

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

2
3 To the editor:

4
5 Rapid advances in DNA sequencing and bioinformatics technologies in the past two decades
6 have substantially improved our understanding of the microbial world. These include our
7 growing understanding of the vast diversity of microorganisms; how our microbiota and
8 microbiomes impact disease¹ and medical treatment²; how microorganisms impact the health of
9 our planet³; and our nascent exploration of the medical⁴, forensic⁵, environmental⁶, and
10 agricultural⁷ applications of microbiome biotechnology. Much of this work has been driven by
11 marker gene surveys (e.g., bacterial/archaeal 16S rRNA genes, fungal internal transcribed spacer
12 (ITS) regions, eukaryal 18S rRNA genes), which profile microbiota with varying degrees of
13 taxonomic specificity and phylogenetic information. The field is now transitioning to integrate
14 other data types, such as metabolite⁸, metaproteome⁹, or metatranscriptome^{9,10} profiles.

15 The QIIME 1 microbiome bioinformatics platform has supported many microbiome
16 studies and gained a broad user and developer community. Interactions with QIIME 1 users in
17 our online support forum, our workshops, and direct collaborations showed the potential to better
18 serve an increasingly diverse array of microbiome researchers in academia, government, and
19 industry. Here, we present QIIME 2, a completely reengineered and rewritten system that will
20 facilitate reproducible and modular analysis of microbiome data to enable the next generation of
21 microbiome science.

22 QIIME 2 is developed based on a plugin architecture (Supplementary Figure 1) that
23 allows third parties to contribute functionality (see <https://library.qiime2.org>). QIIME 2 plugins
24 exist for latest generation tools for sequence quality control from different sequencing platforms
25 (DADA2 (ref. 11) and Deblur¹²), taxonomy assignment¹³, and phylogenetic insertion¹⁴, that
26 quantitatively improve results over QIIME 1 and other tools (detailed in the corresponding tool-
27 specific publications). Plugins also support qualitatively new functionality, including
28 microbiome paired sample and time-series analysis¹⁵ (critical for studying the impact of
29 treatment on the microbiome) and machine learning¹⁶, with the ability to save trained models to
30 not only apply them to new data but also interrogate models to identify important microbiome
31 features. Several recently released plugins, including q2-cscs¹⁷, q2-metabolomics¹⁸, q2-shogun¹⁹,

32 q2-metaphlan2 (ref. 20), and q2-picrust2 (ref. 21), provide initial support for analysis of
33 metabolomics and shotgun metagenomics data. We are currently working with teams developing
34 bioinformatics tools for metatranscriptomics and metaproteomics, and expect to add new plugins
35 supporting these data types to the ecosystem shortly. Additionally, many of the existing
36 ‘downstream’ analysis tools, such as q2-sample-classifier¹⁶, can already work with these data
37 types individually or in combination if they are provided in a feature table. This marks the
38 potential of QIIME 2 to serve not only as a marker gene analysis tool, but also a
39 multidimensional and powerful data science platform that can be rapidly adapted to analyze
40 diverse microbiome features.

41 QIIME 2 provides many new interactive visualization tools facilitating exploratory
42 analyses and result reporting. Static versions of interactive visualizations resulting from four
43 worked examples are provided in Figure 1. QIIME 2 View (<https://view.qiime2.org>) is a unique
44 new service (see Supplementary Methods) that allows users to securely share and interact with
45 results without installing QIIME 2. The QIIME 2 visualizations presented in Figure 1 are
46 provided in Supplementary File 1 for readers to interact with using QIIME 2 View.

47 Corresponding worked QIIME 2 example code is provided in Supplementary Methods.

48 Reproducibility, transparency, and clarity of microbiome data science are guiding
49 principles in the QIIME 2 design. Toward this end, QIIME 2 includes a decentralized data
50 provenance tracking system: details of all analysis steps with references to intermediate data are
51 automatically stored in the results. Users can thus retrospectively determine exactly how any
52 result was generated (Figure 2 illustrates a simplified provenance graph derived from the data
53 provenance of Figure 1c). QIIME 2 also detects corrupted results, indicating that provenance is
54 no longer reliable and the results no longer contain information enabling reproducibility.

55 Provenance of the visualizations presented in Figure 1 can be interactively reviewed by loading
56 the contents of Supplementary File 1 with QIIME 2 View, providing far more detailed
57 information than can typically be provided in Methods text. QIIME 2 results are also
58 semantically typed (Fig. 2) and actions indicate acceptable input types, clarifying the data that
59 actions should be applied to and making complex workflows less error-prone. Complex
60 workflows can be created and shared using Jupyter Notebooks²² or Common Workflow
61 Language (CWL)²³, and support for other workflow engines is currently in development.

62 Finally, QIIME 2 provides a software development kit (see <https://dev.qiime2.org>) that
63 can be used to integrate it as a component of other systems (e.g., such as Qiita²⁴ or Illumina
64 BaseSpace) and to develop interfaces targeted toward users with different levels of
65 computational sophistication (Supplementary Figure 2). QIIME 2 provides the QIIME 2 Studio
66 graphical user interface and QIIME 2 View, interfaces designed for end-user biologists,
67 clinicians, and policymakers; the QIIME 2 application programming interface, designed for data
68 scientists who want to automate workflows or work interactively in Jupyter Notebooks²²; and
69 q2cli and q2cwl, providing a command line interface and CWL²³ wrappers for QIIME 2,
70 designed for high-performance computing experts. At present, computationally expensive steps
71 support parallel computing at the individual action level (for example, many actions including
72 de-noising and taxonomy assignment support multiple threads). We are currently developing
73 deeper integration with parallelism strategies available in third-party workflow engines, and
74 workflow-level parallelism is currently possible through CWL.

75 There are many other powerful open-source software tools for microbiome data science,
76 including mothur²⁵, phyloseq²⁶ and related tools available through Bioconductor²⁷, and the
77 biobakery suite^{20,21,28}. The microbiome bioinformatics platform mothur is often compared to
78 QIIME 1 and QIIME 2. A major difference between mothur and QIIME lies in the interactive
79 visualizations: QIIME 2 provides many interactive visualization tools (several examples are
80 provided in Figure 1), whereas mothur focuses on generating data that can be easily loaded and
81 visualized with other tools. The phyloseq tool focuses on microbiome statistical analysis and
82 generating publication-ready visualizations but, unlike QIIME 2, begins with a feature or OTU
83 table, leaving ‘upstream’ processing steps, such as sequence demultiplexing and quality control,
84 to other processing pipelines, many of which (like phyloseq) are available through Bioconductor.
85 The biobakery suite provides analytic functionality that complements that of QIIME 2, and we
86 are actively working with biobakery developers to support interoperability by making their tools
87 accessible as QIIME 2 plugins (for example, the q2-metaphlan2 plugin allows users to run
88 MetaPhlan2 through QIIME 2). QIIME 2 provides the only Python-based microbiome data-
89 science platform that supports retrospective data provenance tracking to ensure reproducibility,
90 multi-omics analysis support, interfaces geared toward different user types to enhance usability,
91 and an extensibility-focused design through the plugin architecture and software development

92 kit. We share feedback from users of QIIME 2 on these and other features in Supplementary
93 Methods.

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

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

109

110 **Code availability**

111 QIIME 2 is open source and free for all use, including commercial. It is licensed under the BSD
112 three-clause license. Source code is available at <https://github.com/qiime2> . To get help with
113 QIIME 2, visit <https://forum.qiime2.org> .

114

115 **Data availability**

116 Data for the analyses presented in Figure 1 are available as follows: Earth Microbiome Project
117 data in panel (a) was obtained from <ftp://ftp.microbio.me/emp/release1>, and the American Gut
118 Project (AGP) data was obtained from Qiita (<http://qiita.microbio.me>) study ID 10317. Sequence
119 data in panel (b) are available in Qiita under study ID 10249 and EBI under accession number
120 ERP016173. Sequence data in panel (c) are available in Qiita under study ID 925 and the
121 European Bioinformatics Institute (EBI) under accession number ERP022167. Data in panel (d)

122 are available in the q2-ili GitHub repository (<https://github.com/biocore/q2-ili>). Interactive
123 versions of the Figure 1 visualizations can be accessed at <https://github.com/qiime2/paper1> .

124

125

126 **Acknowledgements**

127 *QIIME 2 development was primarily funded by NSF Awards 1565100 to JGC and 1565057 to*
128 *RK. Partial support was also provided from the following grants: NIH U54CA143925 (JGC, TP)*
129 *and U54MD012388 (JGC, TP); grants from the Alfred P. Sloan Foundation (JGC, RK);*
130 *ERCSTG project MetaPG (NS); Strategic Priority Research Program of the Chinese Academy of*
131 *Sciences QYZDB-SSW-SMC021 (YB); from the Australian National Health and Medical*
132 *Research Council APP1085372 (GAH, JGC, Von Bing Yap and RK); and from Natural Sciences*
133 *and Engineering Research Council (NSERC) to DLG; and under the State of Arizona*
134 *Technology and Research Initiative Fund (TRIF), administered by the Arizona Board of Regents,*
135 *through Northern Arizona University. All NCI coauthors were supported by the Intramural*
136 *Research Program of the National Cancer Institute. SMG and ChD were supported by the*
137 *Washington Research Foundation Distinguished Investigator Award. Thanks to the Yellowstone*
138 *Center for Resources for research permit #5664 to JRS for Yellowstone access and sample*
139 *collection. We thank Paul J. McMurdie for helpful discussion on the relationships between*
140 *QIIME 2 and phyloseq. We would like to thank the users of QIIME 1 and 2, whose invaluable*
141 *feedback has shaped QIIME 2. In particular, we would like to thank Ahmed Abdelfattah*
142 *(Stockholm University, Sweden), Rozlyn C.T. Boutin (University of British Columbia, Canada),*
143 *David J. Bradshaw II (Florida Atlantic University Harbor Branch Oceanographic Institute,*
144 *USA), Lorinda Bullington (MPG Ranch, USA), Justine W. Debelius (Karolinska Institutet,*
145 *Sweden), Claire Duvall (Massachusetts Institute of Technology, USA), Erika Korzune Ganda*
146 *(Cornell University, USA), Alexander Mahnert (Medical University of Graz, Austria), Melanie C*
147 *Melendrez (St. Cloud State University, USA), Devon O'Rourke (University of New Hampshire,*
148 *USA), Adam R. Rivers (USDA ARS, USA), Biswarup Sen (Tianjin University, China), Solveig*
149 *Tangedal (Haukeland University Hospital and University of Bergen, Norway), Pedro J. Torres*
150 *(San Diego State University, USA), and Jonathan Warren (National Laboratory Service, UK) for*
151 *writing end user reviews included in Supplementary Methods.*

152

153 **Author contributions**

154 EB, JRR, MRD, NAB, YB, JEB, CJB, AMC, EC, ChD, RD, CFE, MEr, MEs, AG, JMG, DLG,
155 SMG, AKJ, KBK, STK, IK, TK, JL, YL, AVM, JLM, LFN, SBO, DP, AS, SJS, ADS, LRT,
156 PJTo, PJTu, SU, FV, JW, RK, and JGC developed documentation, educational materials, and/or
157 user/developer support content. EB, JRR, MRD, NAB, RK, and JGC wrote the manuscript; all
158 authors assisted with revision of the manuscript. EB, JRR, MRD, NAB, and JGC designed and
159 developed the QIIME 2 framework. DMD, AG, RL, EL, SCM, RS, JRS, WW, CHDW, and RK
160 contributed data used in the manuscript and/or testing of QIIME 2. CCA, CTB, EC, PCD, SH,
161 PK, EL, TP, RS, EV, YW, and RK contributed to the design of analytical methods. EB, JRR,
162 MRD, NAB, GAA, HA, EJA, MA, FA, KB, AB, BJC, JC, GMD, CD, MEr, JF, AG, KG, JG,
163 SMG, BH, HH, CH, GH, SJ, LJ, BK, CRK, DK, JK, MGIL, CL, MM, CM, BM, DM, LJM, JM,
164 ATN, JAN, SLP, MLP, EP, LBR, AR, MSR, PR, NS, MS, PT, AT, JJJV, YV, MV, MW, KCW,
165 ADW, ZZX, JRZ, YZ, QZ, and JGC contributed software to QIIME 2 plugins, interfaces,
166 framework, and/or build and test systems.

167

168 *Editor's Note: This paper has been peer-reviewed*

169

170 *Evan Bolyen^{1,*}, Jai Ram Rideout^{1,*}, Matthew R Dillon^{1,*}, Nicholas A Bokulich^{1,*}, Christian C*
171 *Abnet², Gabriel A Al-Ghalith³, Harriet Alexander^{4,5}, Eric J Alm^{6,7}, Manimozhayan Arumugam⁸,*
172 *Francesco Asnicar⁹, Yang Bai^{10,11,12}, Jordan E Bisanz¹³, Kyle Bittinger^{14,15}, Asker Brejnrod¹⁶,*
173 *Colin J Brislawn¹⁷, C Titus Brown⁵, Benjamin J Callahan^{18,19}, Andrés Mauricio Caraballo-*
174 *Rodríguez²⁰, John Chase¹, Emily K Cope^{1,21}, Ricardo Da Silva²⁰, Christian Diener²², Pieter C*
175 *Dorrestein²⁰, Gavin M Douglas²³, Daniel M Durall²⁴, Claire Duvall⁶, Christian F*
176 *Edwardson²⁵, Madeleine Ernst^{20,26}, Mehrbod Estaki²⁷, Jennifer Fouquier^{28,29}, Julia M Gauglitz²⁰,*
177 *Sean M Gibbons^{22,30}, Deanna L Gibson^{31,32}, Antonio Gonzalez³³, Kestrel Gorlick¹, Jiarong*
178 *Guo³⁴, Benjamin Hillmann³⁵, Susan Holmes³⁶, Hannes Holste^{33,37}, Curtis Huttenhower^{38,39},*
179 *Gavin A Huttley⁴⁰, Stefan Janssen⁴¹, Alan K Jarmusch²⁰, Lingjing Jiang⁴², Benjamin D*
180 *Kaehler^{43,40}, Kyo Bin Kang^{44,20}, Christopher R Keefe¹, Paul Keim¹, Scott T Kelley⁴⁵, Dan*
181 *Knights^{46,35}, Irina Koester^{47,20}, Tomasz Kosciolk⁴⁸, Jorden Kreps¹, Morgan GI Langille⁴⁹,*
182 *Joslynn Lee⁵⁰, Ruth Ley^{51,52}, Yong-Xin Liu^{10,11}, Erika Lofthfield², Catherine Lozupone²⁹, Massoud*
183 *Maher⁵³, Clarisse Marotz³³, Bryan D Martin⁵⁴, Daniel McDonald³³, Lauren J McIver^{38,39}, Alexey*

184 *V Melnik*²⁰, *Jessica L Metcalf*⁵⁵, *Sydney C Morgan*⁵⁶, *Jamie T Morton*^{33,53}, *Ahmad Turan*
185 *Naimey*¹, *Jose A Navas-Molina*^{53,33,57}, *Louis Felix Nothias*²⁰, *Stephanie B Orchanian*⁵⁸, *Talima*
186 *Pearson*¹, *Samuel L Peoples*^{59,60}, *Daniel Petras*²⁰, *Mary Lai Preuss*⁶¹, *Elmar Pruesse*²⁹, *Lasse*
187 *Buur Rasmussen*¹⁶, *Adam Rivers*⁶², *Michael S Robeson, II*⁶³, *Patrick Rosenthal*⁶¹, *Nicola Segata*⁹,
188 *Michael Shaffer*^{29,28}, *Arron Shiffer*¹, *Rashmi Sinha*², *Se Jin Song*³³, *John R Spear*⁶⁴, *Austin D*
189 *Swafford*⁵⁸, *Luke R Thompson*^{65,66}, *Pedro J Torres*⁶⁷, *Pauline Trinh*⁶⁸, *Anupriya Tripathi*^{20,33,69},
190 *Peter J Turnbaugh*⁷⁰, *Sabah Ul-Hasan*⁷¹, *Justin JJ van der Hooft*⁷², *Fernando Vargas*⁶⁹, *Yoshiki*
191 *Vázquez-Baeza*³³, *Emily Vogtmann*², *Max von Hippel*⁷³, *William Walters*⁵¹, *Yunhu Wan*²,
192 *Mingxun Wang*²⁰, *Jonathan Warren*⁷⁴, *Kyle C Weber*^{62,75}, *Charles HD Williamson*¹, *Amy D*
193 *Willis*⁷⁶, *Zhenjiang Zech Xu*³³, *Jesse R Zaneveld*⁷⁷, *Yilong Zhang*⁷⁸, *Qiyun Zhu*³³, *Rob*
194 *Knight*^{33,79,58}, *J Gregory Caporaso*^{1,21,+}

195

196 ¹*Center for Applied Microbiome Science, Pathogen and Microbiome Institute, Northern Arizona*
197 *University, Flagstaff, AZ, USA.*

198 ²*Metabolic Epidemiology Branch, National Cancer Institute, Rockville, MD, USA.*

199 ³*Department of Computer Science and Engineering, University of Minnesota, Minneapolis,*
200 *Minnesota, USA.*

201 ⁴*Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, USA.*

202 ⁵*Department of Population Health and Reproduction, University of California, Davis, CA, USA.*

203 ⁶*Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA,*
204 *USA.*

205 ⁷*Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology,*
206 *Cambridge, MA, USA.*

207 ⁸*University of Copenhagen, Faculty of Health and Medical Sciences, Novo Nordisk Foundation*
208 *Center for Basic Metabolic Research, Copenhagen, Denmark.*

209 ⁹*Centre for Integrative Biology, University of Trento, Trento, Italy.*

210 ¹⁰*State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology,*
211 *Chinese Academy of Sciences, Beijing, China.*

212 ¹¹*Centre of Excellence for Plant and Microbial Sciences (CEPAMS), Institute of Genetics and*
213 *Developmental Biology, Chinese Academy of Sciences & John Innes Centre, Beijing, China.*

214 ¹²*University of Chinese Academy of Sciences, Beijing, China.*

215 ¹³*Department of Microbiology and Immunology, University of California, San Francisco, CA,*
216 *USA.*

217 ¹⁴*Division of Gastroenterology and Nutrition, Children's Hospital of Philadelphia, Philadelphia,*
218 *PA, USA.*

219 ¹⁵*Hepatology, Children's Hospital of Philadelphia, Philadelphia, PA, USA.*

220 ¹⁶*Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and*
221 *Medical Sciences, University of Copenhagen, Denmark.*

222 ¹⁷*Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland,*
223 *WA, USA.*

224 ¹⁸*Department of Population Health & Pathobiology, North Carolina State University, Raleigh,*
225 *NC, USA.*

226 ¹⁹*Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA.*

227 ²⁰*Collaborative mass spectrometry innovation center, Skaggs School of Pharmacy and*
228 *Pharmaceutical Sciences, University of California San Diego, San Diego, CA, USA.*

229 ²¹*Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, USA.*

230 ²²*Institute for Systems Biology, Seattle, WA, USA.*

231 ²³*Department of Microbiology and Immunology, Dalhousie University, Halifax, Nova Scotia,*
232 *Canada.*

233 ²⁴*Irving K. Barber School of Arts and Sciences, University of British Columbia, Kelowna, British*
234 *Columbia, Canada.*

235 ²⁵*A. Watson Armour III Center for Animal Health and Welfare, Aquarium Microbiome Project,*
236 *John G. Shedd Aquarium, Chicago, IL, USA.*

237 ²⁶*Department of Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark.*

238 ²⁷*Department of Biology, University of British Columbia Okanagan, Okanagan, BC, Canada.*

239 ²⁸*Computational Bioscience Program, University of Colorado Anschutz Medical Campus,*
240 *Aurora, CO, USA.*

241 ²⁹*Department of Medicine, Division of Biomedical Informatics and Personalized Medicine,*
242 *University of Colorado Anschutz Medical Campus, Aurora, CO, USA.*

243 ³⁰*eScience Institute, University of Washington, Seattle, WA, USA.*

244 ³¹*Irving K. Barber School of Arts and Sciences, Department of Biology, The University of British*
245 *Columbia, Kelowna, BC, Canada.*

246 ³²*Department of Medicine, The University of British Columbia, Kelowna, BC, Canada.*
247 ³³*Department of Pediatrics, University of California San Diego, La Jolla, CA, USA.*
248 ³⁴*Center for Microbial Ecology, Michigan State University, East Lansing, MI, USA.*
249 ³⁵*Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN,*
250 *USA.*
251 ³⁶*Stanford University, Statistics Department, Palo Alto, CA, USA.*
252 ³⁷*Department of Computer Science and Engineering, University of California San Diego, La*
253 *Jolla, CA, USA.*
254 ³⁸*Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.*
255 ³⁹*Broad Institute of MIT and Harvard, Cambridge, MA, USA.*
256 ⁴⁰*Research School of Biology, The Australian National University, Canberra, ACT, Australia.*
257 ⁴¹*Department of Pediatric Oncology, Hematology and Clinical Immunology, Heinrich-Heine*
258 *University Dusseldorf, Dusseldorf, Germany.*
259 ⁴²*Department of Family Medicine and Public Health, University of California San Diego, La*
260 *Jolla, CA, USA.*
261 ⁴³*School of Science, University of New South Wales, Canberra, ACT, Australia.*
262 ⁴⁴*College of Pharmacy, Sookmyung Women's University, Seoul, Republic of Korea.*
263 ⁴⁵*San Diego State University, Department of Biology, San Diego, CA, USA.*
264 ⁴⁶*Biotechnology Institute, University of Minnesota, Saint Paul, MN, USA.*
265 ⁴⁷*Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA.*
266 ⁴⁸*Department of Pediatrics, University of California San Diego, La Jolla, CA,.*
267 ⁴⁹*Department of Pharmacology, Dalhousie University, Halifax, Nova Scotia, Canada.*
268 ⁵⁰*Science Education, Howard Hughes Medical Institute, Ashburn, VA, USA.*
269 ⁵¹*Department of Microbiome Science, Max Planck Institute for Developmental Biology,*
270 *Tübingen, Germany.*
271 ⁵²*Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA.*
272 ⁵³*Department of Computer Science & Engineering, University of California San Diego, La Jolla,*
273 *CA, USA.*
274 ⁵⁴*Department of Statistics, University of Washington, Seattle, WA, USA.*
275 ⁵⁵*Department of Animal Science, Colorado State University, Fort Collins, CO, USA.*

276 ⁵⁶*Irving K. Barber School of Arts and Sciences, Unit 2 (Biology), University of British Columbia,*
277 *Kelowna, BC, Canada.*

278 ⁵⁷*Mountain View, Google LLC, Mountain View, CA, USA.*

279 ⁵⁸*Center for Microbiome Innovation, University of California San Diego, La Jolla, CA, USA.*

280 ⁵⁹*School of Information Studies, Syracuse University, Syracuse, NY, USA.*

281 ⁶⁰*School of STEM, University of Washington Bothell, Bothell, WA, USA.*

282 ⁶¹*Department of Biological Sciences, Webster University, St Louis, MO, USA.*

283 ⁶²*Agricultural Research Service, Genomics and Bioinformatics Research Unit, United States*
284 *Department of Agriculture, Gainesville, FL, USA.*

285 ⁶³*College of Medicine, Department of Biomedical Informatics, University of Arkansas for*
286 *Medical Sciences, Little Rock, AR, USA.*

287 ⁶⁴*Department of Civil and Environmental Engineering, Colorado School of Mines, Golden, CO,*
288 *USA.*

289 ⁶⁵*Department of Biological Sciences and Northern Gulf Institute, University of Southern*
290 *Mississippi, Hattiesburg, Mississippi, USA.*

291 ⁶⁶*Ocean Chemistry and Ecosystems Division, Atlantic Oceanographic and Meteorological*
292 *Laboratory, National Oceanic and Atmospheric Administration, La Jolla, CA, USA.*

293 ⁶⁷*Department of Biology, San Diego State University, San Diego, CA, USA.*

294 ⁶⁸*Department of Environmental and Occupational Health Sciences, University of Washington,*
295 *Seattle, WA, USA.*

296 ⁶⁹*Division of Biological Sciences, University of California San Diego, San Diego, CA, USA.*

297 ⁷⁰*Department of Microbiology and Immunology, University of California San Francisco, San*
298 *Francisco, CA, USA.*

299 ⁷¹*Quantitative and Systems Biology Graduate Program, University of California Merced,*
300 *Merced, CA, USA.*

301 ⁷²*Bioinformatics Group, Wageningen University, Wageningen, The Netherlands.*

302 ⁷³*Department of Mathematics, University of Arizona, Tucson, AZ, USA.*

303 ⁷⁴*National Laboratory Service, Environment Agency, Starcross, UK.*

304 ⁷⁵*College of Agriculture and Life Sciences, University of Florida, Gainesville, FL, USA.*

305 ⁷⁶*Department of Biostatistics, University of Washington, Seattle, WA, USA.*

306 ⁷⁷*University of Washington Bothell, School of STEM, Division of Biological Sciences, Bothell,*
307 *WA, USA.*

308 ⁷⁸*Merck & Co. Inc., Kenilworth, NJ, USA.*

309 ⁷⁹*Department of Computer Science and Engineering, University of California San Diego, La*
310 *Jolla, California, USA.*

311

312 * *These authors contributed equally to this work.*

313 ⁺ *Correspondence should be addressed to JGC (e-mail: greg.caporaso@nau.edu)*

314

315 1. Smith, M.I. et al. *Science* **339** , 548–554 (2013).

316 2. Gopalakrishnan, V. et al. *Science* **359** , 97–103 (2018).

317 3. Gehring, C.A., Sthultz, C.M., FloresRentería,
318 L., Whipple, A.V. & Whitham, T.G. *Proc. Natl. Acad. Sci. U.*
319 *S. A.* **114** , 11169–11174 (2017).

320 4. Lee, K., Pletcher, S.D., Lynch, S.V., Goldberg, A.N. & Cope, E.K. *Front. Cell. Infect.*
321 *Microbiol.* **8** , 168
322 (2018).

323 5. Metcalf, J.L. et al. *Science* **351** , 158–162 (2016).

324 6. Rubin, R.L. et al. *Ecol. Appl.* **28** , 1594–1605 (2018).

325 7. Pineda, A., Kaplan, I. & Bezemer, T.M. *Trends Plant Sci.* **22** , 770–778 (2017).

326 8. Kaponó, C.A. et al. *Sci. Rep.* **8** , 3669 (2018).

327 9. Verberkmoes, N.C., et al. [ISME J.](https://doi.org/10.1038/ismej.2008.108) **3**, 179–189 (2009). doi: 10.1038/ismej.2008.108

328 10. Barr, T. et al. *Gut Microbes* 1–44 (2018).

329 11. Callahan, B.J. et al. *Nat. Methods* (2016).doi: 10.1038/nmeth.3869

330 12. Amir, A. et al. *mSystems* **2** , (2017).

331 13. Bokulich, N.A. et al. *Microbiome* **6** , 90 (2018).

332 14. Janssen, S. et al. *mSystems* **3** , e00021–18 (2018).

333 15. Bokulich, N.A. et al. *mSystems* **3** , e00219–18 (2018).

334 16. Bokulich, N. et al. *JOSS* **3** , 934 (2018).

335 17. Sedio, B.E., Rojas Echeverri, J.C., Boya P, C.A. & Wright, S.J. *Ecology* **98** , 616–623
336 (2017).

- 337 18. Wang, M. et al. *Nat. Biotechnol.* **34** , 828–837 (2016).
- 338 19. Hillmann, B. et al. *bioRxiv* 320986 (2018).doi: 10.1101/320986
- 339 20. Truong, D.T. et al. *Nat. Methods* **12** , 902–903 (2015).
- 340 21. Langille, M.G.I. et al. *Nat. Biotechnol.* **31** , 814–821 (2013).
- 341 21. Gonzalez, A. et al. *Nat. Methods* **15** , 796–798 (2018).
- 342 22. Amstutz, P. et al. (2016).doi: 10.6084/m9.figshare.3115156.v2
- 343 23. Schloss, P.D. et al. *Appl. Environ. Microbiol.* **75** , 7537–7541 (2009).
- 344 24. McMurdie, P.J. & Holmes, S. *PLoS One* **8** , e61217 (2013).
- 345 25. Huber, W. et al. *Nat. Methods* **12** , 115–121 (2015).
- 346 26. Franzosa, E.A. et al. *Nat. Methods* **15** , 962–968 (2018).
- 347 27. Kluyver, T. et al. Positioning and Power in Academic Publishing: Players, Agents and
- 348 Agendas. *Proceedings of the 20th International Conference on Electronic Publishing* (Loizides,
- 349 F., Schmidt, B, eds) 87–90 (IOS Press, Amsterdam, Netherland, 2016). ISBN print 978-1-
- 350 61499-648-4.
- 351 28. McDonald, D. et al. *Gigascience* **1** , 7 (2012).
- 352

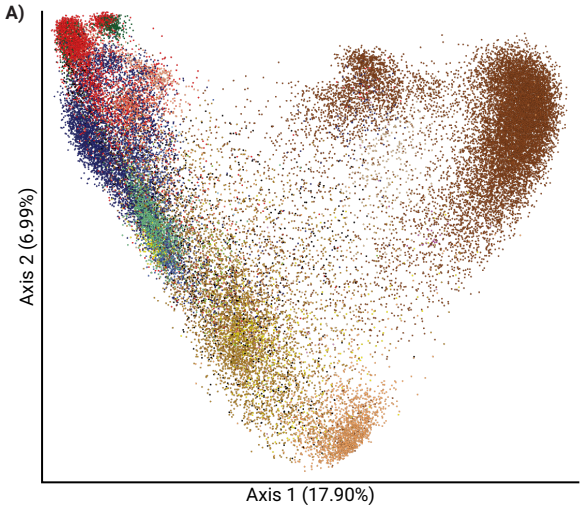
353 **Figure 1:** QIIME 2 provides many interactive visualization tools. The products of four worked
354 examples are presented here, and interactive versions of these screen captures are available in
355 Supplementary File 1 and at <https://github.com/qiime2/paper1> . Detailed descriptions and
356 methods, including the commands used to generate each of these visualizations, are provided in
357 Supplementary Methods. **(a)** Unweighted UniFrac PCoA plot containing 37,680 samples,
358 illustrating the scalability of QIIME 2. Colors indicate sample type, as described by the Earth
359 Microbiome Project ontology (EMPO). **(b)** A feature volatility plot
360 (<https://msystems.asm.org/content/3/6/e00219-18>) illustrating change in *Bifidobacterium*
361 abundance over time in breast-fed and formula-fed infants. Temporally interesting features can
362 be interactively discovered with this visualization. Bar charts rank the importance (predictive
363 power for time point) and mean abundance of all microbial features. These bar charts provide an
364 interface for visualizing volatility plots (line plots) of individual features in the context of their
365 importance and abundance; clicking on a bar will display the volatility plot of that feature and
366 highlight in blue that feature's importance and abundance in the bar charts below. **(c)** Interactive
367 taxonomic composition bar plot illustrating phylum-level composition of microbial mat samples
368 collected along a temperature gradient in Yellowstone National Park Hot Spring outflow
369 channels (Steep Cone Geyser). The many interactive controls available in this plot vastly reduce
370 the burden of exploratory analysis over QIIME 1. **(d)** Molecular cartography of the human skin
371 surface. Colored spots represent the abundance of the small molecule cosmetic, sodium laureth
372 sulfate, on the human skin. Sample data can be interactively visualized on three-dimensional
373 models, supporting the discovery of spatial patterns.
374

375 **Figure 2:** QIIME 2 iteratively records data provenance, ensuring bioinformatics reproducibility.
 376 This simplified diagram illustrates the automatically tracked information about the creation of
 377 the taxonomy barplot presented in Figure 1c. QIIME 2 results (circles) contain network diagrams
 378 illustrating the data provenance stored in the result. Actions (quadrilaterals) are applied to
 379 QIIME 2 results and generate new results. Arrows indicate flow of QIIME 2 results through
 380 actions. TaxonomicClassifier and FeatureData[Sequence] inputs contain independent provenance
 381 (red and blue, respectively) and are provided to a classify action (yellow), which taxonomically
 382 annotates sequences. The result of the classify action, a FeatureData[Taxonomy] result,
 383 integrates the provenance of both inputs with the classify action. This result is then provided to
 384 the barplot action with a FeatureTable[Frequency] input, which shares some provenance with the
 385 FeatureData[Sequence] input as they were generated from the same upstream analysis. The
 386 resulting Visualization (Figure 1c), has the complete data provenance and correctly identifies
 387 shared processing of inputs. This simplified representation was created manually from the
 388 complete provenance graph for the purpose of illustration. An interactive and complete version
 389 of this provenance graph (as well as those for other Figures 1 panels) can be accessed through
 390 Supplementary File 1.

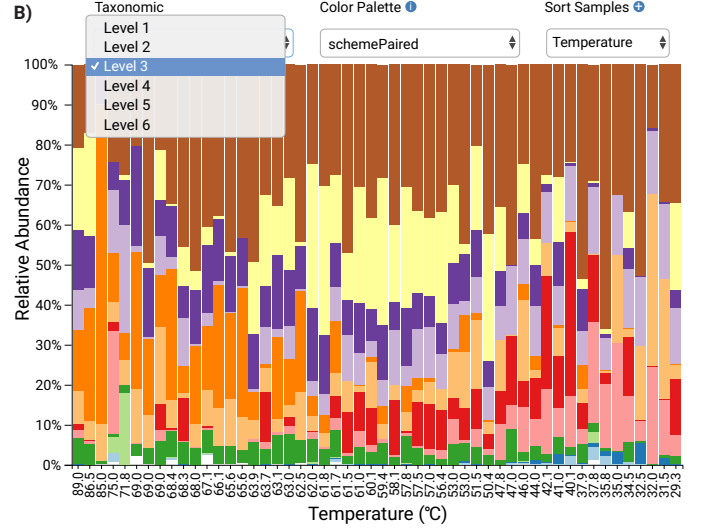
391

<p>ESM (reporting summaries should always come second, unless there is no other ESM file)</p>	<p>File type: PDF Title: Supplementary Information Description: Supplementary Figures 1–3 and Supplementary Methods</p> <p>File type: ZIP Title: Supplementary File 1 Description: Interactive versions of the visualizations presented in Figure 1. These can be viewed using QIIME 2, for example at https://view.qiime2.org.</p>
---	---

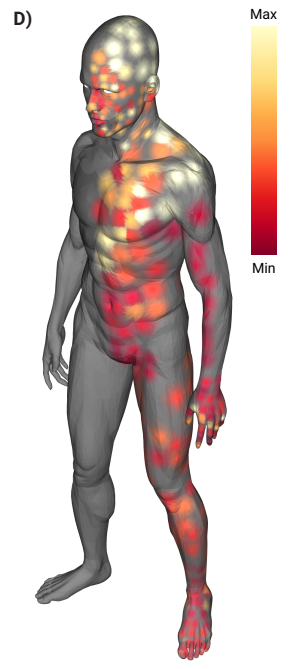
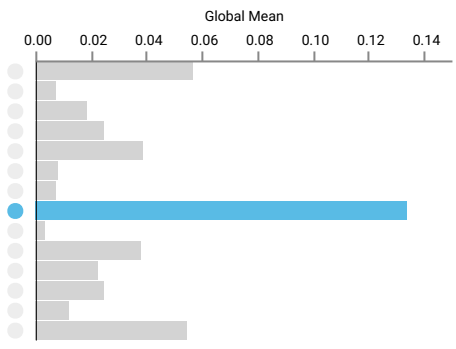
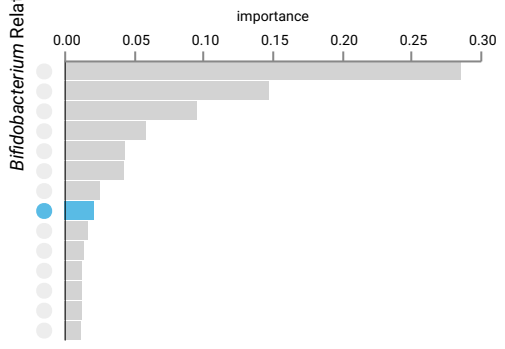
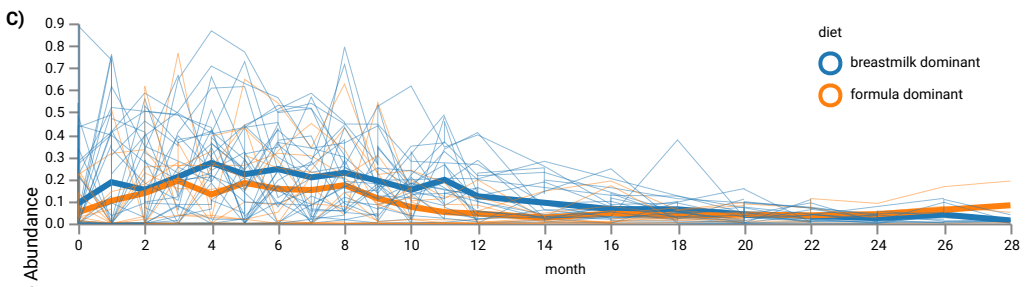
392



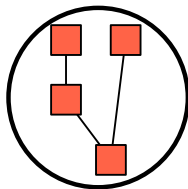
- | | | |
|----------------------|-----------------------|----------------------|
| Animal corpus | Mock community | Soil (non-saline) |
| Animal distal gut | Plant corpus | Sterile water blank |
| Animal proximal gut | Plant rhizosphere | Surface (non-saline) |
| Animal secretion | Plant surface | Surface (saline) |
| Animal surface | Sediment (non-saline) | Water (non-saline) |
| Hypersaline (saline) | Sediment (saline) | Water (saline) |



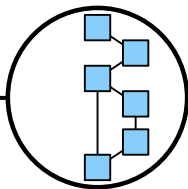
- | | |
|--|------------------------------------|
| D_0__Bacteria;D_1__Cyanobacteria | D_0__Bacteria;D_1__Proteobacteria |
| D_0__Bacteria;D_1__Chloroflexi | D_0__Bacteria;D_1__Firmicutes |
| D_0__Bacteria;D_1__Armatimonadetes | D_0__Bacteria;D_1__Firmicutes |
| D_0__Bacteria;D_1__Bacteroidetes | D_0__Bacteria;D_1__Planctomycetes |
| D_0__Bacteria;D_1__Aquificae | D_0__Bacteria;D_1__Verrucomicrobia |
| D_0__Bacteria;D_1__Deinococcus-Thermus | Other low abundance taxa |
| D_0__Bacteria;D_1__Acidobacteria | |



TaxonomicClassifier

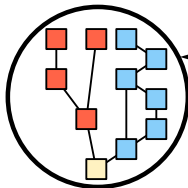


FeatureData[Sequence]

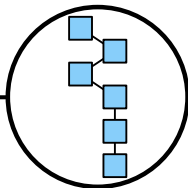


classify

FeatureData[Taxonomy]



FeatureTable[Frequency]



barplot

Visualization

