



# Earth Microbiome Project and Global Systems Biology

Jack A. Gilbert,<sup>a,b</sup> Founding Editor in Chief, *mSystems*, Janet K. Jansson,<sup>c</sup> Senior Editor, *mSystems*, Rob Knight,<sup>d,e,f</sup> Senior Editor, *mSystems*

<sup>a</sup>Biosciences Division, Argonne National Laboratory, Argonne, Illinois, USA

<sup>b</sup>Microbiome Center, Department of Surgery, University of Chicago, Chicago, Illinois, USA

<sup>c</sup>Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington, USA

<sup>d</sup>Department of Pediatrics, University of California San Diego, La Jolla, California, USA

<sup>e</sup>Department of Computer Science & Engineering, University of California San Diego, La Jolla, California, USA

<sup>f</sup>Center for Microbiome Innovation, University of California San Diego, La Jolla, California, USA

**KEYWORDS** amplicon, global, meta-analysis, metadata, microbiome, systems biology

Recently, we published the first large-scale analysis of data from the Earth Microbiome Project (EMP) (1, 2), a truly multidisciplinary research program involving more than 500 scientists and 27,751 samples acquired from 43 countries. These samples represent myriad specimen types and span a wide range of biotic and abiotic factors, geographic locations, and physicochemical properties. The database (<https://qiita.ucsd.edu/emp/>) is still growing, with more than 100,000 amplicon data sets and more than 500 paired metagenomic and metabolomic data sets. Importantly, the techniques, data, and analytical tools are all standardized and publicly accessible, providing a framework to support research at a scale of integration that just 7 years ago seemed impossible.

The project started in 2010, as all good things start, with friends and colleagues meeting to discuss ideas. The Terabase Metagenomics meeting was the brainchild of Rick Stevens from Argonne National Laboratory (3) and provided a platform for brainstorming about approaches that would be necessary to create a database of the microbially diverse populations of the planet. Our goal was that the database would be publically accessible, driven by community engagement, and be able to test fundamental hypotheses around microbial ecology. The community responded. Researchers from around the world were willing to share their samples and their associated metadata (sample and site information), so that the EMP could process material and data through a standardized pipeline (Fig. 1).

Standardization allowed disparate samples to be compared with respect to their microbial compositions and diversity. Before the EMP, this was simply impossible, because different investigators used different DNA extraction methods, sequencing protocols, and analytical tools that had been established over years and optimized for their specific sample types. For example, hundreds of different protocols have been developed for extraction of DNA from soil, each with their own biases (4). This presented a major impediment for cross-sample, cross-site, and cross-study comparisons. The diversity of sample types included in the EMP ranged from soil samples, to human and animal stools and skin samples, to marine and freshwater samples. The EMP even went to extremes to try to collect samples not commonly represented in sequence databases to extend our knowledge of the associated microbiota: examples include bird eggshells, ants, bats, polluted marine sediments, and soil samples from deserts and Antarctica and North American permafrost. The three main labs that established the EMP overcame this hurdle by establishing common protocols for DNA extraction, used the same primers and sequencing protocols—and even performed the

**Published** 10 April 2018

**Citation** Gilbert JA, Jansson JK, Knight R. 2018. Earth microbiome project and global systems biology. *mSystems* 3:e00217-17. <https://doi.org/10.1128/mSystems.00217-17>.

**Copyright** © 2018 Gilbert et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Jack A. Gilbert, [gilbertjack@gmail.com](mailto:gilbertjack@gmail.com).

*The views expressed in this Editorial do not necessarily reflect the views of this journal or of ASM.*



**FIG 1** The Earth Microbiome Project was a result of collaborations with hundreds of scientists around the planet; each with their own research interests.

DNA extractions and sequencing in defined facilities (to overcome known bias between sequencing facilities). In addition, we established QIIME as a common analytical framework and Qiita as a unified database for data and metadata storage. We also established a common metadata platform and required all EMP investigators to provide their metadata before the samples could be sent for DNA extraction and sequencing, greatly facilitating analysis.

The EMP has from the start been challenged to justify this approach and to justify the massive effort required to create this database. The primary justification for this effort is advancing scientific knowledge. We considered but rejected the concept of sampling the earth on a grid, sampling randomly, or sampling with a stratified sampling design, on the grounds that an up-front investment of billions of dollars would be required to do it correctly (although such an investment would be immensely valuable and comparable to existing efforts in earth observation such as satellites). Instead, hypothesis-driven research was an essential part of the EMP from day 1. The EMP is essentially a resource that provides a framework for comparative analysis, so that individual hypothesis-driven studies can be combined to test hypotheses that require bigger data sets or data from a wider range of environments or need to be replicated in multiple systems to establish generality. Therefore, the EMP is a meta-analysis resource that leverages individual hypothesis-driven experiments to ask new questions. As a result, >50 published studies have already generated data through the EMP (<http://www.earthmicrobiome.org/publications/>). These publications represent individual studies from myriad ecosystems, including the microbiota associated with the digestive system of carnivorous pitcher plants (5), time series studies of the dynamics of freshwater lake microbiota (6), the microbiota of saliva in mice who drink wine (lean and obese men) (7), and the microbiota of permafrost soils from boreal forests in Alaska (8). Many large consortia also worked with the EMP to derive data that could aid in the development and testing of many novel hypotheses, such as the phylogenetic distribution of microbiota across sponges around the world (9). These studies underpin the meta-analysis that was recently published (2).

So what did this meta-analysis uncover? First and foremost, the volume and environmental distribution of these data enabled us to test fundamental hypotheses in biogeography, including determining patterns that have previously been possible only for “macrobial” ecology. In addition, the ecological trends demonstrated key organizing principles whereby ecosystems with less diversity maintained taxa that were found in samples with greater diversity. This nested ecology extended our understanding of microbial seed banks that support global distribution networks (10, 11). Importantly, the data generated have also been used to explore the factors that underpin global diversity trends (12) and by using informatics techniques that highlight the local adaptation and therefore environmental specificity of subspecies.

So where do we go from here? Although the EMP has provided a tremendous resource to the scientific community, it is focused on the taxonomic representation of the community members. What is lacking is understanding of the functional roles that they carry out. Therefore, the next step will be to use a combination of metagenome sequencing and complementary omics technologies (e.g., metatranscriptomics, metaproteomics, and metabolomics) to determine the functional complement of genes, which genes are expressed and translated into proteins and what metabolic processes are carried out. A multi-omics approach has already been used to evaluate microbiomes in targeted habitats. Examples from our own research include permafrost before and after thaw (13), the deep-sea oil plume (14) and contaminated sediments (15) resulting from the Deepwater Horizon oil spill in the Gulf of Mexico, and the human gut following a resistant starch diet (16). An EMP-type approach is now being planned to standardize multi-omics across sample types in order to glean functional comparisons of microbiomes across the planet. To begin with, 500 EMP samples that cover a variety of diverse sample types have been selected for metagenomics and metabolomics. This next step on the EMP horizon will be to provide the foundation for gaining a mechanistic understanding of the roles carried out by earth’s microbiomes and will allow us to test the limits of our ability to extrapolate from taxonomy to function, as well as to integrate observations into models of microbiome change at different levels.

Additionally, with the amplicon protocols well established, we invite the scientific community to generate their own data using the EMP protocols and contribute the data to the Qiita database at <https://qiita.ucsd.edu/>. We will provide a streamlined user interface for incorporating these new data sets into the EMP, capturing standardized as

well as study-specific metadata, checking them against the wealth of published studies for quality control and potential issues of sampling handling, and adding them to the global catalog of microbial knowledge. This rapidly expanding community resource will be of immense value in understanding the global distribution of microbial life and uncovering new ecological principles as well as confirming the generality of existing ones across spatial and temporal scales.

This resource is therefore of direct relevance to the microbial systems biology community and all researchers involved in the exploration of relationships between the components of a microbial cell or community. A system is a set of connected components forming a complex whole. A systems component, in a microbiological context, can comprise any biological or even physicochemical element that connects to produce the observed characteristics of the system. By defining the coassociation and distribution of microbial taxonomic components across planet earth, we provide a framework that contextualizes the ecology of those elements, which can in turn aid interpretation of their properties in other systems. Demonstrating that a particular taxonomic unit of relevance to a plant-root system is broadly distributed across defined physicochemical gradients provides evidence for the adaptation of that lineage, which can aid interpretation of its ecological dynamics. We hope this resource will benefit the community.

## REFERENCES

- Gilbert JA, Jansson JK, Knight R. 2014. The Earth Microbiome Project: successes and aspirations. *BMC Biol* 12:69. <https://doi.org/10.1186/s12915-014-0069-1>.
- Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vázquez-Baeza Y, González A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Kanbar J, Zhu Q, Jin Song S, Kosciulek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R, Earth Microbiome Project Consortium. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551:457–463. <https://doi.org/10.1038/nature24621>.
- Gilbert JA, Meyer F, Antonopoulos D, Balaji P, Brown CT, Brown CT, Desai N, Eisen JA, Evers D, Field D, Feng W, Huson D, Jansson J, Knight R, Knight J, Kolker E, Konstantidis K, Kostka J, Kyrpides N, Mackelprang R, McHardy A, Quince C, Raes J, Sczyrba A, Shade A, Stevens R. 2010. Meeting report: the Terabase Metagenomics Workshop and the vision of an Earth Microbiome Project. *Stand Genomic Sci* 3:243–248. <https://doi.org/10.4056/signs.1433550>.
- Delmont TO, Prestat E, Keegan KP, Faubladier M, Robe P, Clark IM, Pelletier E, Hirsch PR, Meyer F, Gilbert JA, Le Paslier D, Simonet P, Vogel TM. 2012. Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J* 6:1677–1687. <https://doi.org/10.1038/ismej.2011.197>.
- Armitage DW. 2017. Linking the development and functioning of a carnivorous pitcher plant's microbial digestive community. *ISME J* 11: 2439–2451. <https://doi.org/10.1038/ismej.2017.99>.
- Kara EL, Hanson PC, Hu YH, Winslow L, McMahon KD. 2013. A decade of seasonal dynamics and co-occurrences within freshwater bacterioplankton communities from eutrophic Lake Mendota, WI, USA *ISME J* 7:680–684. <https://doi.org/10.1038/ismej.2012.118>.
- Piombino P, Genovese A, Esposito S, Moio L, Cutolo PP, Chambery A, Severino V, Moneta E, Smith DP, Owens SM, Gilbert JA, Ercolini D. 2014. Saliva from obese individuals suppresses the release of aroma compounds from wine. *PLoS One* 9:e85611. <https://doi.org/10.1371/journal.pone.0085611>.
- Taş N, Prestat E, McFarland JW, Wickland KP, Knight R, Berhe AA, Jorgenson T, Waldrop MP, Jansson JK. 2014. Impact of fire on active layer and permafrost microbial communities and metagenomes in an upland Alaskan boreal forest. *ISME J* 8:1904–1919. <https://doi.org/10.1038/ismej.2014.36>.
- Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, Olson JB, Erwin PM, López-Legentil S, Luter H, Chaves-Fonnegra A, Costa R, Schupp PJ, Steindler L, Erpenbeck D, Gilbert J, Knight R, Ackermann G, Victor Lopez J, Taylor MW, Thacker RW, Montoya JM, Hentschel U, Webster NS. 2016. Diversity, structure and convergent evolution of the global sponge microbiome. *Nat Commun* 7:11870. <https://doi.org/10.1038/ncomms11870>.
- Lennon JT, Jones SE. 2011. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat Rev Microbiol* 9:119–130. <https://doi.org/10.1038/nrmicro2504>.
- Gibbons SM, Caporaso JG, Pirrung M, Field D, Knight R, Gilbert JA. 2013. Evidence for a persistent microbial seed bank throughout the global ocean. *Proc Natl Acad Sci U S A* 110:4651–4655. <https://doi.org/10.1073/pnas.1217767110>.
- Locey KJ, Lennon JT. 2016. Scaling laws predict global microbial diversity. *Proc Natl Acad Sci U S A* 113:5970–5975. <https://doi.org/10.1073/pnas.1521291113>.
- Hultman J, Waldrop MP, Mackelprang R, David MM, McFarland J, Blaze-wicz SJ, Harden J, Turetsky MR, McGuire AD, Shah MB, VerBerkmoes NC, Lee LH, Mavrommatis K, Jansson JK. 2015. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* 521:208–212. <https://doi.org/10.1038/nature14238>.
- Mason OU, Hazen TC, Borglin S, Chain PSG, Dubinsky EA, Fortney JL, Han J, Holman H-YN, Hultman J, Lamendella R, Mackelprang R, Malfatti S, Tom LM, Tringe SG, Woyke T, Zhou J, Rubin EM, Jansson JK. 2012. Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J* 6:1715–1727. <https://doi.org/10.1038/ismej.2012.59>.
- Handley KM, Piceno YM, Hu P, Tom LM, Mason OU, Andersen GL, Jansson JK, Gilbert JA. 2017. Metabolic and spatio-taxonomic response of uncultivated seafloor bacteria following the Deepwater Horizon oil spill. *ISME J* 11:2569–2583. <https://doi.org/10.1038/ismej.2017.110>.
- Maier TV, Lucio M, Lee LH, VerBerkmoes NC, Brislawn CJ, Bernhardt J, Lamendella R, McDermott JE, Bergeron N, Heinzmann SS, Morton JT, González A, Ackermann G, Knight R, Riedel K, Krauss RM, Schmitt-Kopplin P, Jansson JK. 2017. Impact of dietary resistant starch on the human gut microbiome, metaproteome, and metabolome. *mBio* 8:e01343-17. <https://doi.org/10.1128/mBio.01343-17>.