

MIT Open Access Articles

*Reproducible, interactive, scalable and extensible
microbiome data science using QIIME 2*

The MIT Faculty has made this article openly available. **Please share**
how this access benefits you. Your story matters.

Citation: Bolyen, Evan et al. "Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2." *Nature Biotechnology* 37, 8 (July 2019): 852–857 © 2019 Springer Nature

As Published: <http://dx.doi.org/10.1038/s41587-019-0209-9>

Publisher: Springer Science and Business Media LLC

Persistent URL: <https://hdl.handle.net/1721.1/125325>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike





Published in final edited form as:

Nat Biotechnol. 2019 August ; 37(8): 852–857. doi:10.1038/s41587-019-0209-9.

Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2

A full list of authors and affiliations appears at the end of the article.

To the Editor — Rapid advances in DNA-sequencing and bioinformatics technologies in the past two decades have substantially improved understanding of the microbial world. This growing understanding relates to the vast diversity of microorganisms; how microbiota and microbiomes affect disease¹ and medical treatment²; how microorganisms affect the health of the planet³; and the nascent exploration of the medical⁴, forensic⁵, environmental⁶ and agricultural⁷ applications of microbiome biotechnology. Much of this work has been driven by marker-gene surveys (for example, bacterial/archaeal 16S rRNA genes, fungal internal-transcribed-spacer regions and eukaryotic 18S rRNA genes), which profile microbiota with varying degrees of taxonomic specificity and phylogenetic information. The field is now transitioning to integrate other data types, such as metabolite⁸, metaproteome⁹ or metatranscriptome^{9,10} profiles.

The QIIME 1 microbiome bioinformatics platform has supported many microbiome studies and gained a broad user and developer community. Interactions with QIIME 1 users in our online support forum, our workshops and direct collaborations have shown the platform's potential to serve an increasingly diverse array of microbiome researchers in academia, government and industry. Here, we present QIIME 2, a completely reengineered and rewritten system that is expected to facilitate reproducible and modular analysis of microbiome data to enable the next generation of microbiome science.

QIIME 2 was developed on the basis of a plugin architecture (Supplementary Fig. 1) that allows third parties to contribute functionality (<https://library.qiime2.org>). QIIME 2 plugins exist for latest-generation tools for sequence quality control from different sequencing platforms (DADA2 (ref.¹¹) and Deblur¹²), taxonomy assignment¹³ and phylogenetic insertion¹⁴, which quantitatively improve the results over QIIME 1 and other tools (as detailed in the corresponding tool-specific publications). The plugins also support

* greg.caporaso@nau.edu.

Author contributions

E.B., J.R.R., M.R.D., N.A.B., Y.B., J.E.B., C.J.B., A.M.C.-R., E.K.C., C. Diener, R.D., C.F.E., M. Ernst, M. Estaki, A.G., J.M.G., D.L.G., S.M.G., A.K.J., K.B.K., S.T.K., I.K., T.K., J.L., Y.-X.L., A.V.M., J.L.M., L.F.N., S.B.O., D.P., A.S., S.J.S., A.D.S., L.R.T., P. J. Torres, P. J. Turnbaugh, S.U.-H., F.V., J.W., R.K. and J.G.C. developed documentation, educational materials and/or user/developer support content. E.B., J.R.R., M.R.D., N.A.B., R.K. and J.G.C. wrote the manuscript; all authors assisted with revision of the manuscript. E.B., J.R.R., M.R.D., N.A.B. and J.G.C. designed and developed the QIIME 2 framework. D.M.D., A.G., R.L., E.L., S.C.M., R.S., J.R.S., W.W., C.H.D.W. and R.K. contributed data used in the manuscript and/or testing of QIIME 2. C.C.A., C.T.B., E.K.C., P.C.D., S.H., P.K., E.L., T.P., R.S., E.V., Y.W. and R.K. contributed to the design of analytical methods. E.B., J.R.R., M.R.D., N.A.B., G.A.A.-G., H.A., E.J.A., M.A., F.A., K.B., A.B., B.J.C., J.C., G.M.D., C. Duvallet, M. Ernst, J.F., A.G., K.G., J.G., S.M.G., B.H., H.H., C.H., G.H., S.J., L.J., B.D.K., C.R.K., D.K., J.K., M.G.I.L., C.L., M.M., C.M., B.D.M., D.M., L.J.M., J.T.M., A.T.N., J.A.N.-M., S.L.P., M.L.P., E.P., L.B.R., A.R., M.S.R., P.R., N.S., M.S., P.T., A.T., J.J.J.v.d.H., Y.V.-B., M.V., M.W., K.C.W., A.D.W., Z.Z.X., J.R.Z., Y.Z., Q.Z. and J.G.C. contributed software to QIIME 2 plugins, interfaces, framework and/or build and test systems.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41587-019-0209-9>.

qualitatively new functionality, including microbiome paired-sample and time-series analysis¹⁵ (which are critical for studying the effects of treatments on the microbiome), and machine learning¹⁶. Trained machine learning models can be saved for application to new data and interrogated to identify important microbiome features. Several recently released plugins, including q2-cscs¹⁷, q2-metabolomics¹⁸, q2-shogun¹⁹, q2-metaphlan2 (ref.²⁰) and q2-picrust2 (ref.²¹), provide initial support for analysis of metabolomics and shotgun metagenomics data. We are currently working with teams developing bioinformatics tools for metatranscriptomics and metaproteomics, and we expect to add new plugins supporting these data types to the ecosystem shortly. Additionally, many of the existing ‘downstream’ analysis tools, such as q2-sample-classifier¹⁶, can already work with these data types individually or in combination if they are provided in a feature table. Thus, QIIME 2 has the potential to serve not only as a marker-gene analysis tool but also a multidimensional and powerful data science platform that can be rapidly adapted to analyze diverse microbiome features.

QIIME 2 provides many new interactive visualization tools facilitating exploratory analyses and result reporting. Static versions of interactive visualizations resulting from four worked examples are provided in Fig. 1. QIIME 2 View (<https://view.qiime2.org>) is a unique new service (Supplementary Methods) that allows users to securely share and interact with results without installing QIIME 2. The QIIME 2 visualizations presented in Fig. 1 are provided in Supplementary File 1 to allow readers to interact with QIIME 2 View. Corresponding worked QIIME 2 example code is provided in the Supplementary Methods.

Reproducibility, transparency and clarity of microbiome data science are guiding principles in QIIME 2 design. To this end, QIIME 2 includes a decentralized data-provenance tracking system: details of all analysis steps with references to intermediate data are automatically stored in the results. Users can thus retrospectively determine exactly how any result was generated (Fig. 2 illustrates a simplified provenance graph derived from the data provenance of Fig. 1b). QIIME 2 also detects corrupted results indicating that the provenance is no longer reliable and the results no longer contain information enabling reproducibility. The provenance of the visualizations presented in Fig. 1 can be interactively reviewed by loading the contents of Supplementary File 1 with QIIME 2 View, providing far more detailed information than can typically be provided in Methods text. QIIME 2 results are also semantically typed (Fig. 2), and actions indicate acceptable input types, clarifying the data that actions should be applied to and making complex workflows less error prone. Complex workflows can be created and shared by using Jupyter Notebooks²² or Common Workflow Language (CWL)²³, and support for other workflow engines is currently in development.

Finally, QIIME 2 provides a software-development kit (<https://dev.qiime2.org>) that can be used to integrate it as a component of other systems (such as Qiita²⁴ or Illumina BaseSpace) and to develop interfaces targeted toward users with different levels of computational sophistication (Supplementary Fig. 2). QIIME 2 provides the QIIME 2 Studio graphical user interface and QIIME 2 View, interfaces designed for end-user biologists, clinicians and policy-makers; the QIIME 2 application programming interface, designed for data scientists who want to automate workflows or work interactively in Jupyter Notebooks²²; and q2cli and q2cwl, providing a command-line interface and CWL²³ wrappers for QIIME 2,

designed for experts in high-performance computing. At present, computationally expensive steps support parallel computing at the individual-action level (for example, many actions including de-noising and taxonomy assignment support multiple threads). We are currently developing deeper integration with parallelism strategies available in third-party workflow engines, and workflow-level parallelism is currently possible through CWL.

There are many other powerful open-source software tools for microbiome data science, including mothur²⁵, phyloseq²⁶ and related tools available through Bioconductor²⁷, and the biobakery suite^{20,21,28}. The microbiome bioinformatics platform mothur is often compared to QIIME 1 and QIIME 2. A major difference between mothur and QIIME lies in the interactive visualizations: QIIME 2 provides many interactive visualization tools (several examples are provided in Fig. 1), whereas mothur focuses on generating data that can be easily loaded and visualized with other tools. The phyloseq tool focuses on microbiome statistical analysis and generating publication-ready visualizations but, unlike QIIME 2, begins with a feature or operational-taxonomic-unit table, leaving ‘upstream’ processing steps, such as sequence demultiplexing and quality control, to other processing pipelines, many of which (like phyloseq) are available through Bioconductor. The biobakery suite provides analytic functionality that complements that of QIIME 2, and we are actively working with biobakery developers to support interoperability by making their tools accessible as QIIME 2 plugins (for example, the q2-metaphlan2 plugin allows users to run MetaPhlAn2 through QIIME 2). QIIME 2 provides the only Python-based microbiome data-science platform that supports retrospective data-provenance tracking to ensure reproducibility, multi-omics analysis support, interfaces geared toward different user types to enhance usability and an extensibility-focused design through the plugin architecture and software-development kit. We share feedback from users of QIIME 2 on these and other features in Supplementary Methods.

The tools described in the preceding paragraph are all interoperable through plugins, exchange of files in standard formats or using multi-language environments, such as Jupyter Notebooks²². For example, the BIOM format²⁹ is supported by all of them. A diverse ecosystem of interoperable software is beneficial for the field, because it allows both experienced users to obtain multiple perspectives on their data and novice bioinformaticians to work in the programming environments that they are most comfortable with (for example, phyloseq allows users to work in R, whereas QIIME 2 allows users to work in Python). We plan to continue working with the developers of these tools, and with organizations such as the Genomics Standards Consortium, on plugins and standards to ensure interoperability, as well as developing tools to automatically import data from microbiome data-sharing platforms such as Qiita, the European Bioinformatics Institute (EBI) European Read Archive and the National Center for Biotechnology Information (NCBI) Sequence Read Archive.

Advances in microbiome research promise to improve many aspects of health and the world, and QIIME 2 will help drive those advances by enabling accessible, community-driven microbiome data science.

Data availability

Data for the analyses presented in Fig. 1 are available as follows: Earth Microbiome Project data in Fig. 1a were obtained from <ftp://ftp.microbio.me/emp/release1>, and the American Gut Project (AGP) data were obtained from Qiita (<http://qiita.microbio.me>) study ID 10317. Sequence data in Fig. 1c are available in Qiita under study ID 10249 and the EBI under accession number ERP016173. Sequence data in Fig. 1b are available in Qiita under study ID 925 and the EBI under accession number ERP022167. Data in Fig. 1d are available in the q2-ili GitHub repository (<https://github.com/biocore/q2-ili>). Interactive versions of the Fig. 1 visualizations can be accessed at <https://github.com/qiime2/paper1>.

Code availability

QIIME 2 is open source and free for all use, including commercial. It is licensed under a BSD three-clause license. Source code is available at <https://github.com/qiime2>. Help for QIIME 2 is provided at <https://forum.qiime2.org>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Evan Bolyen^{1,80}, Jai Ram Rideout^{1,80}, Matthew R. Dillon^{1,80}, Nicholas A. Bokulich^{1,80}, Christian C. Abnet², Gabriel A. Al-Ghalith³, Harriet Alexander^{4,5}, Eric J. Alm^{6,7}, Manimozhiyan Arumugam⁸, Francesco Asnicar⁹, Yang Bai^{10,11,12}, Jordan E. Bisanz¹³, Kyle Bittinger^{14,15}, Asker Brejnrod⁸, Colin J. Brislawn¹⁶, C. Titus Brown⁵, Benjamin J. Callahan^{17,18}, Andrés Mauricio Caraballo-Rodríguez¹⁹, John Chase¹, Emily K. Cope^{1,20}, Ricardo Da Silva¹⁹, Christian Diener²¹, Pieter C. Dorrestein¹⁹, Gavin M. Douglas²², Daniel M. Durall²³, Claire Duvallet⁶, Christian F. Edwards²⁴, Madeleine Ernst^{19,25}, Mehrbod Estaki²⁶, Jennifer Fouquier^{27,28}, Julia M. Gauglitz¹⁹, Sean M. Gibbons^{21,29}, Deanna L. Gibson^{30,31}, Antonio Gonzalez³², Kestrel Gorlick¹, Jiarong Guo³³, Benjamin Hillmann³⁴, Susan Holmes³⁵, Hannes Holste^{32,36}, Curtis Huttenhower^{37,38}, Gavin A. Huttley³⁹, Stefan Janssen⁴⁰, Alan K. Jarmusch¹⁹, Lingjing Jiang⁴¹, Benjamin D. Kaehler^{39,42}, Kyo Bin Kang^{19,43}, Christopher R. Keefe¹, Paul Keim¹, Scott T. Kelley⁴⁴, Dan Knights^{34,45}, Irina Koester^{19,46}, Tomasz Kosciolk⁴⁷, Jorden Kreps¹, Morgan G. I. Langille⁴⁸, Joslynn Lee⁴⁹, Ruth Ley^{50,51}, Yong-Xin Liu^{10,11}, Erika Lofffield², Catherine Lozupone²⁸, Massoud Maher⁵², Clarisse Marotz³², Bryan D. Martin⁵³, Daniel McDonald³², Lauren J. McIver^{37,38}, Alexey V. Melnik¹⁹, Jessica L. Metcalf⁵⁴, Sydney C. Morgan⁵⁵, Jamie T. Morton^{32,52}, Ahmad Turan Naimey¹, Jose A. Navas-Molina^{32,52,56}, Louis Felix Nothias¹⁹, Stephanie B. Orchanian⁵⁷, Talima Pearson¹, Samuel L. Peoples^{58,59}, Daniel Petras¹⁹, Mary Lai Preuss⁶⁰, Elmar Priesse²⁸, Lasse Buur Rasmussen⁸, Adam Rivers⁶¹, Michael S. Robeson II⁶², Patrick Rosenthal⁶⁰, Nicola Segata⁹, Michael Shaffer^{27,28}, Arron Shiffer¹, Rashmi Sinha², Se Jin Song³², John R. Spear⁶³, Austin D. Swafford⁵⁷, Luke R. Thompson^{64,65},

Pedro J. Torres⁶⁶, Pauline Trinh⁶⁷, Anupriya Tripathi^{19,32,68}, Peter J. Turnbaugh⁶⁹, Sabah Ul-Hasan⁷⁰, Justin J. J. vander Hooff⁷¹, Fernando Vargas⁶⁸, Yoshiki Vázquez-Baeza³², Emily Vogtmann², Max von Hippel⁷², William Walters⁵⁰, Yunhu Wan², Mingxun Wang¹⁹, Jonathan Warren⁷³, Kyle C. Weber^{61,74}, Charles H. D. Williamson⁷⁵, Amy D. Willis⁷⁶, Zhenjiang Zech Xu³², Jesse R. Zaneveld⁷⁷, Yilong Zhang⁷⁸, Qiyun Zhu³², Rob Knight^{32,57,79}, J. Gregory Caporaso^{1,20,*}

Affiliations

¹Center for Applied Microbiome Science, Pathogen and Microbiome Institute, Northern Arizona University, Flagstaff, AZ, USA. ²Metabolic Epidemiology Branch, National Cancer Institute, Rockville, MD, USA. ³Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA. ⁴Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. ⁵Department of Population Health and Reproduction, University of California, Davis, Davis, CA, USA. ⁶Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁷Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁸Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ⁹Centre for Integrative Biology, University of Trento, Trento, Italy. ¹⁰State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. ¹¹Centre of Excellence for Plant and Microbial Sciences (CEPAMS), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences & John Innes Centre, Beijing, China. ¹²University of Chinese Academy of Sciences, Beijing, China. ¹³Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, CA, USA. ¹⁴Division of Gastroenterology and Nutrition, Children's Hospital of Philadelphia, Philadelphia, PA, USA. ¹⁵Hepatology, Children's Hospital of Philadelphia, Philadelphia, PA, USA. ¹⁶Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA, USA. ¹⁷Department of Population Health & Pathobiology, North Carolina State University, Raleigh, NC, USA. ¹⁸Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA. ¹⁹Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, San Diego, CA, USA. ²⁰Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, USA. ²¹Institute for Systems Biology, Seattle, WA, USA. ²²Department of Microbiology and Immunology, Dalhousie University, Halifax, Nova Scotia, Canada. ²³Irving K. Barber School of Arts and Sciences, University of British Columbia, Kelowna, British Columbia, Canada. ²⁴A. Watson Armour III Center for Animal Health and Welfare, Aquarium Microbiome Project, John G. Shedd Aquarium, Chicago, IL, USA. ²⁵Department of Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. ²⁶Department of Biology, University of British Columbia Okanagan, Okanagan, British Columbia, Canada. ²⁷Computational Bioscience Program, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. ²⁸Department of Medicine, Division of Biomedical Informatics and Personalized Medicine, University of

Colorado Anschutz Medical Campus, Aurora, CO, USA. ²⁹eScience Institute, University of Washington, Seattle, WA, USA. ³⁰Irving K. Barber School of Arts and Sciences, Department of Biology, University of British Columbia, Kelowna, British Columbia, Canada. ³¹Department of Medicine, University of British Columbia, Kelowna, British Columbia, Canada. ³²Department of Pediatrics, University of California San Diego, La Jolla, CA, USA. ³³Center for Microbial Ecology, Michigan State University, East Lansing, MI, USA. ³⁴Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA. ³⁵Statistics Department, Stanford University, Palo Alto, CA, USA. ³⁶Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA. ³⁷Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. ³⁸Broad Institute of MIT and Harvard, Cambridge, MA, USA. ³⁹Research School of Biology, The Australian National University, Canberra, Australian Capital Territory, Australia. ⁴⁰Department of Pediatric Oncology, Hematology and Clinical Immunology, Heinrich-Heine University Dusseldorf, Dusseldorf, Germany. ⁴¹Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA, USA. ⁴²School of Science, University of New South Wales, Canberra, Australian Capital Territory, Australia. ⁴³College of Pharmacy, Sookmyung Women's University, Seoul, Republic of Korea. ⁴⁴Department of Biology, San Diego State University, San Diego, CA, USA. ⁴⁵Biotechnology Institute, University of Minnesota, Saint Paul, MN, USA. ⁴⁶Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA. ⁴⁷Department of Pediatrics, University of California San Diego, La Jolla, California, USA. ⁴⁸Department of Pharmacology, Dalhousie University, Halifax, Nova Scotia, Canada. ⁴⁹Science Education, Howard Hughes Medical Institute, Ashburn, VA, USA. ⁵⁰Department of Microbiome Science, Max Planck Institute for Developmental Biology, Tübingen, Germany. ⁵¹Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA. ⁵²Department of Computer Science & Engineering, University of California San Diego, La Jolla, CA, USA. ⁵³Department of Statistics, University of Washington, Seattle, WA, USA. ⁵⁴Department of Animal Science, Colorado State University, Fort Collins, CO, USA. ⁵⁵Irving K. Barber School of Arts and Sciences, Unit 2 (Biology), University of British Columbia, Kelowna, British Columbia, Canada. ⁵⁶Google LLC, Mountain View, CA, USA. ⁵⁷Center for Microbiome Innovation, University of California San Diego, La Jolla, CA, USA. ⁵⁸School of Information Studies, Syracuse University, Syracuse, NY, USA. ⁵⁹School of STEM, University of Washington Bothell, Bothell, WA, USA. ⁶⁰Department of Biological Sciences, Webster University, St. Louis, MO, USA. ⁶¹Agricultural Research Service, Genomics and Bioinformatics Research Unit, United States Department of Agriculture, Gainesville, FL, USA. ⁶²College of Medicine, Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, USA. ⁶³Department of Civil and Environmental Engineering, Colorado School of Mines, Golden, CO, USA. ⁶⁴Department of Biological Sciences and Northern Gulf Institute, University of Southern Mississippi, Hattiesburg, MS, USA. ⁶⁵Ocean Chemistry and Ecosystems Division, Atlantic Oceanographic and Meteorological Laboratory, National Oceanic

and Atmospheric Administration, La Jolla, CA, USA. ⁶⁶Department of Biology, San Diego State University, San Diego, CA, USA. ⁶⁷Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA, USA. ⁶⁸Division of Biological Sciences, University of California San Diego, San Diego, CA, USA. ⁶⁹Department of Microbiology and Immunology, University of California San Francisco, San Francisco, CA, USA. ⁷⁰Quantitative and Systems Biology Graduate Program, University of California Merced, Merced, CA, USA. ⁷¹Bioinformatics Group, Wageningen University, Wageningen, the Netherlands. ⁷²Department of Mathematics, University of Arizona, Tucson, AZ, USA. ⁷³National Laboratory Service, Environment Agency, Starcross, UK. ⁷⁴College of Agriculture and Life Sciences, University of Florida, Gainesville, FL, USA. ⁷⁵Pathogen and Microbiome Institute, Northern Arizona University, Flagstaff, AZ, USA. ⁷⁶Department of Biostatistics, University of Washington, Seattle, WA, USA. ⁷⁷School of STEM, Division of Biological Sciences, University of Washington Bothell, Bothell, WA, USA. ⁷⁸Merck & Co. Inc., Kenilworth, NJ, USA. ⁷⁹Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA. ⁸⁰These authors contributed equally: Evan Bolyen, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich.

Acknowledgements

QIIME 2 development was primarily funded by NSF Awards 1565100 to J.G.C. and 1565057 to R.K. Partial support was also provided by the following: grants NIH U54CA143925 (J.G.C. and T.P.) and U54MD012388 (J.G.C. and T.P.); grants from the Alfred P. Sloan Foundation (J.G.C. and R.K.); ERCSTG project MetaPG (N.S.); the Strategic Priority Research Program of the Chinese Academy of Sciences QYZDB-SSW-SMC021 (Y.B.); the Australian National Health and Medical Research Council APP1085372 (G.A.H., J.G.C., Von Bing Yap and R.K.); the Natural Sciences and Engineering Research Council (NSERC) to D.L.G.; and the State of Arizona Technology and Research Initiative Fund (TRIF), administered by the Arizona Board of Regents, through Northern Arizona University. All NCI coauthors were supported by the Intramural Research Program of the National Cancer Institute. S.M.G. and C. Diener were supported by the Washington Research Foundation Distinguished Investigator Award. Thanks to the Yellowstone Center for Resources for research permit no. 5664 to J.R.S. for Yellowstone access and sample collection. We thank P. J. McMurdie for helpful discussion on the relationships between QIIME 2 and phyloseq. We would like to thank the users of QIIME 1 and 2, whose invaluable feedback has shaped QIIME 2. In particular, we would like to thank A. Abdelfattah (Stockholm University, Sweden), R. C. T. Boutin (University of British Columbia, Canada), D. J. Bradshaw II (Florida Atlantic University Harbor Branch Oceanographic Institute, USA), L. Bullington (MPG Ranch, USA), J. W. Debelius (Karolinska Institutet, Sweden), C. Duvall (Massachusetts Institute of Technology, USA), E. Korzune Ganda (Cornell University, USA), A. Mahner (Medical University of Graz, Austria), M. C. Melendrez (St. Cloud State University, USA), D. O'Rourke (University of New Hampshire, USA), A. R. Rivers (USDA ARS, USA), B. Sen (Tianjin University, China), S. Tangedal (Haukeland University Hospital and University of Bergen, Norway), P. J. Torres (San Diego State University, USA) and J. Warren (National Laboratory Service, UK) for writing end-user reviews included in the Supplementary Methods.

References

1. Smith MI et al. *Science* 339, 548–554 (2013). [PubMed: 23363771]
2. Gopalakrishnan V et al. *Science* 359, 97–103 (2018). [PubMed: 29097493]
3. Gehring CA, Sthultz CM, Flores-Rentería L, Whipple AV & Whitham TG *Proc. Natl Acad. Sci. USA* 114, 11169–11174 (2017). [PubMed: 28973879]
4. Lee K, Pletcher SD, Lynch SV, Goldberg AN & Cope EK *Front. Cell. Infect. Microbiol* 8, 168 (2018). [PubMed: 29876323]
5. Metcalf JL et al. *Science* 351, 158–162 (2016). [PubMed: 26657285]
6. Rubin RL et al. *Ecol. Appl* 28, 1594–1605 (2018). [PubMed: 29989265]

7. Pineda A, Kaplan I & Bezemer TM *Trends Plant Sci.* 22, 770–778 (2017). [PubMed: 28757147]
8. Kapono CA et al. *Sci. Rep* 8, 3669 (2018). [PubMed: 29487294]
9. Verberkmoes NC et al. *ISME J.* 3, 179–189 (2009). [PubMed: 18971961]
10. Barr T et al. *Gut Microbes* 9, 338–356 (2018). [PubMed: 29517944]
11. Callahan BJ et al. *Nat. Methods* 13, 581–3 (2016). [PubMed: 27214047]
12. Amir A et al. *mSystems* 2, e00191–16 (2017). [PubMed: 28289731]
13. Bokulich NA et al. *Microbiome* 6, 90 (2018). [PubMed: 29773078]
14. Janssen S et al. *mSystems* 3, e00021–18 (2018). [PubMed: 29719869]
15. Bokulich NA et al. *mSystems* 3, e00219–18 (2018).
16. Bokulich N et al. *J. Open Source Softw* 3, 934 (2018).
17. Sedio BE, Rojas Echeverri JC, Boya PCA & Wright SJ *Ecology* 98, 616–623 (2017). [PubMed: 27984635]
18. Wang M et al. *Nat. Biotechnol* 34, 828–837 (2016). [PubMed: 27504778]
19. Hillmann B et al. *mSystems* 3, e00069–18 (2018). [PubMed: 30443602]
20. Truong DT et al. *Nat. Methods* 12, 902–903 (2015). [PubMed: 26418763]
21. Langille MGI et al. *Nat. Biotechnol* 31, 814–821 (2013). [PubMed: 23975157]
22. Kluyver T et al. Positioning and power in academic publishing: players, agents and agendas. in *Proc 20th International Conference on Electronic Publishing* (eds Loizides F & Schmidt B) 87–90 (IOS Press, 2016).
23. Amstutz P et al. 10.6084/m9.figshare.3115156.v2 (2016).
24. Gonzalez A et al. *Nat. Methods* 15, 796–798 (2018). [PubMed: 30275573]
25. Schloss PD et al. *Appl. Environ. Microbiol* 75, 7537–7541 (2009). [PubMed: 19801464]
26. McMurdie PJ & Holmes S *PLoS One* 8, e61217 (2013). [PubMed: 23630581]
27. Huber W et al. *Nat. Methods* 12, 115–121 (2015). [PubMed: 25633503]
28. Franzosa EA et al. *Nat. Methods* 15, 962–968 (2018). [PubMed: 30377376]
29. McDonald D et al. *Gigascience* 1, 7 (2012). [PubMed: 23587224]

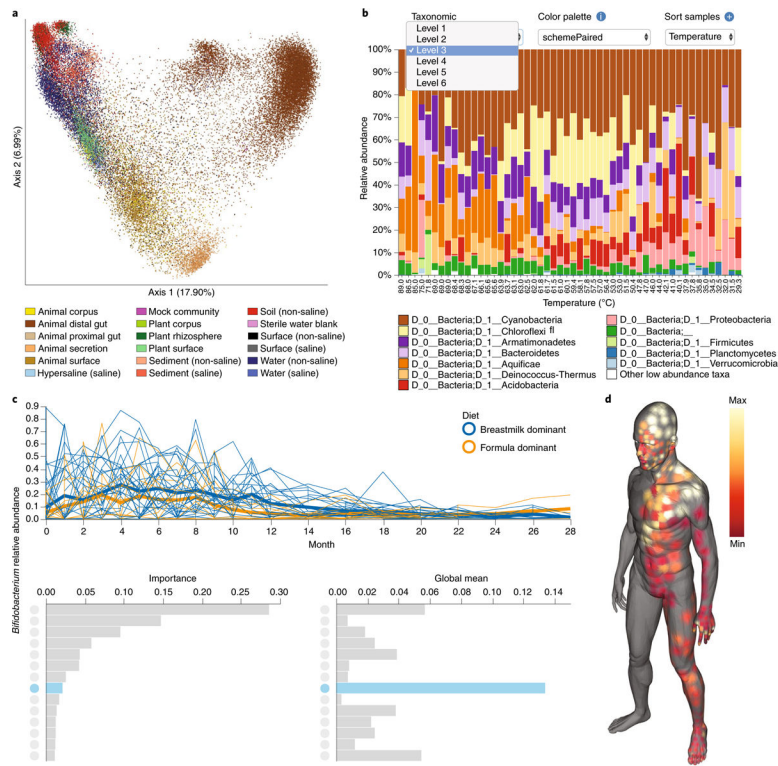


Fig. 1 | QIIME 2 provides many interactive visualization tools.

The products of four worked examples are presented here, and interactive versions of these screen captures are available in Supplementary File 1 and at <https://github.com/qiime2/paper1>. Detailed descriptions and methods, including the commands used to generate each of these visualizations, are provided in Supplementary Methods. **a**, Unweighted UniFrac principal coordinate analysis plot containing 37,680 samples, illustrating the scalability of QIIME 2. Colors indicate sample type, as described by the Earth Microbiome Project ontology (EMPO). **b**, Interactive taxonomic composition bar plot illustrating the phylum-level composition of microbial-mat samples collected along a temperature gradient in Yellowstone National Park Hot Spring outflow channels (Steep Cone Geyser). The many interactive controls available in this plot vastly decrease the burden of exploratory analysis over QIIME 1. **c**, Feature volatility plot (<https://msystems.asm.org/content/3/6/e00219-18>) illustrating the change in *Bifidobacterium* abundance over time in breast-fed and formula-fed infants. Temporally interesting features can be interactively discovered with this visualization. Bar charts rank the importance (predictive power for time point) and mean abundance of all microbial features. These bar charts provide an interface for visualizing volatility plots (line plots) of individual features in the context of their importance and abundance; clicking on a bar will display the volatility plot of that feature and highlight in blue that feature's importance and abundance in the bar charts below. **d**, Molecular cartography of the human skin surface. Colored spots represent the abundance of the small-molecule cosmetic ingredient sodium laureth sulfate on the human skin. Sample data can be interactively visualized in three-dimensional models, thus supporting the discovery of spatial patterns.

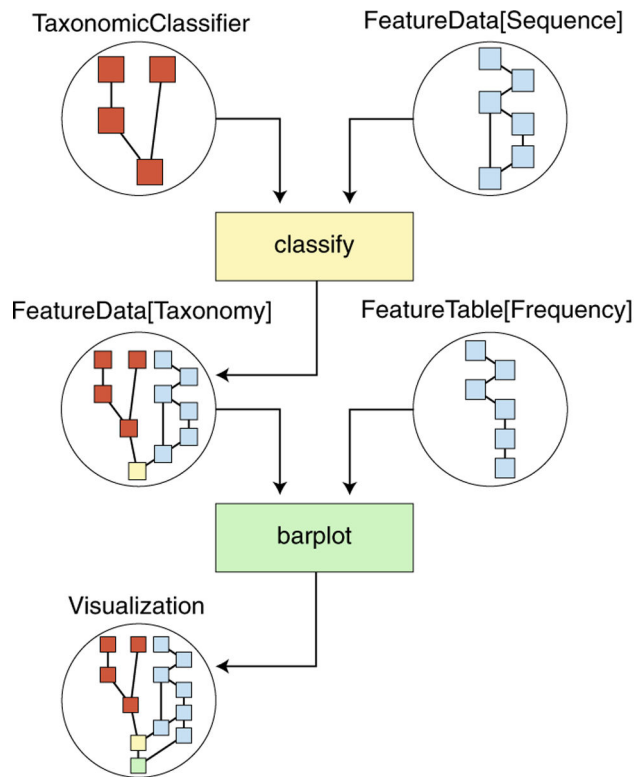


Fig. 2 |. QIIME 2 iteratively records data provenance, ensuring bioinformatics reproducibility. This simplified diagram illustrates the automatically tracked information regarding the creation of the taxonomy bar plot presented in Fig. 1b. QIIME 2 results (circles) contain network diagrams illustrating the data provenance stored in the result. Actions (quadrilaterals) are applied to QIIME 2 results and generate new results. Arrows indicate the flow of QIIME 2 results through actions. TaxonomicClassifier and FeatureData[Sequence] inputs contain independent provenance (red and blue, respectively) and are provided to a classify action (yellow), which taxonomically annotates sequences. The result of the classify action, a FeatureData[Taxonomy] result, integrates the provenance of both inputs with the classify action. This result is then provided to the barplot action with a FeatureTable[Frequency] input, which shares some provenance with the FeatureData[Sequence] input, because they were generated from the same upstream analysis. The resulting visualization (Fig. 1b) has the complete data provenance and correctly identifies shared processing of inputs. This simplified representation was created manually from the complete provenance graph for the purpose of illustration. An interactive and complete version of this provenance graph (as well as those for other Fig. 1 panels) can be accessed through Supplementary File 1.