



OPEN

Quantifying the impact of workshops promoting microbiome data standards and data stewardship

Julia M. Kelliher^{1,14}✉, Francisca E. Rodriguez^{1,14}✉, Leah Y. D. Johnson¹, Simon Roux², Montana Smith³, Alicia Clum², Wendi Lynch², Candace Hope Bias^{4,5}, Sarai S. Finks⁶, Ishi Keenum⁷, E. Anders Kiledal⁸, Heng-An Lin⁹, Reid Longley¹, Ryan McDonald⁴, Thomas M. Pitot¹⁰, Josué Rodríguez-Ramos³, Jiaxian Shen¹¹, Daniel D. Sprockett¹², Joel Swift¹³, Archana Yadav² & Emiley A. Eloë-Fadrosch²✉

The field of microbiome research continues to grow at a rapid pace, with multi-omics approaches becoming widely used to interrogate diverse microbiome samples. However, due to lagging awareness and implementation of standards and data stewardship, many datasets are produced that are not comparable, reproducible, or reusable. In 2021, the National Microbiome Data Collaborative launched its Ambassador Program, which utilizes a community-learning model to annually train a cohort of early-career researchers in microbiome data stewardship best practices. These Ambassadors then host workshops and other events to communicate these themes to their respective microbiome research communities. To quantify the impact of this learning model for promoting awareness of and experience with microbiome data, we conducted a survey of workshop participants from events hosted by the 2023 Ambassador cohort. The 2023 cohort of 13 National Microbiome Data Collaborative Ambassadors collectively hosted 21 events, reaching over 550 researchers. The Ambassadors distributed an anonymous post-workshop survey to their event participants to quantify the effectiveness of the training materials, the workshop format, and the thematic content. From the 21 events, survey results were successfully collected for 15 of those events from a total of 122 researchers. Overall, 122 participants working with a range of microbiome types and from a variety of institutions responded to the survey and reported overwhelmingly positive experiences with the workshop content and materials, with 98% of respondents reporting that they gained knowledge from the event. Participants across the events also reported an increase in their post-workshop understanding of metadata standards, principles for microbiome data management and reporting, and the importance of standardization in microbiome data processing. Participants also expressed a willingness to apply what they learned about microbiome data stewardship to their own research. The results of this study demonstrate the effectiveness of hands-on workshops and community-learning for communicating data stewardship best practices to microbiome researchers. The lessons learned and details about the implementation of this cohort-based learning model contained herein are intended to assist other groups in their efforts to create or improve similar learning strategies.

Keywords Community learning, Microbiome, Standards, Open science, FAIR data, Retrospective methodology, Workshop, Cohort-based learning

Abbreviations

BIPOC	Black, indigenous, people of color
EDGE	Empowering the development of genomics expertise
FAIR	Findable, accessible, interoperable, and reusable
GSC	Genomic standards consortium
MiXS	Minimum information about any (x) sequence
NIST	National institute of standards and technology

NMDC	National microbiome data collaborative
NSURP	National summer undergraduate research project
STORMS	Strengthening the organization and reporting of microbiome studies

¹Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM, USA. ²Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ³Pacific Northwest National Laboratory, Richland, WA, USA. ⁴U.S. Food and Drug Administration, College Park, MD, USA. ⁵Oak Ridge Institute for Science and Education, Oak Ridge, TN, USA. ⁶Department of Biology, The Pennsylvania State University, University Park, PA, USA. ⁷Michigan Technological University, Houghton, MI, USA. ⁸Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI, USA. ⁹Texas A&M University, College Station, TX, USA. ¹⁰Department of Biochemistry, Microbiology and Bioinformatics, Université Laval, Québec, QC, Canada. ¹¹Northwestern University, Evanston, IL, USA. ¹²Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA. ¹³Kansas Biological Survey & Center for Ecological Research, University of Kansas, Lawrence, KS, USA. ¹⁴Julia M. Kelliher and Francisca E. Rodriguez contributed equally to this work. ✉email: jkelliher@lanl.gov; francisca@lanl.gov; eaeloefadros@lbl.gov

Microbiome research is an exponentially growing field that spans diverse domains ranging from human health to agriculture to aquatic system functioning^{1–5}. Increasingly, researchers utilize multi-omics approaches to generate large, complex datasets in an effort to understand the genomic composition and functional potential of microbial communities^{6–8}. Much of this data is currently generated and documented in non-standardized ways across researchers and organizations, creating challenges for data reuse and ultimately limiting the return on investment for microbiome studies^{9,10}. Efforts have been made by several groups to establish standards and guidelines for microbiome research best practices, such as the genomic standards consortium (GSC) reporting standards, the Strengthening The Organization and Reporting of Microbiome Studies (STORMS) guidelines for human microbiome research, and the controls developed by the national institute of standards and technology (NIST)^{10–14}. While these efforts have established a solid foundation for microbiome data stewardship best practices, the limited awareness and adoption of these standards across the microbiome research community remains a significant hindrance¹⁵.

To address this gap in awareness and training, we launched the National Microbiome Data Collaborative (NMDC) Ambassador Program in 2021 as an annual program preliminarily focused on hosting training workshops for metadata standards. Using a cohort-based learning approach, the pilot cohort of Ambassadors were trained on the benefits of findable, accessible, interoperable, and reusable (FAIR) data principles¹⁶, metadata standards, and how to use the GSC's Minimum Information about any (x) Sequence (MIxS) standard¹⁴. This pilot cohort of Ambassadors then hosted 23 events over 9 months, reaching over 800 researchers¹⁷. While the events ranged from panel discussions to town halls, most events were hands-on workshops focused on providing practical experiences with the MIxS templates to workshop participants. Similar to other successful workshops focused on bioinformatics tools and data science resources^{18–20}, we found this training approach to be highly effective¹⁷. We also observed the 'train-the-trainer' approach had additional value as a means for emerging researchers in the field to serve as leaders within their domain-specific networks and be part of a coordinated national program. This echoes the successes seen with other implementations of this community engagement format²¹.

Leveraging the lessons learned during the pilot phase, we expanded the NMDC Ambassador Program beyond metadata standards to other aspects of data stewardship within microbiome research and to include content about the three NMDC products: the NMDC Submission Portal, NMDC EDGE²², and the NMDC Data Portal²³. This expansion also included the option for Ambassadors to choose one or a combination of three outlined content paths to focus their events on: 1) Microbiome data stewardship, data management, and the NMDC Data Portal; 2) Microbiome metadata standards, metadata templates, and the NMDC Submission Portal; or 3) Multi-omics data processing, standardized bioinformatics workflows, and NMDC EDGE. To better quantify the training activities and impact, we designed a post-event survey to be completed by event participants to evaluate if the educational approach, training materials, and content were effective. The survey contained retrospective questions to capture participant assessments of their own knowledge about the topics covered prior to the event and after the event, while minimizing response shift biases and incomplete datasets that can occur when participants are given separate pre- and post-event surveys^{24–26}. Overall, the 2023 Ambassador cohort of 13 Ambassadors hosted 21 events in three countries (USA, Canada, and Japan) and in three languages (English, Spanish, and French), reaching over 550 participants over the course of the year^{27,28}. Herein, we report on the anonymized survey results from 122 participants (22% of the total reported event participants) who responded to the post-event survey. The survey results were analyzed to assess the quantitative impact of this learning model for microbiome researchers, and these insights and lessons learned can serve as a guide for implementing similar training programs in other fields of research.

Methods

Survey design

The post-event survey was designed by the NMDC team and reviewed and approved by the Human Subjects Committee at Lawrence Berkeley National Laboratory as an exempt IRB protocol under #394NR002. Informed consent was obtained from all participants. All methods were carried out in accordance with relevant guidelines and regulations for human subjects research. The post-event survey included the following questions regarding participant event experience (<https://doi.org/10.6084/m9.figshare.25045667.v1>). The survey questions were grouped into five sections: i) Background; ii) Event experience; iii) Standardization, FAIR data, and Data stewardship; iv) NMDC products; and v) Next steps.

For the first section, *Background*, we sought to capture information about participant career stage, institution type, and microbiome research domain to better understand who was attending these workshops and if the content was applicable across diverse research backgrounds and expertise levels. This section included three questions, two of which were checkboxes (institution type and microbiome research domain) and one of which was multiple choice (career stage). The second section, *Event experience*, included six questions focused on assessing participant experiences and overall opinions about the event. This section utilized Likert scales to balance positive and negative options and included one multiple choice question and one long answer question to capture any additional comments about participant event experiences²⁹. The third section, *Standardization, FAIR data, and Data Stewardship*, included four Likert scale questions based on a retrospective survey design structure. These questions were displayed as multiple choice matrices where participants were asked to rank their familiarity with a topic and their perceived importance of a topic before and after the event to determine gained knowledge throughout the course of the event^{24–26}. These questions provided background information on the participants' knowledge of the topics presented as well as the perceived knowledge gained from the event.

The *NMDC Products* section included nine subsections, each focused on one of the NMDC products and any hands-on activities included in the event about the products. Participants were only directed to these specific questions if their event included detailed information about any of these products. The last section, *Next steps*, included three questions to assess how the microbiome community intends to incorporate data stewardship principles into their own work, how they intend to stay connected with the NMDC, and to capture any feedback that may not have been otherwise addressed by the survey. This section was used to evaluate how actionable the event content was and determine potential methods for continued engagement with event participants.

Survey distribution

The approved event survey questions were added to Google Forms for distribution, and survey links for each Ambassador event were created. Ambassadors were provided with the links and QR codes for the survey to distribute to participants following their events. Ambassadors sent the survey link directly to any virtual participants or displayed the link or QR code for participants to access the Form. Many participants utilized the QR code option and answered the survey on their mobile devices. It was estimated that the survey would take around five minutes to complete based on preliminary testing by the survey designers. Participants were given ample time at the end of most of the events to complete the survey and were encouraged to complete it while still at the workshop, but participants with the survey link were able to provide their answers at any time following the event. Participants were provided with the IRB information at the start of the survey and were not required to complete the survey. No survey questions were required to be answered, and all responses were anonymous. Respondents were not compensated for completing the survey, but the benefits of the survey were explained in the context of how the results would help to improve the Ambassador Program and the NMDC products, and that the results would be summarized in a publication. For the six events without survey results, the Ambassadors either ran out of time, the event type was not conducive for this type of assessment (e.g., they participated in a town hall rather than hosted a workshop), or participants were unable to take the survey due to technical challenges (e.g., participants did not have a phone or computer, could not access Google Forms, internet problems). Therefore, the number of respondents does not match the total number of event attendees for this cohort.

Survey results & statistical analyses

Each Ambassador event had a unique survey link and event identifier to enable event-specific response analyses if desired. However, after the completion of the program, all survey data was collated to better assess the impact of the entire Ambassador cohort rather than individual Ambassadors or events. NMDC team members approved to work on this IRB protocol removed or anonymized any potentially identifiable information from the combined results, including the date, time, and other information included in the long answer responses (e.g., participants mentioning their Ambassador event host's name in a comment).

As not all questions were required to be answered, unanswered questions did not disqualify an entire participant's survey response from being included in the larger combined dataset, but any unanswered question responses were removed prior to statistical analyses. For the retrospective questions within the *Standardization, FAIR data, and Data Stewardship* section, participant responses were only included if they answered for both the pre- and post-event options.

Microsoft Excel was used for basic data organization and calculations. For the retrospective survey questions, the results were analyzed using two-tailed paired sample t-tests with a significance cut-off of $p < 0.05$. Results were transformed into figures using R (Version 4.3.1), RStudio (Version 2023.06.1 + 524)³⁰, ggplot2 (Version 3.4.3)³¹, and UpSetR (Version 1.4.0)³² R packages.

Results

The 2023 cohort of 13 NMDC Ambassadors hosted 21 events throughout their term (May 2023 to December 2023). Of the 13 Ambassadors, 3 were graduate students, 8 were postdoctoral fellows, and 2 were research scientists. To foster collaborative and networking opportunities, Ambassadors were encouraged to co-host events with other Ambassadors, and seven of the 21 events were co-hosted by at least two Ambassadors. One Ambassador extended the 'train-the-trainers' model into their own events, holding a smaller hands-on workshop with graduate students to give them the tools needed to then co-host an event for undergraduates. Two Ambassadors co-hosted a session within the national summer undergraduate research project (NSURP) aimed at providing rewarding microbiology research opportunities for black, indigenous, people of color (BIPOC) and Latinx students³³. Two Ambassadors modified their workshop content to accommodate the primary language of their event attendees, with one Ambassador translating their workshop slides into Spanish for a workshop with

undergraduates at the Universidad de Puerto Rico²⁸ and another Ambassador presenting in French for an event at the Université du Québec à Chicoutimi.

Of the 21 total Ambassador events, 15 included the distribution of a post-event survey to capture participant responses about their event experiences, their assessments of the presented materials, and what they gained from the event. Collectively, these 15 events captured responses from 122 participants. None of the questions were required to be answered, therefore, some questions had more participant responses than others.

The Ambassador events reached researchers primarily from academic institutions (76.2%, $n=93$), although several events also included participants from government (15.6%, $n=19$), industry (2.5%, $n=3$), and a combination of the academia and government sectors (5.7%, $n=7$) (Fig. 1A). Within these sectors, the majority of event participants were graduate students (39.3%, $n=48$), with the remainder spanning all career

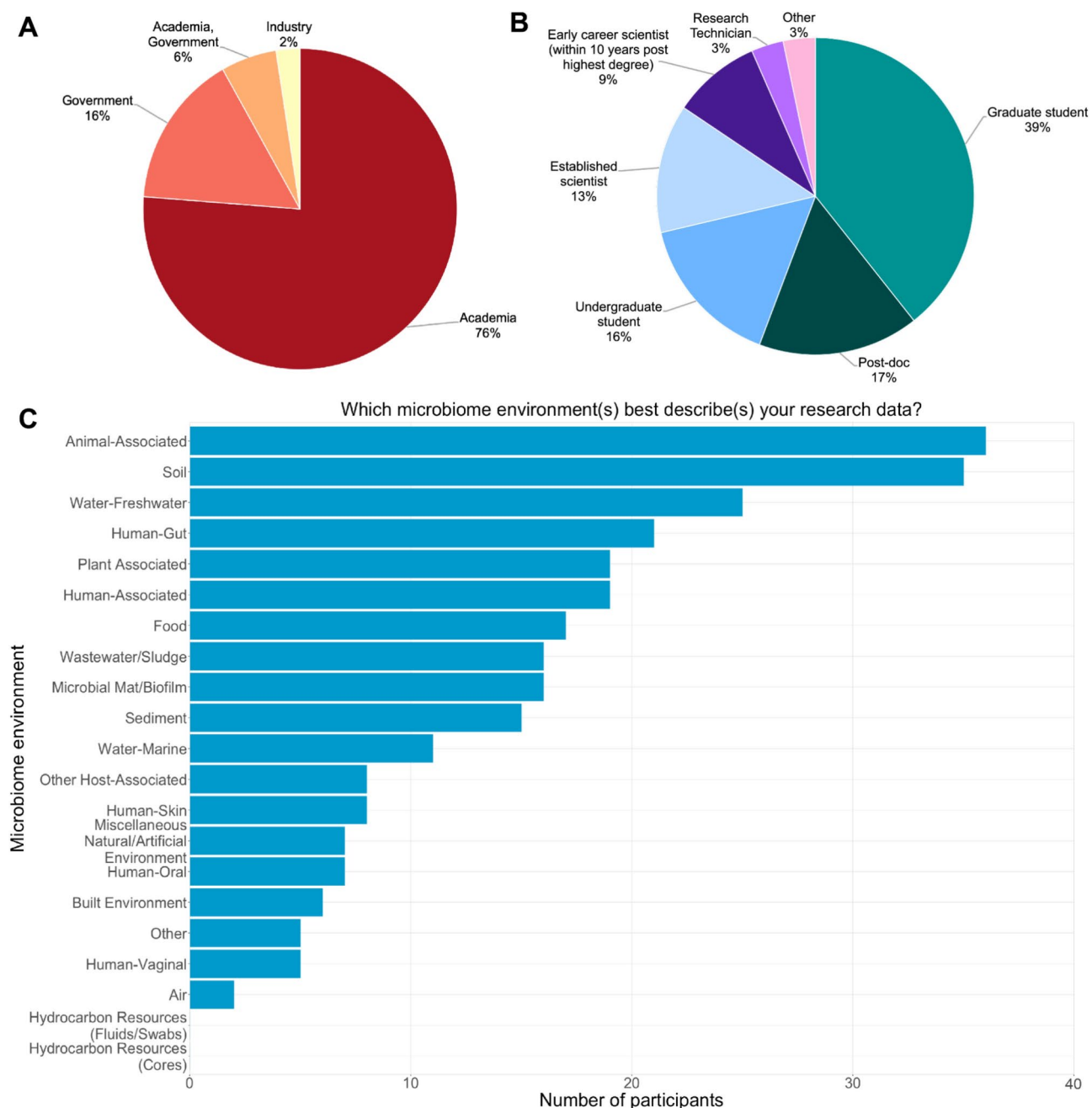


Fig. 1. Self-reported responses from event participants regarding their demographic information and research background. **(A)** Sector of participant primary institution, as reported from a checkbox question with the prompt “Which sector best describes your primary institution? Check all that apply” (total responses $n=122$). **(B)** Participant career stage, as reported from a multiple choice question with the prompt “What career stage do you identify with?” (total responses $n=122$). **(C)** Participant microbiome environment research domain as defined using the GSC’s MxS environment extensions (total responses $n=120$).

stage options in the survey, from Undergraduate student (15.6%, $n = 19$) to Established scientist (13.1%, $n = 16$) (Fig. 1B). Respondents also reported working with diverse microbiome environments, ranging from human skin to freshwater, with the majority focused on animal-associated and soil microbiomes, and more than half (51.7%, $n = 62$) reporting that they work with at least two of the listed microbiome environment types (Fig. 1C).

Four questions about overall event experience were asked using a Likert scale format from 1 (Strongly disagree) to 5 (Strongly agree) to assess participants' overall thoughts, feelings, and takeaways from the events. For the question, "The content of the event was useful and appropriate to my existing level of knowledge about the subject", 95% ($n = 116$) of respondents reported a 3 (Agree), 4, or 5 (Strongly agree) (Fig. 2A). Participants were asked if they learned something new from the event, and 98% ($n = 119$) responded with a 3 (Agree), 4, or 5 (Strongly agree) (Fig. 2B). This spanned all career stages, including established scientists, where 94% of respondents reporting this career stage ($n = 15$) selected a 3 (Agree), 4, or 5 (Strongly agree), indicating that these topics and learning methods were even relevant to those more senior in their field. Participants also provided high ratings for "The materials for this event were effective for learning the content" (97%, $n = 118$ reported a 3, 4, or 5) and "I felt that my contributions and questions were welcome during the event" (98%, $n = 119$ reported a 3, 4, or 5), indicating that the training materials were effective and that the Ambassadors fostered an inclusive and collaborative environment for discussion (Fig. 2C,D).

Event attendees were provided with a series of retrospective questions where they were asked to rate their familiarity with or their perceived importance of concepts before and after the event to assess knowledge gained over the course of the event from the Ambassadors and the training content. When asked about their familiarity with the FAIR data principles, 86% ($n = 101$) of respondents reported increased familiarity after the event (Fig. 3A). Nine of the 17 respondents who did not report an increase started off with a 5 (Very familiar) rating and ended at a 5. The mode of this dataset shifted from 1 for prior to the event to 4 for after the event, and the

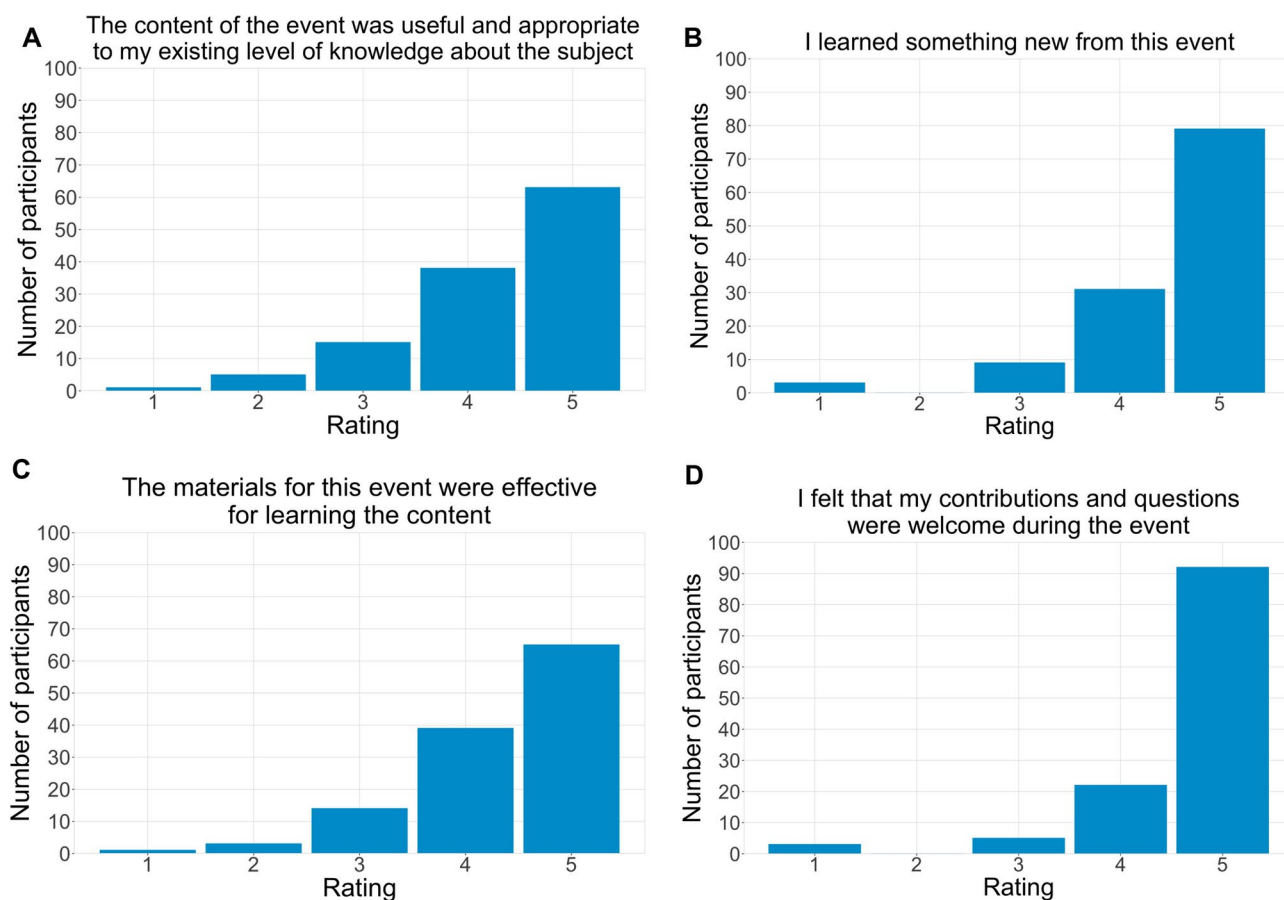


Fig. 2. Histograms representing the number of participants that reported each rating in response to Likert scale questions about overall event experience. Rating scale ranged from 1 (Strongly disagree) to 5 (Strongly agree) ($n = 122$ for total responses for all four questions). The entire *Event experience* section of the survey included six questions. The two questions that are not reported here were a yes/no/other multiple choice for a prompt about if participants could access the technology and materials (e.g., Zoom, the NMDC websites, the activity materials), and a free-form long answer text question prompting for any additional comments about overall event experience. (A) Prompt: *The content of the event was useful and appropriate to my existing level of knowledge about the subject.* (B) Prompt: *I learned something new from this event.* (C) Prompt: *The materials for this event were effective for learning the content.* (D) Prompt: *I felt that my contributions and questions were welcome during the event.*

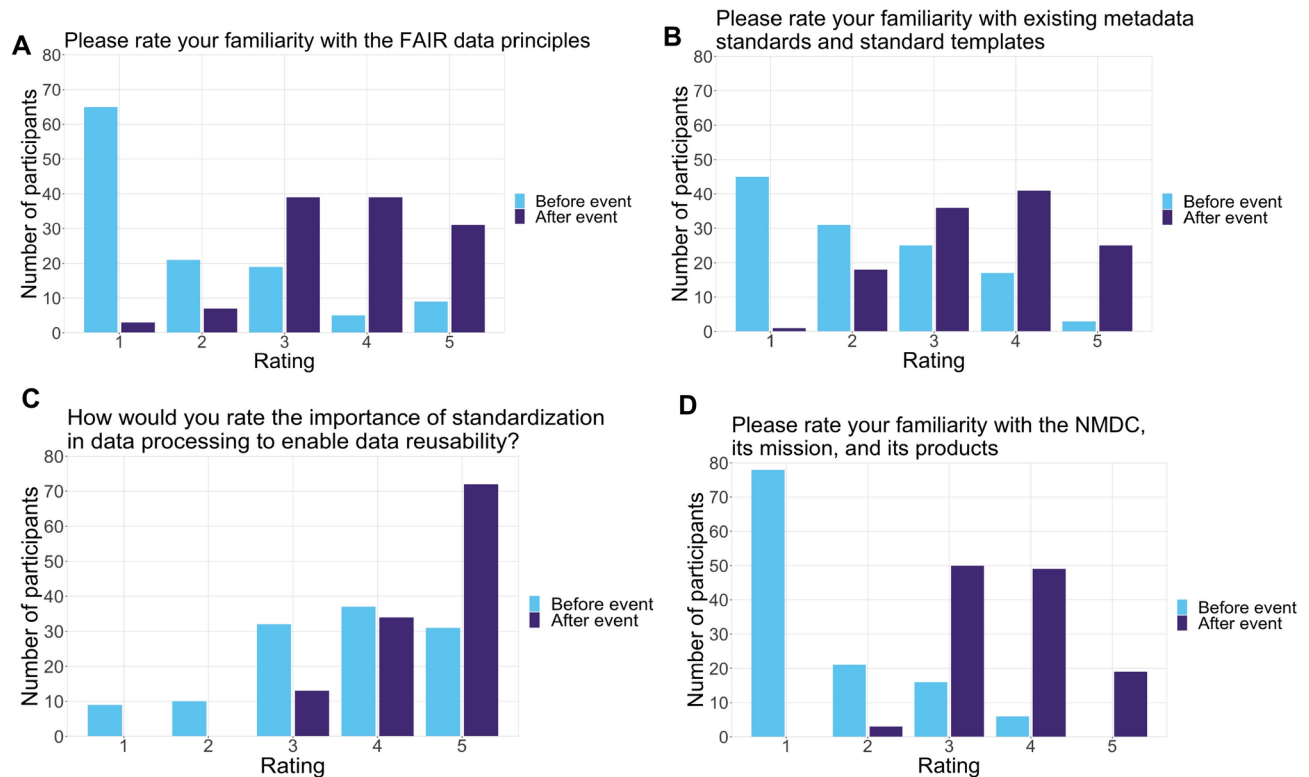


Fig. 3. Grouped histograms representing the retrospective survey data where participants were asked to rank their familiarity with a topic or their perceived importance of a topic before and after the Ambassador-led event. **(A)** Prompt: *Please rate your familiarity with the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles*; Rating scale ranged from 1 (Not familiar) to 5 (Very familiar); Response $n = 117$. **(B)** Prompt: *Please rate your familiarity with existing metadata standards and standard templates*; Rating scale ranged from 1 (Not familiar) to 5 (Very familiar); Response $n = 119$. **(C)** Prompt: *How would you rate the importance of standardization in data processing to enable data reusability?*; Rating scale ranged from 1 (Not important) to 5 (Very important); Response $n = 119$. **(D)** Prompt: *Please rate your familiarity with the NMDC, its mission, and its products*; Rating scale ranged from 1 (Not familiar) to 5 (Very familiar); Response $n = 120$.

two-tailed, paired sample t-test results indicated a significant shift in the mean from 1.93 to 3.76 ($p = 1.39E-35$). Participants were then asked to rate their familiarity with existing metadata standards and standard templates, and even though only nine of the 16 Ambassador events covered this topic in great detail, 82% ($n = 98$) of attendees reported an increase in familiarity (Fig. 3B). The mode for the responses to this question also shifted from 1 to 4, and the mean from 2.18 to 3.59 ($p = 1.31E-30$). To gauge the microbiome community's recognition of the importance of standardization in multi-omics data processing, participants were asked to rate the importance of standardization in data processing to enable data reusability (Fig. 3C). This question received the lowest percentage (58%, $n = 69$) of rating increase, as many of the event attendees already had a high level of awareness and recognition for this concept before the events (84%, $n = 100$ started at a 3, 4, or 5). However, the improvements in ratings were still significant, with a 3.60 mean for before the event and 4.50 mean for after the event ($p = 1.02E-17$).

To assess how the Ambassador events impacted recognition of the NMDC and associated activities, participants were asked to rate their familiarity with the NMDC, its mission, and its products both before and after the event (Fig. 3D). This question had the highest percentage (93%; $n = 112$) that reported an increase in their familiarity ratings. The mean response rating significantly increased from 1.58 to 3.70 ($p = 1.39E-47$). No respondent gave a 1 (Not familiar) rating after the event.

To assess the potential lasting impact of these workshops and the perception of the utility of the presented concepts, participants were asked to respond with a Likert scale rating from 1 (Strongly disagree) to 5 (Strongly agree) to the prompt, "I plan to incorporate the concepts of FAIR microbiome data, data reuse, proper data management, data stewardship, and/or data standards into my work" (Fig. 4A). Of the total responses, 99% ($n = 116$) indicated a 3, 4, or 5. Attendees were also asked about their potential continued involvement with the NMDC and its products. Seventy-one respondents indicated that they plan to continue their engagement with the NMDC by using one of the products, and twenty-one attendees expressed interest in applying for a future cohort of the Ambassador Program or the NMDC Champions Program.

To better understand the interests and needs of the microbiome research community regarding resources and training materials, participants were surveyed about the bioinformatics workflows and -omics data types they would most likely use moving forward. Participants indicated the most interest in utilizing the NMDC

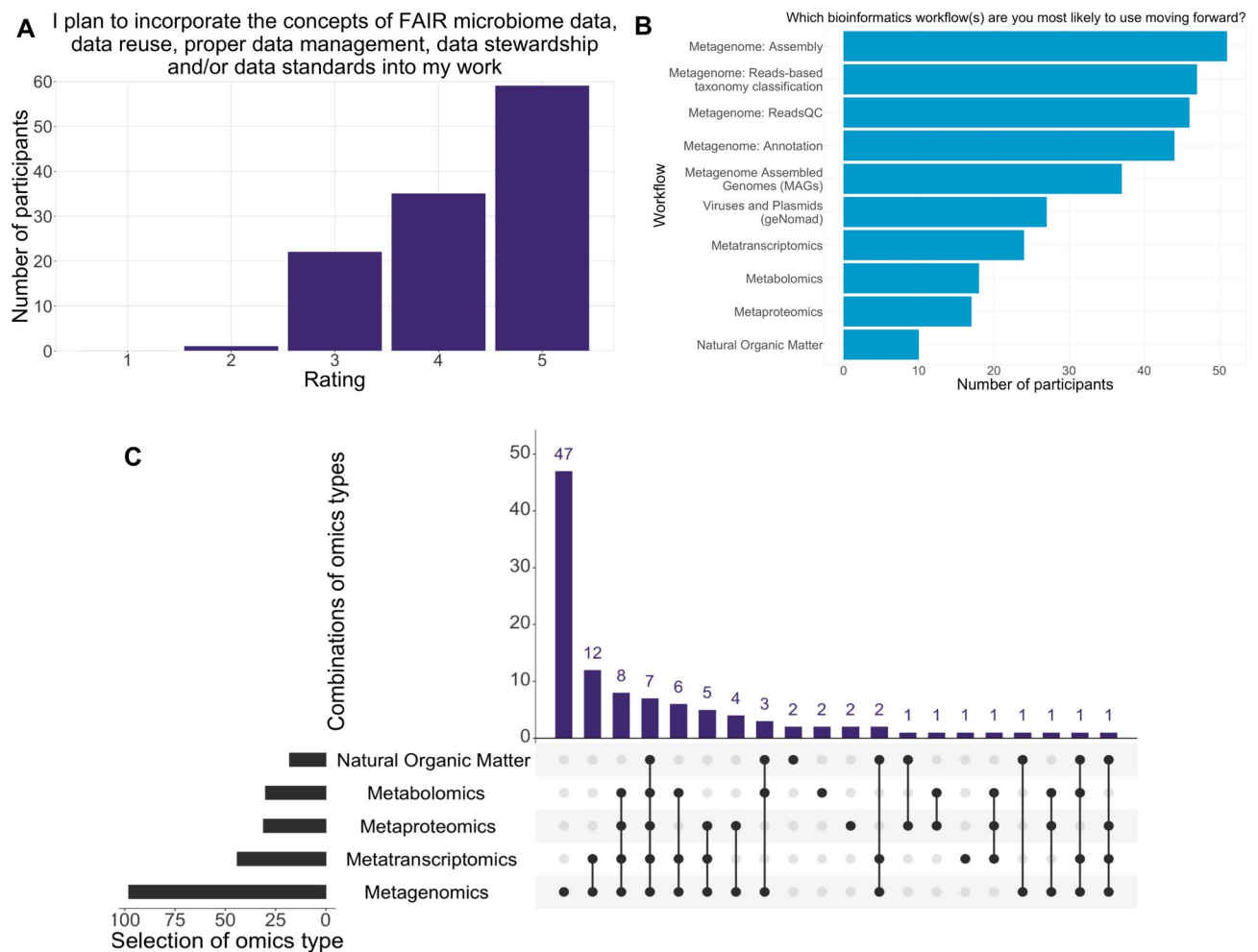


Fig. 4. Insights into participant usage of -omics data and data stewardship best practices following the event. **(A)** Histogram of Likert scale [1 (Strongly disagree) to 5 (Strongly agree)] participant responses to the prompt, “I plan to incorporate the concepts of FAIR microbiome data, data reuse, proper data management, data stewardship, and/or data standards into my work” (total responses: $n = 117$); **(B)** Number of participants that selected each response to the prompt, “Which bioinformatics workflow(s) are you most likely to use moving forward? Check all that apply” (total responses: $n = 68$); **(C)** Upset plot depicting the -omics types and combinations of -omics types selected by participants in response to the prompt “What -omics data types would you be interested in searching for/reusing from the data portal? Check all that apply” (total responses: $n = 108$).

standardized metagenome bioinformatics workflows in the future (Fig. 4B). The next most popular choice was the viruses & plasmids workflow, which takes in assemblies and provides reports about any viruses or plasmids detected in the sample, including taxonomic, quality, and antimicrobial resistance gene information³⁴. Participants were also asked, “What omics data types would you be interested in searching for/reusing from the Data Portal?” (Fig. 4C). Similarly, most respondents reported that they are most likely to reuse metagenomics data (90%, $n = 97$), and 56 of the 108 participants that responded to this question selected multiple -omics types, suggesting an interest in reusing multi-omics datasets.

While the aim of this study was to quantify the impact of the Ambassador Program, the survey also captured non-quantitative attendee feedback in the free-form text prompts. Highlights of responses include “Very well done, it was easy to follow along”; “Really nice presentation by [Ambassador]! Enjoyed and learned a lot”; “Great presentation in our language”; “Great job answering all the questions!”; “Excellent information shared, and I hope to connect with the presenters!”; “Loved the [Data Portal] scavenger hunt [activity], [it] was a very useful way to see if I actually understood”; “The organization and initiatives were great. Hope it can expand to various levels of academic research! Great resources!”. The majority of free-form text responses were positive, with a handful of neutral or more critical responses that include “I struggled to understand the material since this is my first experience with research”; “Please present the best research, not only the tools”; “Data standardization is generally a pretty dry (albeit important) topic. To compensate, I think the presentation should be more engaging. Maybe go through some processes available”. All combined survey responses are available at <https://doi.org/10.6084/m9.figshare.25045667.v1>.

Discussion

Existing educational resources for microbiome research are typically focused on technical aspects of microbiome data analysis and rarely emphasize data standards and stewardship principles^{35,36}. To expand the awareness and adoption of best practices in data stewardship, management, and standards, there is a need for resources and coordinated programs focused on these core themes that encompass the interdisciplinary nature of microbiome research. To have the greatest impact on microbiome research practices for future projects, these programs should ideally target early-career researchers and diverse audiences, and be intentionally structured to maximize the learning of these concepts throughout the community. The NMDC Ambassador Program previously demonstrated its utility for expanding the reach of these training materials¹⁷, and the quantitative results described herein demonstrate the value of community learning models towards increasing awareness across the field. All of the post-event survey retrospective question responses showed an overwhelmingly positive trend towards an increased understanding of metadata standards, bioinformatics workflows, standardization, the FAIR principles, and the NMDC program following the Ambassador events. This included researchers from various career stages, institutions, and microbiome environments, demonstrating the Program's efficacy across experience levels and backgrounds.

Several challenges and lessons learned were encountered and documented during the 2023 Ambassador workshops. While hybrid formats can make events more equitable and accessible, they presented challenges for the Ambassadors similar to what others in the field have encountered³⁷. Virtual attendees reported issues hearing questions and answers and experiencing a lack of virtual engagement. To mitigate some of these challenges with virtual and hybrid events, Ambassadors were able to request support from the NMDC team, and team members attended many of these events to assist with monitoring questions and providing links. Not all events allotted adequate time for the survey, and some participants had issues accessing the survey form (e.g., could not connect to the internet, could not access Google Forms due to a firewall). Although the survey responses indicate that the content was broadly applicable across career stages, there was some constructive feedback, for example: "I struggled to understand the material since this is my first experience with research", that led to discussions on how to best modify event content depending on experience levels. Check-in calls occurred with the Ambassadors throughout their term to facilitate discussion, leading to valuable insights that enabled adaptation throughout the year to subsequent workshops. This prompted exchanges between Ambassadors ranging from "Make sure to tell participants to bring their laptops for hands-on activities" to "I underestimated how much coffee to order". Overall, this study was limited by a lack of participation in the post-event surveys and its reliance on self-reporting. Further, we acknowledge that our analysis pooled the survey results from diverse events and workshop types with a range of participants (e.g., 4 to 45 participants). Future efforts will focus on gathering survey information from more participants to improve the representation of feedback. Additionally, the retrospective survey methodology does have known limitations, but it was chosen to minimize the number of incomplete datasets from participants only answering either a pre- or post-event survey, and to minimize biases with what participants think they know about a topic prior to the educational event^{24–26}.

Beyond the broader concepts, the results also indicate that this 'train-the-trainers' model is effective for communicating the NMDC mission and products to diverse audiences, and expands upon the number and reach of events that would have been possible by NMDC team members alone. The development of the NMDC products relies on user input to ensure we are addressing community needs, and the Ambassador events and survey provided valuable feedback to the team from diverse groups and research domains outside of our direct network. The data about bioinformatics workflows and usage of multi-omics data is indispensable for understanding current and future user needs. The fact that over twenty respondents indicated interest in applying to the NMDC Ambassador or Champions Programs also highlighted the success of these events for promoting interest in continued engagement across microbiome research communities.

Conclusions

Educational efforts such as the NMDC Ambassador Program will continue to be invaluable to microbiome science, as they expand the distribution of microbiome research best practices to diverse audiences and institutions. Here, we quantitatively measured how the 2023 Ambassador events increased awareness of standardization efforts and data stewardship practices. Implementing and adhering to microbiome data stewardship best practices, FAIR data principles, and standardization efforts will undoubtedly lead to overall improvements in the generation and utilization of microbiome data, thus increasing scientific outputs and innovation both in the short term and in years to come. Hands-on workshops like those presented here will continue to be critical in microbiome workforce development and will contribute to training the next generation of microbiome scientists. Insights from the NMDC Ambassador Program and its post-event survey can be a generally useful model to consider in other training programs.

Data availability

The survey utilized in this study as well as the survey results are available through Figshare, <https://doi.org/http://doi.org/10.6084/m9.figshare.25045667.v2>. The raw survey result data was anonymized and partially redacted to protect participant privacy.

Received: 20 November 2024; Accepted: 10 February 2025

Published online: 22 March 2025

References

1. Abreu, A. et al. Priorities for ocean microbiome research. *Nat. Microbiol.* **7**, 937–947 (2022).

2. Huttenhower, C. et al. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
3. Pfister, C. A. et al. Conceptual exchanges for understanding free-living and host-associated microbiomes. *mSystems* **7**, e0137421 (2022).
4. Stulberg, E. et al. An assessment of US microbiome research. *Nat. Microbiol.* **1**, 1–7 (2016).
5. Suman, J. et al. Microbiome as a key player in sustainable agriculture and human health. *Front. Soil. Sci.* <https://doi.org/10.3389/fsoil.2022.821589> (2022).
6. Ferrocino, I. et al. The need for an integrated multi-OMICS approach in microbiome science in the food system. *Comp. Rev. Food Sci. Food Safe.* **22**, 1082–1103 (2023).
7. Wensel, C. R., Pluznick, J. L., Salzberg, S. L. & Sears, C. L. Next-generation sequencing: insights to advance clinical investigations of the microbiome. *J. Clin. Invest.* <https://doi.org/10.1172/JCI154944> (2022).
8. Zhang, X., Li, L., Butcher, J., Stintzi, A. & Figey, D. Advancing functional and translational microbiome research using meta-omics approaches. *Microbiome* **7**, 154 (2019).
9. Hornung, B. V. H., Zwiittink, R. D. & Kuijper, E. J. Issues and current standards of controls in microbiome research. *FEMS Microbiol. Ecol.* **95**, f045 (2019).
10. Huttenhower, C., Finn, R. D. & McHardy, A. C. Challenges and opportunities in sharing microbiome data and analyses. *Nat. Microbiol.* **8**, 1960–1970 (2023).
11. Amos, G. C. A. et al. Developing standards for the microbiome field. *Microbiome* **8**, 98 (2020).
12. Mandal, R. et al. Workshop report: Toward the development of a human whole stool reference material for metabolomic and metagenomic gut microbiome measurements. *Metabolomics* **16**, 119 (2020).
13. Mirzayi, C. et al. Reporting guidelines for human microbiome research: the STORMS checklist. *Nat. Med.* **27**, 1885–1892 (2021).
14. Yilmaz, P. et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIXS) specifications. *Nat. Biotechnol.* **29**, 415–420 (2011).
15. Vangay, P. et al. Microbiome metadata standards: Report of the national microbiome data collaborative's workshop and follow-on activities. *mSystems* **6**, e01194-e1220 (2021).
16. Wilkinson, M. D. et al. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
17. Kelliher, J. M. et al. Cohort-based learning for microbiome research community standards. *Nat. Microbiol.* **8**, 751–753 (2023).
18. Dillon, M. R. et al. Experiences and lessons learned from two virtual, hands-on microbiome bioinformatics workshops. *PLoS Computat. Biol.* **17**, e1009056 (2021).
19. Shade, A., Dunivin, T. K., Choi, J., Teal, T. K. & Howe, A. C. Strategies for building computing skills to support microbiome analysis: a five-year perspective from the EDAMAME workshop. *mSystems* <https://doi.org/10.1128/mSystems.00297-19> (2019).
20. Teal, T. K. et al. Data carpentry: Workshops to increase data literacy for researchers. *International Journal of Digital Curation.* **10**, 135–143 (2015).
21. Wilson G. *Software Carpentry: lessons learned*. *F1000Research*. <https://f1000research.com/articles/3-62> (2016).
22. Kelliher, J. M. et al. Standardized and accessible multi-omics bioinformatics workflows through the NMDC EDGE resource. *Computational and Structural Biotechnology Journal*, **23**, 3575–3583. <https://doi.org/10.1016/j.csbj.2024.09.018> (2024).
23. Elloe-Fadrosh, E. A. et al. The national microbiome data collaborative data portal: an integrated multi-omics microbiome data resource. *Nucleic Acids Res.* **50**, D828–D836 (2022).
24. Howard, G. S. et al. Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. *Appl. Psychol. Meas.* **3**, 1–23 (1979).
25. Pratt, C. C., McGuigan, W. M. & Katzev, A. R. Measuring program outcomes: Using retrospective pretest methodology. *Am. J. Eval.* **21**, 341–349 (2000).
26. Raidl, M., Johnson, S., Gardiner, K. & Denham, M. Use retrospective surveys to obtain complete data sets and measure impact in extension programs. *The Journal of Extension*, **42**(2), 13 (2004).
27. Kelliher, J., Rodriguez, F., Johnson, L., Ockert, I., Roux, S., Elloe-Fadrosh, E., et al. 2023 NMDC ambassador presentations. <https://zenodo.org/records/10015793> (2023).
28. Rodríguez-Ramos, J., Kelliher, J., Rodríguez, F., Johnson, L. & Elloe-Fadrosh, E. Standardized workflows and NMDC EDGE training: Spanish translation. <https://zenodo.org/records/10014901> (2023).
29. Likert, R. A technique for the measurement of attitudes. *Archives of psychology* (1932).
30. Racine, J. S. Rstudio: A platform-independent ide for R and sweave. *J. Appl. Econ.* **27**, 167–172 (2012).
31. Wickham, H. *ggplot2* (International Publishing, 2016).
32. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).
33. Johnson, M. D. L. & Knox, C. J. National summer undergraduate research project (NSURP): A virtual research experience to deliver REAL (retention, equity, access, and life-changing) outcomes for underrepresented minorities in STEM. *J. Microbiol. Biol. Educ.* **23**, e00335-e421 (2022).
34. Camargo, A. P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., et al. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* **1–10** (2023).
35. Mitchell, K. et al. PUMAA: A platform for accessible microbiome analysis in the undergraduate classroom. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2020.584699/full> (2020).
36. Rosen, G. L. & Hammrich, P. Teaching microbiome analysis: From design to computation through inquiry. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2020.528051/full> (2020).
37. Fulcher, M. R. et al. Broadening participation in scientific conferences during the era of social distancing. *Trends Microbiol.* **28**, 949–952 (2020).

Author contributions

JK and FR ran the 2023 Ambassador Program, wrote the manuscript draft, performed all analyses, and generated the figures. LJ, SR, MS, AC, WL, and EEF supported the Ambassador Program activities, contributed to training materials, assisted with Program logistics, and helped with the original draft of the manuscript. CHB, SF, IK, EAK, HAL, RL, RM, TP, JRR, JS, DS, JS, and AY are the 2023 NMDC Ambassadors who hosted the referenced workshops, administered the survey, and contributed to the drafting and editing of the manuscript. All authors read and approved the manuscript.

Funding

The work conducted by the National Microbiome Data Collaborative (<https://ror.org/05cwx3318>) 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), and DE-AC05-76RL01830 (PNNL). LA-UR-24-27381.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

The survey was reviewed and approved by the Human Subjects Committee at Lawrence Berkeley National Laboratory as an exempt IRB protocol under #394NR002.

Additional information

Correspondence and requests for materials should be addressed to J.M.K., F.E.R. or E.A.E.-F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025