

Microbiome Research Strategy: Infrastructure, Computation, and Discovery

A Policy Paper by the Engineering Biology Research Consortium

Compiled and edited by Wilson Sinclair, EBRC Postdoctoral Scholar, and Becky Mackelprang, EBRC Associate Director for Security Programs

July 2023

Microbiome science is vast and complex, spanning multiple scientific disciplines with applications in several sectors essential to the current and future U.S. bioeconomy.¹ Microbiomes – communities or consortia of diverse microbes – are ubiquitous in nature and exist in, on, and around living hosts (including plants, animals, and humans) as well as soil, water, and built materials.² Recent advances in engineering biology and data science have substantially accelerated discoveries in microbiome research. Scientists’ capabilities are rapidly expanding to engineer natural microbiomes, design synthetic microbiomes, and introduce novel microbes into an existing microbiome. These efforts enable academic researchers and industry to interrogate host-microbe and microbe-microbe interactions, perform complex tasks impossible by a single engineered microbial species alone, and improve host and environmental outcomes.³ These seemingly disparate applications of engineered microbiomes highlight how basic discoveries and common tools can be shared to elevate the entire discipline and progress toward solving national and global challenges.

To enable both fundamental and applied research and development of engineered microbiomes, the Federal Government should: i) fund the development of technical methods and computational tools to enable advanced microbiome manipulation; ii) invest in the infrastructure to test engineered microbiomes at scale, in environmentally relevant contexts; iii) develop a central data repository for pre-competitive data sharing of microbiome discoveries. It is notable that these recommendations align with and would enable several agencies’ [Bold Goals](#) for biotechnology and biomanufacturing compiled by OSTP. These efforts should be coordinated across the Federal Government to capitalize on momentum and bolster productivity throughout the field with various applications.

Increase support for the development of scalable engineering techniques, next-generation -omics technologies, and accessible AI/ML bioinformatics methods to enable advanced microbiome manipulation.

While established techniques in synthetic biology are routinely used to genetically engineer microbes, further development of technical methods for engineering, testing, and evaluating *microbiomes* is needed. Engineering biology techniques that work on single-cells or a small number of species in co-culture do not always yield expected outcomes when applied to a native or native-like microbiome.⁴ Techniques are even less likely to

¹ Engineering Biology Research Consortium (2020). *Microbiome Engineering: A Research Roadmap for the Next-Generation Bioeconomy*. Retrieved from <http://roadmap.ebrc.org>. doi: 10.25498/E4QP4T.

² *Interagency Strategic Plan for Microbiome Research, FY 2018-2022*. 2018. doi:10.2172/1471707.

³ Leggieri, P. et al. (2021). *Integrating Systems and Synthetic Biology to Understand and Engineer Microbiomes*. *Annu. Rev. Biomed. Eng.* 2021. 23:169–201.

⁴ Brooks, S.M., Alper, H.S. Applications, challenges, and needs for employing synthetic biology beyond the lab. *Nat Commun* 12, 1390 (2021). <https://doi.org/10.1038/s41467-021-21740-0>.

succeed when applied to non-model microbial components; thus, this is a priority area for research. Approaches such as *in situ* exogenous editing of a single targeted species within a microbiome, narrow-spectrum antibiotic control, and metabolic enmeshment between a host and microbiome could facilitate important advances across the field and would benefit from continued investment.⁵ Next-generation -omics technologies are improving scientists' capabilities to quantify and track nucleic acids, proteins, and metabolites in order to better interrogate microbial function, activity, and interactions in their native communities. These characterizations help to describe microbiome composition and dynamics in their natural environment and/or when interfacing with another organism, such as in the plant rhizosphere or in the human gut. Such tools can be used to better understand system homeostasis and disease states and determine what host and microbial factors drive microbiome composition and the engraftment or rejection of additional member organisms, whether beneficial or detrimental to the host system.

Another important goal to enable microbiome engineering remains to determine what phenotypes occur with unique combinations of microbes and under specific conditions, similar to genome-wide association studies.⁶ Such bioinformatic predictions are most accurate when based on the same or closely related species which have been characterized in the laboratory. Yet bioinformatic predictions of function based on minimal homology tenuously ascribe links between function and single species within a microbiome. This uncertainty may result in unpredictable behaviors when attempting to engineer organisms toward desired outcomes within an existing system.⁷ Due to the vast amount of data processing this requires, continued investment for collaboration between microbiome researchers and bioinformaticians specializing in artificial intelligence and machine learning (AI/ML) is necessary. Ultimately, user-friendly programs and accessible algorithms could expand the ability of researchers at all levels of computing expertise to analyze generated data, keeping pace as access to next-generation -omics technologies continue to grow. With a better understanding of how certain phenotypes occur, engineering biology techniques can be used to control or reengineer those phenotypes for a specific application.

Invest in nationally accessible infrastructure necessary to test engineered microbiomes at scale in application-relevant environments.

Akin to a clinical trial design, careful and gradual steps are important for engineering validation, safety, and efficacy when translating between a laboratory-engineered microbiome and its final application in a natural environment. Investment in accessible animal, field, and water testing sites at incremental scales would make it more feasible to establish functional profiles and understand the limitations and vulnerabilities of an engineered product. Enabling more precise control over the microbiota to which an engineered microbe is being introduced could help answer fundamental scientific questions to track long-term microbe persistence, gene transfer within the microbiome, and quantify “dosage” of a small molecule or therapeutic released from an engineered microbe. Furthermore, advancements into understanding and predicting how engineered changes perform and persist in increasingly complex environments are needed. This is especially evident in recent work on biorefining through anaerobic digestion from wastewater and using synthetic consortia for biofuel and chemical production or other critical [carbon conversion applications](#) highlighted in the [EBRC Roadmap of Engineering Biology for Climate and Sustainability](#), as promising discoveries have had difficulty

⁵ Lawson, C.E., Harcombe, W.R., Hatzenpichler, R. *et al.* Common principles and best practices for engineering microbiomes. *Nat Rev Microbiol* **2019**, 17, 725–741.

⁶ See data science section of [An Assessment of Short-Term Milestones in EBRC's 2019 Roadmap, Engineering Biology](#).

⁷ Pérez-Enciso, M., Zingaretti, L.M., Ramayo-Caldas, Y. *et al.* Opportunities and limits of combining microbiome and genome data for complex trait prediction. *Genet Sel Evol* **2021**, 53, 65. <https://doi.org/10.1186/s12711-021-00658-7>.

transitioning to real-world samples and environmental field testing with non-model species.^{8,9} Understanding and predicting performance and persistence in complex environments is also important for realistic [biosafety and biosecurity assessments](#) for engineered microbiomes. Such exploratory, yet essential questions have received critical investment from agencies including the [DOE](#) and [NSF](#), which could serve as a model for future funding mechanisms.

Interagency collaboration in construction for testing microbiome research at several scales could ensure that scalable infrastructure is sustainably funded and focused in a coordinated, non-redundant manner. While this infrastructure could be built by larger companies and universities in-house, it may be inaccessible to startups or smaller universities and impractical for those institutions to build, potentially causing valuable innovations to be abandoned. In general, smaller-scale controlled spaces with appropriate environmental conditions are needed to investigate the applications of novel technologies yet remain unaffordable to many. Some university-industry partnerships and national labs have invested in field-testing infrastructure and clean animal facilities to perform microbiome experiments with explicit control over microbial composition. One successful example that could be replicated to expand access nationally is the [Harvard Digestive Disease Center](#), which has illustrated how investment in clean animal facilities can effectively test hypotheses of precisely constructed microbiota in animal models of disease and treatment. These examples demonstrate the substantial advances that can be accomplished with sustained federal investment for microbiome testing, but efforts are still needed to expand their accessibility across the U.S.

House a central data repository with standardized reporting formats for pre-competitive data sharing.

With current -omics technologies, nearly all microbiome researchers are capable of generating a modest-sized dataset that could be useful if shared with others. The hoarding of data is a pressing issue in several areas of biology, yet it is especially notable in microbiome research where sharing real-world sampling data could build a more collective and cohesive picture of various microbiomes to aid all researchers. As researchers and funders invest more into [genotype-to-phenotype](#) and systemic physiology experiments, resulting findings have the potential to be impactful for a wide range of microbiome technologies in virtually every sector of the bioeconomy. While the research community recognizes the value of data sharing, and many journals and funders require it, there is no agreed format or universally accessible tool to collect and analyze findings in a manner usable by other scientists. As a result, investigators and new companies carry an undue burden of building data sets from scratch that may already exist elsewhere. Larger companies invest significant capital in building large datasets and have no incentive to share after it has been collected and stored in-house. Yet if a company fails, this valuable intellectual property can be lost. Early-stage investigators would benefit if there were a pre-competitive data sharing agreement to a central repository maintained at the federal level and accessible by all those who wish to analyze it.¹⁰ However, in doing so, there must also be agreed standards on the types of data collected and the formats they are reported so that anyone can analyze that data. It is important that early findings from such nationally funded endeavors be cataloged sufficiently and disseminated widely to inform efforts across the microbiome research enterprise.

⁸ Carr, S.; Aldridge, J.; Buan, N.R. Isoprene Production from Municipal Wastewater Biosolids by Engineered Archaeon *Methanosarcina acetivorans*. *Appl. Sci.* **2021**, *11*, 3342.

⁹ Vaccaro F, Cangioli L, Mengoni A, Fagorzi C. Synthetic plant microbiota challenges in nonmodel species. *Trends Microbiol.* **2022** Oct; 30 (10):922-924.

¹⁰ Hodgson, A., Alper, J., Maxon, M.E. **2022**. The U.S. Bioeconomy: Charting a Course for a Resilient and Competitive Future. New York, New York: Schmidt Futures. <https://doi.org/10.55879/d2hrs7zwc>.

Several data initiatives have been made at various agencies,¹¹ but continued investment, interagency collaboration, and decreased barriers to scientists working across agency boundaries are necessary to ensure their success. The [National Microbiome Data Collaborative](#), [Protein Data Bank](#), and [EMBL-EBI MGnify](#) program could serve as successful models for constructing databases and standardizing comprehensive data collection, but coordinated communication of these resources would be incredibly beneficial. If a central repository were federally operated, plug-and-play analytical tools could also be developed, and microbiome researchers could be trained to use and report to the repository. Ultimately, user-friendly programs and accessible algorithms could expand the ability of researchers at all levels of computing expertise to analyze generated data, keeping pace as access to next-generation -omics technologies continue to grow. The American Society for Microbiology and other scientific consortia could be consulted for the types of standards required, building a culture of reporting, and advertising publicly available tools. It is important to recognize that microbiome data sets and analytical tools can be informative across sectors, so separating health microbiome applications from agriculture or energy applications, for example, may be inefficient.

Conclusion

The capabilities of engineered microbiomes are rapidly accelerating, but there are still fundamental questions that should be answered. Coordinated policy and investment in microbiome engineering tool development, computing, and scalable infrastructure can be established to promote the safe expansion of the US bioeconomy. Engineered microbiomes can serve as living tools to interrogate the systems and environments and have several beneficial applications, including in biosensing, bioremediation, agriculture, and therapeutic synthesis and delivery. When the discoveries of microbiome researchers are shared, the possibilities of engineered microbiota remain bright to improve human and planetary health.

¹¹ See the [National Microbiome Data Collaborative](#), NIH [Common Fund Data Ecosystems](#), NSF [CyVerse](#) and [Environmental Data Science Innovation and Inclusion Lab](#), and DOE [Integrated Microbial Genomes and Microbiomes](#)