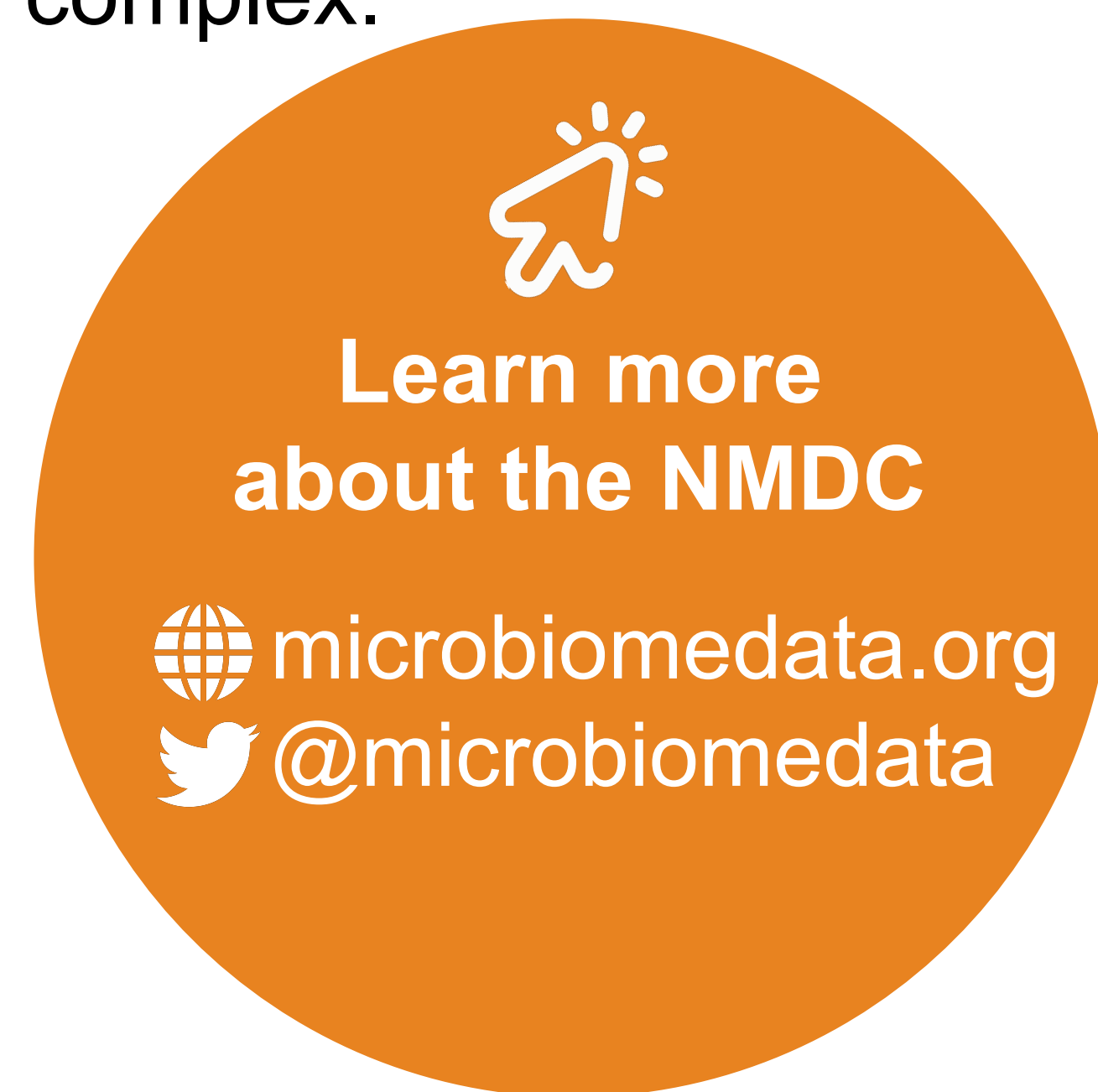
Project: The National Microbiome Data Collaborative is a pilot initiative launched to support microbiome data exploration and discovery through a collaborative, integrative science gateway. With a community-centered design approach, the NMDC team is building an integrated data science ecosystem that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

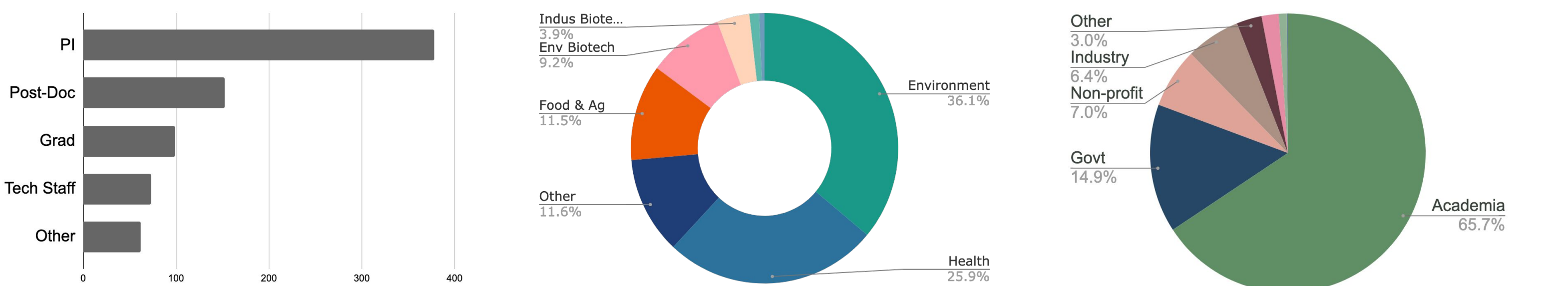
Methods

- 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)

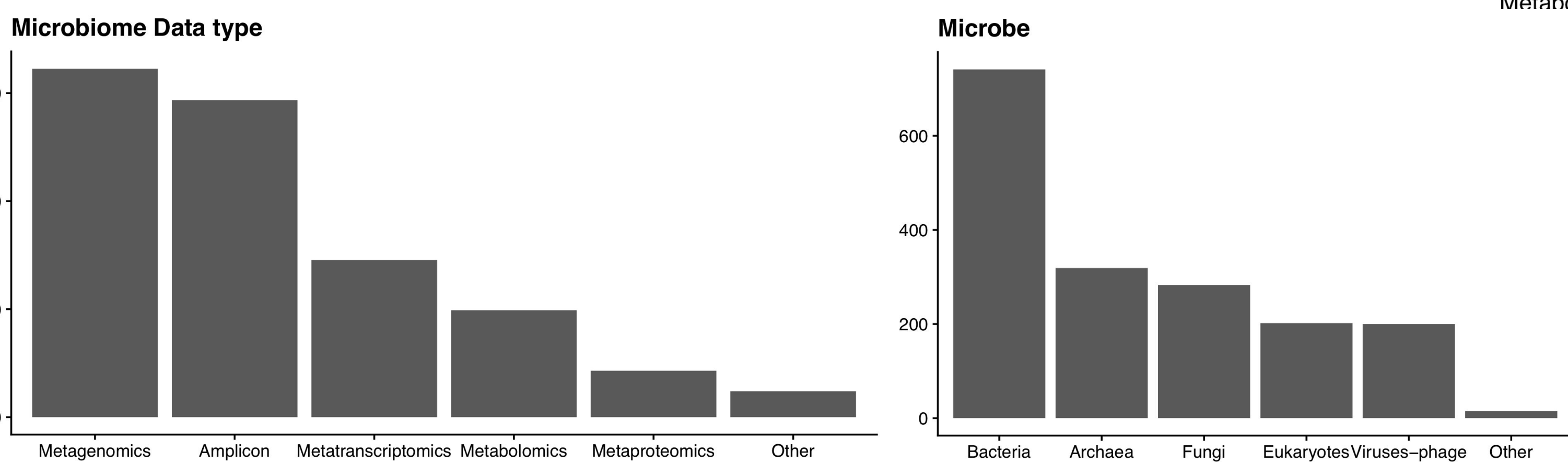


Results

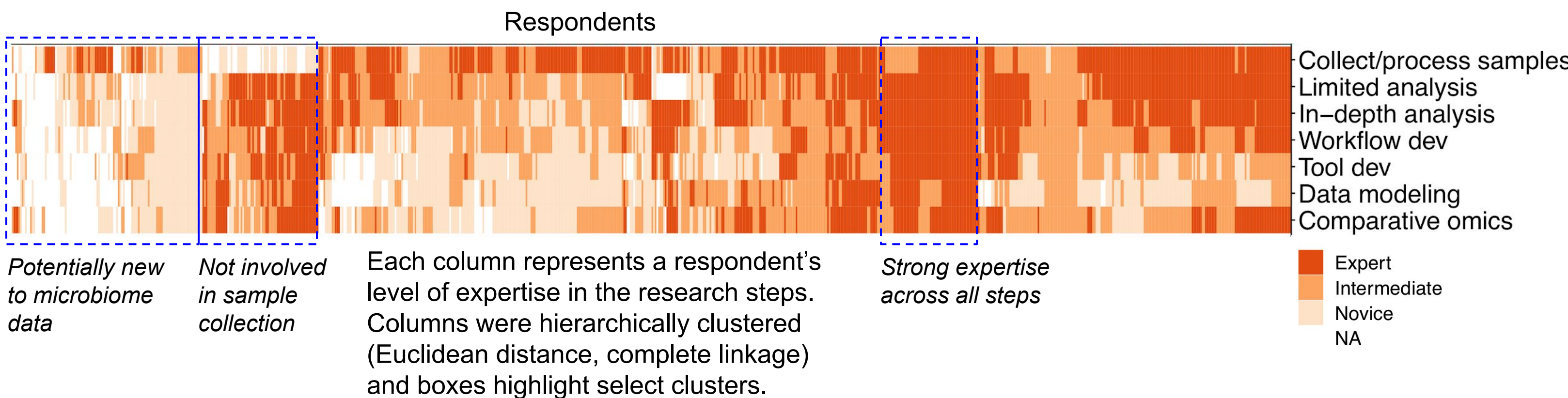
Survey respondents (n=773) are primarily principal investigators in the environmental sciences and health sciences, who are based in academia.



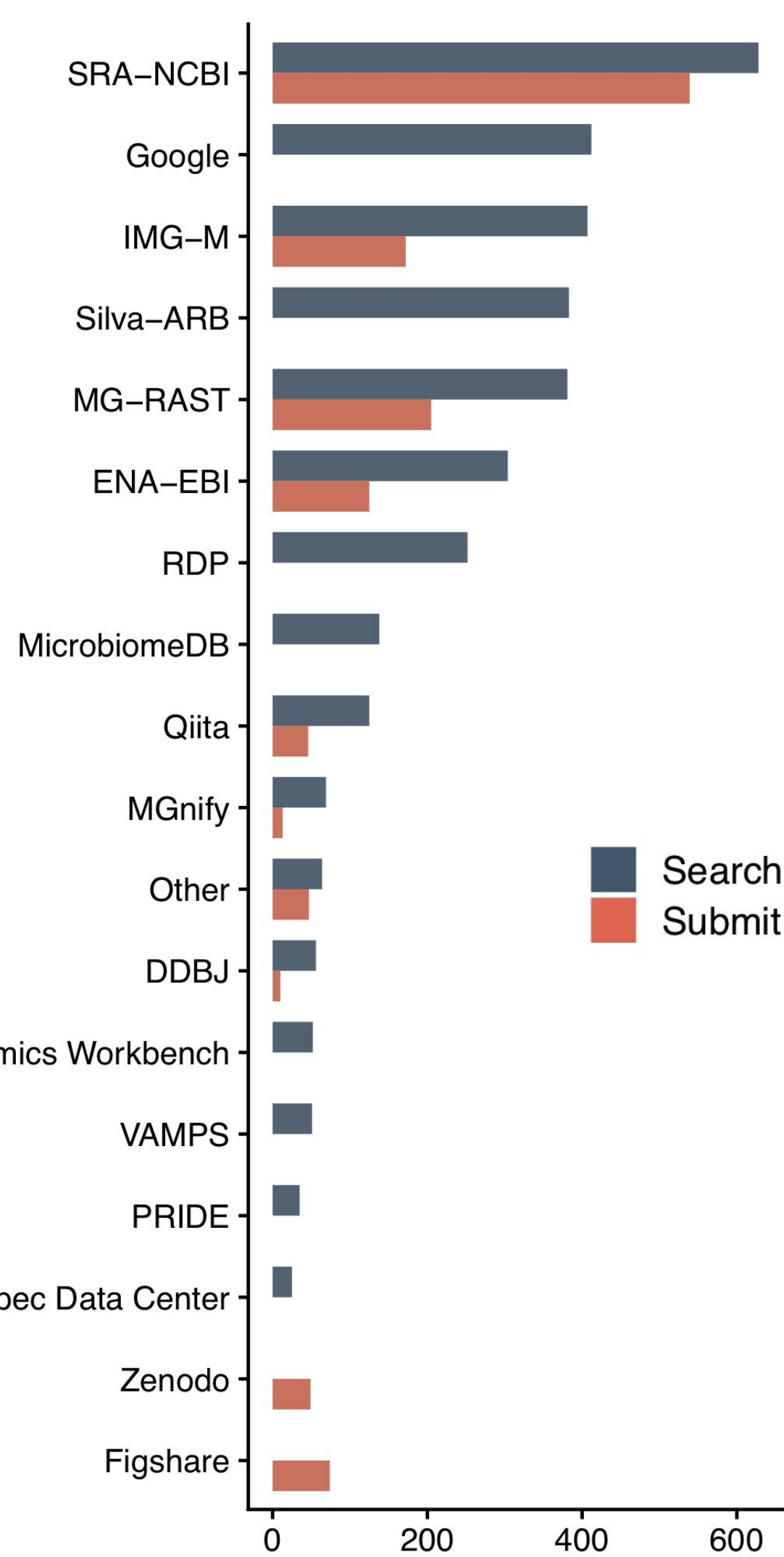
The majority of respondents work with metagenomics and amplicon data to study bacteria.



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

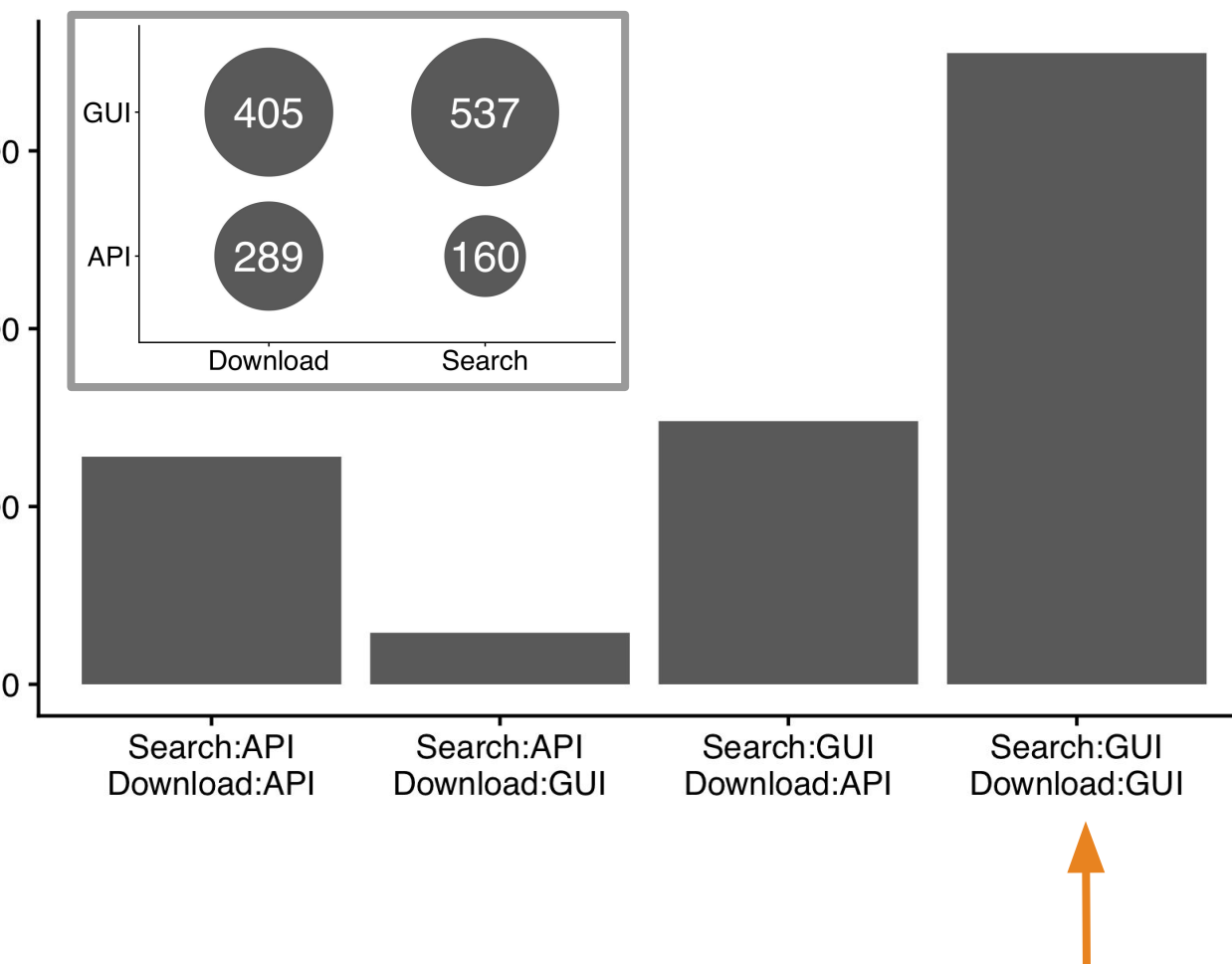


Search and submission resources



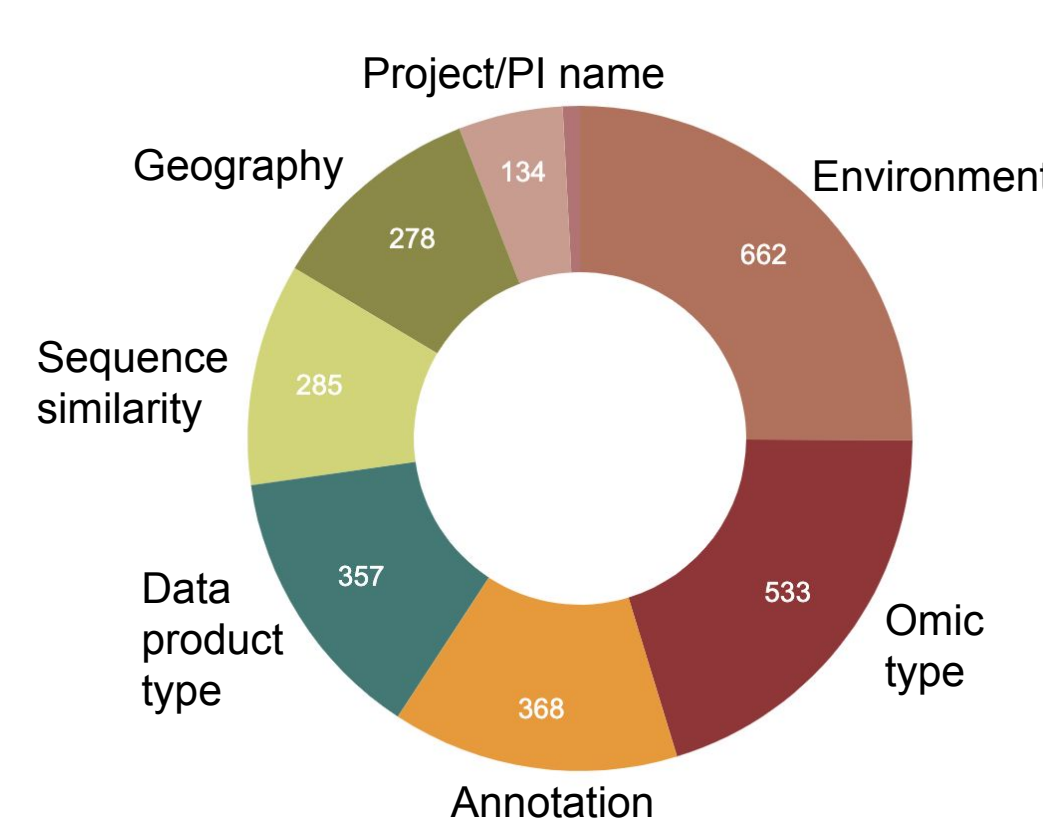
NCBI/SRA is the most commonly used resource for searching and sharing of microbiome data

Search and download interfaces

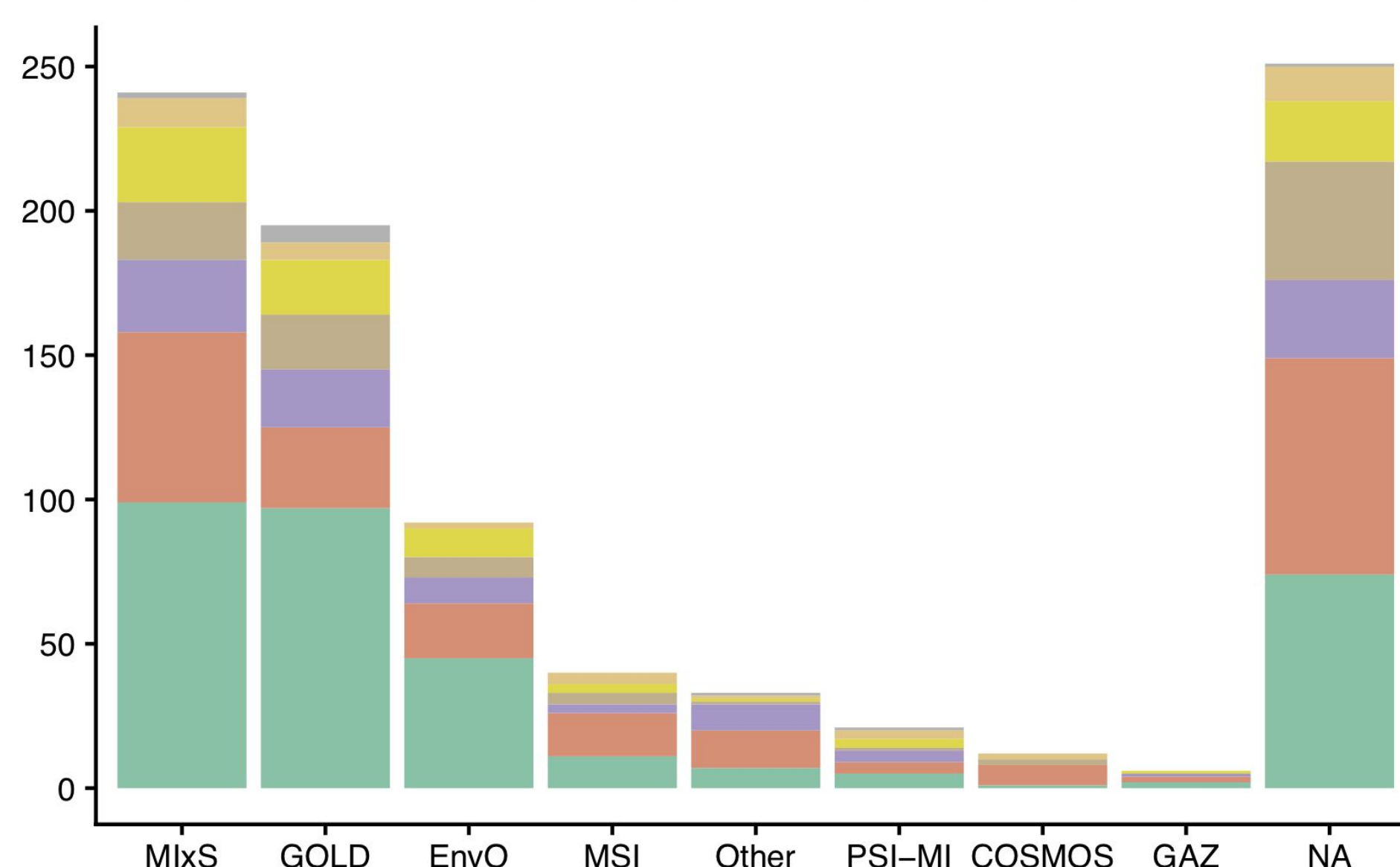


Most people prefer to use a graphical user interface for both searching and downloading of datasets
**many noted it depends on the size of the dataset*

Search criteria for microbiome datasets



Metadata standards across research fields



1/3 of respondents have used the MIxS standard

1/3 of respondents have not used any metadata standard

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), 89233218CNA000001 (LANL), DE-AC05-00OR22725 (ORNL), and DE-AC05-76RL01830 (PNNL).

