

The Architecture of Theory and Data in Microbiome Design: towards an S-matrix for microbiomes

Shreya Arya¹, Ashish B. George^{2,3}, and James O'Dwyer^{2,✉}

¹Department of Physics, University of Illinois, Urbana-Champaign, Urbana, IL 61801, USA

²Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

³Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

Designing microbiomes for applications in health, bioengineering, and sustainability is intrinsically linked to a fundamental theoretical understanding of the rules governing microbial community assembly. Microbial ecologists have used a range of mathematical models to understand, predict, and control microbiomes, ranging from mechanistic models, putting microbial populations and their interactions as the focus, to purely statistical approaches, searching for patterns in empirical and experimental data. We review the success and limitations of these modeling approaches when designing novel microbiomes, especially when guided by (inevitably) incomplete experimental data. Although successful at predicting generic patterns of community assembly, mechanistic and phenomenological models tend to fall short of the precision needed to design and implement specific functionality in a microbiome. We argue that to effectively design microbiomes with optimal functions in diverse environments, ecologists should combine data-driven techniques with mechanistic models—a middle, third way for using theory to inform design.

Introduction

Microbiomes are complex systems, comprising multiple taxa interacting with their biotic and abiotic environment, and can be found from inhospitable cold, heat, and acidity (1), to the familiar host-associated environments of humans, plants, and animals (2). Across these environments, microbiomes modulate a broad range of functions, from biogeochemical cycles to host health (3–6). This breadth of possible microbiome function has highlighted the potential for us to engineer microbiomes to produce novel, beneficial outcomes, or even just to keep pace with the demands of a changing global environment (7). Humans as engineers of microbiomes have a long history: with records going as far back as 10000 years ago, human populations have been developing fermented food and drink, including wine, beer, and dairy products (8), extending more recently to produce biofuel, enhance treatment of infections and inflammatory conditions, increase crop yield, and waste-water treatment (9, 10, 18, 19). Microbial communities can exhibit emergent functions that no one microbe can perform, and there is also evidence for greater stability and resilience of diverse consortia (12, 13). While we can quantify communities in unprecedented detail (14–17), the question is whether we can go beyond merely documenting communities, and harness these data to inform design and engineering of functionally-relevant microbiomes.

A brute-force search for novel or optimized functions is unlikely to be efficient. Theoretical frameworks have often provided guiding principles in ecology and beyond, suggesting that the right theory might allow us to characterize the outcome of microbiome assembly, and thus predict function. The interplay between prediction and theoretical modeling mirrors a related goal in a seemingly distant area of science: calculating the scattering- or S-matrix in theoretical physics. The S-matrix provides a framework for mapping initial particles and momenta to predict late-time outcomes following interactions among those incoming particles. While the component parts differ, this resembles the goal of mapping taxonomic and nutrient initial conditions to late-time community function. We'll review several microbiome modeling approaches below (20–23). But the difficulty in matching predictive power with available taxonomic, genomic, and metabolomic data means that we will likely require a re-thinking of modeling in ecology, and potentially need to embrace data-driven design and control alongside traditional approaches. This may reveal principles of microbiome design that are independent of any specific theoretical model.

Mechanistic and Phenomenological Models

Models assuming pairwise interactions among microbial taxa (24–26) are characterized by a community matrix, $A = \{a_{ij}\}$, where the strength and the sign of the coefficient a_{ij} denotes the effect that species j has on the growth of species i . The generalized Lotka-Volterra (gLV) model is the standard way to model pairwise interactions, and is able to accommodate many kinds of observed behaviors: amensalism, commensalism, exploitation, competition, and mutualism; this flexibility as much as anything else has positioned gLV as a foundational model in studying microbial communities. This approach can be extended to include higher-order interactions (27–30) or to include the role of space (31). In particular, inspired by Robert May's use of random interaction matrices to probe the diversity-stability relationships in ecology, and using tools from physics of disordered systems, random gLV models have been explored extensively to inform equilibria, community coexistence, stability, and feasibility outcomes (25, 26, 32).

The gLV model has been used to make predictions for statistical patterns in microbiomes, as well as coarse-grained properties like the dependence of community stability on community diversity. But the number of parameters required to model a specific community grows quadratically with the size of the pool, and so a major challenge is posed by the number of experiments needed to infer these parameters accurately. When considering in-vivo communities like the human gut microbiome, this data suffers from non-uniformity of sampling times, and experimental noise, although studies have incorporated various algorithms to address these issues (33–35). Many studies have applied gLV models to infer parameters for in-vitro and in-vivo communities, and for explanation and prediction of community function (17, 26, 27, 35–38). Though when community function is something that depends in a non-trivial way on species composition (for example, the production of a metabolite) additional hypotheses must then be generated and tested for the relationship between composition and function, and this mapping might well be idiosyncratic to a given community and context. In summary, the strength of the gLV model lies

in the potential for explaining generic, statistical assembly patterns of microbiomes, and in making specific predictions, but only when the diversity of a community is relatively small and the relationship between late-time taxonomic composition and function can be independently estimated.

One of the main drawbacks of the gLV modelling framework is that it does not account for chemical mediators and resources and metabolites—some of the key mechanisms underlying species interactions. As a result, gLV parameters inferred from a system with a given resource and nutrient condition are unlikely to extend to arbitrarily-different resource environments. Second, the addition of new taxa may change the effective strength of existing pairwise interactions through modification of resources available to other species (39, 40). This makes pairwise models like the gLV unsuitable for design of consortia whenever there are strong effects of this type. This issue has been addressed using adaptations of the classic MacArthur’s consumer-resource model (41) that include explicit dynamics for both abiotic resources and species dynamics (42, 43). More generally, chemically-mediated interactions via toxins and anti-microbial production (44) can also be included. Consumer-resource models also hold a lot of promise in incorporating the various aspects of microbial metabolism. Constraints on and flexibility of microbial metabolism leads to diverse phenotypes and behaviors of consumer-resource dynamics.

These additional mechanisms also generally increase the number of parameters, which makes accurate parameter inference challenging. This issue is compounded by the fact that it is generally harder to measure and interpret resource concentrations. These twin difficulties in fitting consumer-resource models means that they have so far been used less often than gLV models for parameter inference from experiments. More often these models have also been used to explain statistical patterns in microbial communities, including stability-diversity relationships (42, 43, 45, 46), and qualitative patterns in large, varied data such as the Human Microbiome and the Earth Microbiome Projects (46, 47). But as estimating metabolite concentrations becomes more tractable, we expect a continuing move towards fitting consumer-resource model parameters using combined resource and taxonomic data. Gowda et al (48) used a regression model to map gene content to consumer-resource model parameters and then used the consumer-resource model to predict metabolite dynamics in consortia, while Ho et al (16) used a coarse-grained consumer-resource model to characterize communities assembled from a 15-species pool. This work highlights how conservative metabolomics and spent-media experiments pave the way for using coarse-grained CR models to make predictions for community function.

In addition to the generalized Lotka-Volterra model and consumer-resource dynamics, more detailed metabolic models also provide a way of understanding, designing, and controlling microbiomes. Consumer-resource models themselves have successfully captured and explained the presence of auxotrophy as a stabilizing force in microbiomes, how diauxic shifts explain coexistence, and the behaviors of slow growing communities. But sequencing across a broad taxonomic range means that we are able to map the precise metabolic pathways of many individual taxa. Genome-scale metabolic models for individual taxa can be integrated to build models that capture metabolic functions in whole communities and also summarize interactions between species. However, most of these consortia comprise only a few strains, bringing scalability into question. While both consumer-resource and genome-scale models speak to the use of resources, they lie at either end of a mechanistic spectrum. It could be that intermediate, coarse-grained descriptions of internal metabolism might provide a bridge, trading off data overhead (the potentially prohibitive requirements for appropriate training data) and mechanistic accuracy. Metabolically informed consumer-resource models can give insights into small consortia, correctly capturing dynamics and interaction mechanisms (20, 49–52)

In summary, every mechanistic model we know of has its drawbacks, centering around difficulties in accurately formulating and fitting these models given limited data. The full spectrum of mechanisms which underpin microbiome stability also includes spatial structure and dynamics resulting from spatial heterogeneity, adding further complexity to the application of these models. Any more exhaustive range of mechanistic models would have many of the same (or exacerbated) issues of data overhead and model specification.

Data-driven modelling and model-free approaches

Mechanistic models serve as excellent starting points to make sense of microbiomes, but have these clear limitations. Integration of these models with other approaches may provide a key step needed for effective design and control of microbiomes. For example, it is possible to combine pairwise genome-scale models with generalized Lotka-Volterra models, as shown recently in the context predicting engraftment of probiotics in the human gut microbiome (53). A broad range of other machine learning and statistical methods have also been used to analyze and predict microbiome composition and function. In the context of colonization prediction, mechanism-agnostic approaches like logistic regression, random forest regression, and a compositional neural net have been used (23). It may also be possible to guide design of communities and understand co-variation in host phenotypes and community compositions using generative models that are based off of mechanistic

frameworks like the consumer-resource model (54). A way to improve model performance, wherever possible, is to physically constrain the model architecture to allow for biologically sensible outcomes (55). These include obvious constraints, like ensuring that species abundances are non-negative and species which were absent in an assemblage are not predicted to be present in the steady-state outcome. When the target variable for prediction is a metabolite, the target value can also be constrained by known biochemical limits and stoichiometry. Furthermore, search for low-dimensional representation of data via auto-encoders (56) or using graph neural networks that build on microbial interactions as a network (57) may be employed for prediction of community structures. This area current lacks guiding principles, but certainly a range of methods have been employed to predict (58–61) and understand interactions amongst microbes (62).

The downside to ML-based and deep learning methods is that interpretability might be compromised—we might be able to predict outcomes within a given parameter range, but we don't know *why* we are predicting these well, or where this prediction might begin to break down. Post-hoc methods like LIME (63) and SHAP (64) have been used to partially address this issue by building interaction networks, and thus providing both data stratification and insights on biological mechanisms (65). Even so, these models may be tailored for the specific datasets they were built for, and maybe need to be re-hauled in terms of hyperparameter tuning or model architecture when translating to other data, undermining their robustness.

Recent work has made progress using ‘middle-ground’, statistical approaches which straddle the divide between mechanistic models and deep learning methods. To this end, community-function and community abundance landscapes have been defined and studied (66–68). Inspired by fitness landscapes in evolutionary biology which map genotypes to phenotypes, these ecological landscapes map community assembly to steady-state composition and function. With this mapping between composition and function, it has been possible to efficiently navigate the landscape of functions looking for functional optima. Within the scope of this kind of ‘landscape thinking’, studies demonstrate that across a variety of environments, both abundances and functions at steady-state can be predicted from species presence/absence data (69, 70). The success of these approaches suggest that microbial community assembly is constrained: with only a few effective, statistical interactions. These regression models also show signals of sparsity in the landscape interaction space, akin to sparse higher-order epistasis in protein and genomics datasets (71). This sparsity of higher-order statistical interactions can be leveraged to infer entire landscapes from limited experimental data using regression techniques with weights. We think that these signatures of sparsity may be used to a) aid design of experiments, b) inform and refine mechanism and interaction based models like gLV and CR, and c) define the role of environmental heterogeneity when probing the same sets of species across habitats. Further development of these approaches may need to generalize the input-output map to incorporate environmental conditions and nutrient availability, along side taxonomic composition, as part of the ‘input’, leading to community function as an output. The landscape might then incorporate regimes where nutrients determine function independent of taxonomy (72), regimes where taxonomy is key (70), and regimes where both act in concert to determine function.

Discussion

Approaches to the control and design of microbiomes with optimal functions must wrestle with the question of what's important: no single model can capture and integrate the complexity posed by all the fluctuating biotic and abiotic components of an ecosystem. While phenomenological and mechanistic models like the generalized Lotka-Volterra and the consumer-resource models can be, and have been used to design and test functionally beneficial communities, their success in making accurate predictions is compromised by either model limitations or the number of parameters to fit. This means that models like generalized Lotka-Volterra, Consumer-resource models, and even metabolic models mostly enjoy success when employed to understand and explain statistical patterns of communities, rather than precise predictions of function—at least in cases where these communities comprise large numbers of diverse taxa. On the other hand, approaches which rely on machine learning methods tend to sideline mechanisms and hence may not be straightforwardly translated to datasets beyond those they were developed for. In between these two sets of approaches, methods which rely on finding low-dimensional representations of microbial community-function landscapes may prove to be useful and worthy of further exploration.

We think that some recent successes lay out a potential roadmap, integrating traditional modeling and statistical learning to understand, predict, and design microbiomes. The key here may be two-fold: first, identifying the right units of analysis. Coarse-grained taxonomic units, or functional groups, will likely play a role in making microbiome analysis tractable. But the importance of metabolism and consumer-resource dynamics suggests that we may need to go beyond just finding the right taxonomic units: we will likely need to identify coarse-grained “modules” of consumers and resources, encoding the key processes, not just the key players in a microbial community. In short, we need to identify functional groups in natural communities using methods that go beyond taxonomy and phylogeny.

The end-point of this back-and-forth might resemble progress in other fields. The particle physics S-matrix was at one

point proposed to be uniquely-determined by fundamental principles (73). Later, it was thought of as predictable, but only given knowledge of the specific combinations of symmetries and particles in nature, and modeled using the language of field theory (74). While now the S-matrix is sometimes thought of again as an object of study in its own right, with certain fundamental features that are independent of any specific, mechanistic models (75). The map from input nutrients and taxa to late-time community function might turn out to be the S-matrix of microbiome ecology. Understanding this input-output map may similarly benefit from back-and-forth between mechanistic models and the identification of fundamental, inevitable structures. Simplifying features in the composition-function landscape certainly suggest that there may be robust rules that govern this microbial S-matrix. But it could also be that we will only be able to leverage those rules for microbiome design if we formulate microbiome composition and dynamics in the right language, and using the right variables.

Author Contribution Statement: All authors contributed equally to this manuscript.

Bibliography

1. Wen-Sheng Shu and Li-Nan Huang. Microbial diversity in extreme environments. *Nature Reviews Microbiology*, 20(4):219–235, 2022.
2. Courtney J Robinson, Brendan JM Bohannan, and Vincent B Young. From structure to function: the ecology of host-associated microbial communities. *Microbiology and Molecular Biology Reviews*, 74(3):453–476, 2010.
3. Paul G Falkowski, Tom Fenchel, and Edward F Delong. The microbial engines that drive earth's biogeochemical cycles. *science*, 320(5879):1034–1039, 2008.
4. Siyu Chen, Shuyan Luo, and Chao Yan. Gut microbiota implications for health and welfare in farm animals: A review. *Animals*, 12(1):93, 2021.
5. Hirokazu Toju, Kabir G Peay, Masato Yamamichi, Kazuhiko Narisawa, Kei Hiruma, Ken Naito, Shinji Fukuda, Masayuki Ushio, Shinji Nakaoka, Yusuke Onoda, et al. Core microbiomes for sustainable agroecosystems. *Nature plants*, 4(5):247–257, 2018.
6. Kaijian Hou, Zhuo-Xun Wu, Xuan-Yu Chen, Jing-Quan Wang, Dongya Zhang, Chuanxing Xiao, Dan Zhu, Jagadish B Koya, Liuya Wei, Jilin Li, et al. Microbiota in health and diseases. *Signal transduction and targeted therapy*, 7(1):1–28, 2022.
7. Ricardo Cavicchioli, William J Ripple, Kenneth N Timmis, Farooq Azam, Lars R Bakken, Matthew Baylis, Michael J Behrenfeld, Antje Boetius, Philip W Boyd, Aimée T Classen, et al. Scientists' warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology*, 17(9):569–586, 2019.
8. Jean-Luc LeGras, Didier Merdinoglu, Jean-Marie Cornuet, and Francis Karst. Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Molecular ecology*, 16(10):2091–2102, 2007.
9. Thomas Louie, Yoav Golan, Sahil Khanna, Dmitri Bobilev, Nathalie Erpelding, Candida Fratazzi, Meg Carini, Rajita Menon, Mary Ruisi, Jason M Norman, et al. Ve303, a defined bacterial consortium, for prevention of recurrent clostridioides difficile infection: a randomized clinical trial. *Jama*, 329(16):1356–1366, 2023.
10. Sudarshan Paramsothy, Michael A Kamn, Nadeem O Kaakoush, Alissa J Walsh, Johan van den Bogaerde, Douglas Samuel, Rupert WL Leong, Susan Connor, Watson Ng, Ramesh Paramsothy, et al. Multidonor intensive faecal microbiota transplantation for active ulcerative colitis: a randomised placebo-controlled trial. *The Lancet*, 389(10075):1218–1228, 2017.
11. Matti Gralka. Searching for Principles of Microbial Ecology Across Levels of Biological Organization. *Integrative and Comparative Biology*, page icad060, June 2023. ISSN 1540-7063. doi: 10.1093/icb/icad060.
12. Jasmine Shong, Manuel Rafael Jimenez Diaz, and Cynthia H Collins. Towards synthetic microbial consortia for bioprocessing. *Current Opinion in Biotechnology*, 23(5):798–802, 2012.
13. Wael Sabra, David Dietz, Donna Tjahjajari, and An-Ping Zeng. Biosystems analysis and engineering of microbial consortia for industrial biotechnology. *Engineering in Life Sciences*, 10(5):407–421, 2010.
14. Curtis Huttenhower, Dirk Gevers, Rob Knight, Sahar Abubucker, Jonathan H Badger, Asif T. Chinwalla, Heather H. Creasy, Ashlee M. Earl, Michael G. FitzGerald, Robert S. Fulton, Michelle G. Giglio, Kimberlie Hallsworth-Pepin, Elizabeth A. Lobos, Ramana Madupu, Vincent Magrini, John C. Martin, Makedonka Mitreva, Donna M. Muzny, Erica J. Sodergren, James Versalovic, Aye M. Wollam, Kim C. Worley, Jennifer R. Wortman, Sarah K. Young, Qiandong Zeng, Kjersti M. Aagaard, Olukemi O. Abulode, Emma Allen-Vercoe, Eric J. Alm, Lucia Alvarado, Gary L. Andersen, Scott Anderson, Elizabeth Appelbaum, Harindra M. Arachchi, Gary Armitage, Cesar A. Arze, Tulin Ayvaz, Carl C. Baker, Lisa Begg, Tsegahiwot Belachew, Veena Bhonegiri, Monika Bihan, Martin J. Blaser, Toby Bloom, Vivien Bonazzi, J. Paul Brooks, Gregory A. Buck, Christian J. Buhay, Dana A. Busam, Joseph L. Campbell, Shane R. Canon, Brandi L. Cantarel, Patrick S. G. Chain, I-Min A. Chen, Lei Chen, Shailla Chhibba, Ken Chu, Dawn M. Ciulla, Jose C. Clemente, Sandra W. Clifton, Sean Conlan, Jonathan Crabtree, Mary A. Cutting, Noam J. Davidovics, Catherine C. Davis, Todd Z. DeSantis, Carolyn Deal, Kimberley D. Delehaunty, Floyd E. Dewhirst, Elena Deych, Yan Ding, David J. Dooling, Shannon P. Dugan, Wm Michael Dunne, A. Scott Durkin, Robert C. Edgar, Rachel L. Erlich, Candace N. Farmer, Ruth M. Farrell, Karoline Faust, Michael Feldgarden, Victor M. Felix, Sheila Fisher, Anthony A. Fodor, Larry J. Forney, Leslie Foster, Valentina Di Francesco, Jonathan Friedman, Dennis C. Friedrich, Catrina C. Fronick, Lucinda L. Fulton, Hongyu Gao, Nathalia Garcia, Georgia Giannoukos, Christina Giblin, Maria Y. Giovanni, Jonathan M. Goldberg, Johannes Goll, Antonio Gonzalez, Allison Griggs, Sharvari Gujja, Susan Kinder Haake, Brian J. Haas, Holli A. Hamilton, Emily L. Harris, Theresa A. Hepburn, Brandi Herter, Diane E. Hoffmann, Michael E. Holder, Clinton Howarth, Katherine H. Huang, Susan M. Huse, Jacques Izard, Janet K. Jansson, Huaiyang Jiang, Catherine Jordan, Vandita Joshi, James A. Katancik, Wendy A. Keitel, Scott T. Kelley, Cristyn Kells, Nicholas B. King, Dan Knights, Heidi H. Kong, Omry Koren, Sergey Koren, Karthik C. Kota, Christie L. Kovar, Nikos C. Kyrpides, Patricio S. La Rosa, Sandra L. Lee, Katherine P. Lemon, Niall Lennon, Cecil M. Lewis, Lora Lewis, Ruth E. Ley, Kelvin Li, Konstantinos Liolios, Bo Liu, Yue Liu, Chien-Chi Lo, Catherine A. Lozupone, R. Dwayne Lunsford, Tessa Madden, Anup A. Mahurkar, Peter J. Mannon, Elaine R. Mardis, Victor M. Markowitz, Konstantinos Mavromatis, Jamison M. McCarrison, Daniel McDonald, Jean McEwen, Amy L. McGuire, Pamela McInnes, Teena Mehta, Kathie A. Mihindukulasuriya, Jason R. Miller, Patrick J. Minx, Irene Newsham, Chad Nusbaum, Michelle O'Laughlin, Joshua Orvis, Ioanna Pagani, Krishna Palaniappan, Shital M. Patel, Matthew Pearson, Jane Peterson, Mircea Podar, Craig Pohl, Katherine S. Pollard, Mihai Pop, Margaret E. Priest, Lita M. Proctor, Xiang Qin, Jeroen Raes, Jacques Ravel, Jeffrey G. Reid, Mina Rho, Rosamond Rhodes, Kevin P. Riehl, Maria C. Rivera, Beltran Rodriguez-Mueller, Yu-Hui Rogers, Matthew C. Ross, Carsten Russ, Ravi K. Sanka, Pamela Sankar, J. Fah Sathirapongsasuti, Jeffery A. Schloss, Patrick D. Schloss, Thomas M. Schmidt, Matthew Scholz, Lynn Schriml, Alyxandria M. Schubert, Nicola Segata, Julia A. Segre, William D. Shannon, Richard R. Sharp, Thomas J. Sharpton, Narmada Shenoy, Nihar U. Sheth, Gina A. Simone, Indresh Singh, Christopher S. Smillie, Jack D. Sobel, Daniel D. Sommer, Paul Spicer, Granger G. Sutton, Sean M. Sykes, Diana G. Tabbaa, Mathangi Thiagarajan, Chad M. Tomlinson, Manolito Torralba, Todd J. Treangen, Rebecca M. Truty, Tatiana A. Vishnivetskaya, Jason Walker, Lu Wang, Zhengyuan Wang, Doyle V. Ward, Wesley Warren, Mark A. Watson, Christopher Wellington, Kris A. Wetterstrand, James R. White, Katarzyna Wilczek-Boney, Yuanqing Wu, Kristine M. Wylie, Todd Wylie, Chandri Yandava, Liang Ye, Yuzhen Ye, Shibu Yooseph, Bonnie P. Youmans, Lan Zhang, Yanjiao Zhou, Yiming Zhu, Laurie Zoloth, Jeremy D. Zucker, Bruce W. Birren, Richard A. Gibbs, Sarah K. Highlander, Barbara A. Methé, Karen E. Nelson, Joseph F. Petrosino, George M. Weinstock, Richard K. Wilson, Owen White, and The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. 486(7402):207–214. ISSN 1476-4687. doi: 10.1038/nature11234. Publisher: Nature Publishing Group.
15. Luke R. Thompson, Jon G. Sanders, Daniel McDonald, Amnon Amir, Joshua Ladau, Kenneth J. Locey, Robert J. Prill, Anupriya Tripathi, Sean M. Gibbons, Gail Ackermann, Jose A. Navas-Molina, Stefan Janssen, Evgenia Kopylova, Yoshiki Vázquez-Baeza, Antonio González, James L. Morton, Siavash Mirarab, Zhenjiang Zech Xu, Lingjing Jiang, Mohamed F. Haroon, Jad Kanbar, Qiyun Zhu, Se Jin Song, Tomasz Kosciolk, Nicholas A. Bokulich, Joshua Lefler, Colin J. Brislawn, Gregory Humphrey, Sarah M. Owens, Jarrod Hampton-Marcell, Donna Berg-Lyons, Valerie McKenzie, Noah Fierer, Jed A. Fuhrman, Aaron Clauset, Rick L. Stevens, Ashley Shade, Katherine S. Pollard, Kelly D. Goodwin, Janet K. Jansson, Jack A. Gilbert, and Rob Knight. A communal catalogue reveals earth's multiscale microbial diversity. 551(7681):457–463. ISSN 1476-4687. doi: 10.1038/nature24621. Publisher: Nature Publishing Group.
16. Po-Yi Ho, Taylor H Nguyen, Juan M Sanchez, Brian C DeFelice, and Kerwyn Casey Huang. Resource competition predicts assembly of gut bacterial communities in vitro. *Nature Microbiology*, pages 1–13, 2024.
17. Ryan L. Clark, Bryce M. Connors, David M. Stevenson, Susan E. Hromada, Joshua J. Hamilton, Daniel Amador-Noguez, and Ophelia S. Venturielli. Design of synthetic human gut microbiome assembly and butyrate production. *Nat Commun*, 12(1):3254, December 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-22938-y.
18. Posy E Busby, Chinmay Soman, Maggie R Wagner, Maren L Friesen, James Kremer, Alison Bennett, Mustafa Morsy, Jonathan A Eisen, Jan E Leach, and Jeffery L Dangl. Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS biology*, 15(3):e2001793, 2017.
19. Yuancai Chen, Che-Jen Lin, Gavin Jones, Shiyu Fu, and Huaiyu Zhan. Enhancing biodegradation of wastewater by microbial consortia with fractional factorial design. *Journal of hazardous materials*, 171(1-3):948–953, 2009.
20. Matti Gralka, Shaul Pollak, and Otto X Cordero. Genome content predicts the carbon catabolic preferences of heterotrophic bacteria. *Nature Microbiology*, 8(10):1799–1808, 2023.
21. Yang-Yu Liu. Controlling the human microbiome. *Cell Systems*, 14(2):135–159, February 2023. ISSN 2405-4712. doi: 10.1016/j.cels.2022.12.010.
22. Yili Qian, Freeman Lan, and Ophelia S Venturielli. Towards a deeper understanding of microbial communities: integrating experimental data with dynamic models. *Current Opinion in Microbiology*, 62:84–92, August 2021. ISSN 1369-5274. doi: 10.1016/j.mib.2021.05.003.
23. Shengbo Wu, Zheping Qu, Danlei Chen, Hao Wu, Qinggele Caiyin, and Jianjun Qiao. Deciphering and designing microbial communities by genome-scale metabolic modelling. *Computational and Structural Biotechnology Journal*, 2024.
24. Wenping Cui, Robert Marsland III, and Pankaj Mehta. Les houches lectures on community ecology: From niche theory to statistical mechanics. *ArXiv*, 2024.
25. Matthieu Barbier, Jean-François Arnoldi, Guy Bunin, and Michel Loreau. Generic assembly patterns in complex ecological communities. *Proceedings of the National Academy of Sciences*, 115(9):2156–2161, 2018.
26. Jiliang Hu, Daniel R Amor, Matthieu Barbier, Guy Bunin, and Jeff Gore. Emergent phases of ecological diversity and dynamics mapped in microcosms. 2022.
27. Aamir Faisal Ansari, Yugandhar BS Reddy, Janhavi Raut, and Narendra M Dixit. An efficient and scalable top-down method for predicting structures of microbial communities. *Nature Computational Science*, 1(9):619–628, 2021.
28. Mayank Baranwal, Ryan L Clark, Jaron Thompson, Zeyu Sun, Alfred O Hero, and Ophelia S Venturielli. Recurrent neural networks enable design of multifunctional synthetic human gut microbiome dynamics. *eLife*, 11:e73870, June 2022. ISSN 2050-084X. doi: 10.7554/eLife.73870.
29. Theo Gibbs, Gabriel Gellner, Simon A. Levin, Kevin S. McCann, Alan Hastings, and Jonathan M. Levine. Can higher-order interactions resolve the species coexistence paradox?, June 2023. Pages: 2023.06.19.545649 Section: New Results.
30. Theo Gibbs, Simon A. Levin, and Jonathan M. Levine. Coexistence in diverse communities with higher-order interactions. *Proceedings of the National Academy of Sciences*, 119(43):e2205063119, October 2022. doi: 10.1073/pnas.2205063119. Publisher: Proceedings of the National Academy of Sciences.
31. Michael T Pearce, Alish Agarwala, and Daniel S Fisher. Stabilization of extensive fine-scale diversity by ecologically driven spatiotemporal chaos. *Proceedings of the National Academy of Sciences*, 117(25):14572–14583, 2020.
32. Guy Bunin. Ecological communities with lotka-volterra dynamics. *Physical Review E*, 95(4):042414, 2017.
33. Vanni Bucci, Belinda Tzen, Ning Li, Matt Simmons, Takeshi Tanoue, Elijah Bogart, Luxue Deng, Vladimir Yeliseyev, Mary L Delaney, Qing Liu, et al. Mdsine: Microbial dynamical systems inference engine for microbiome time-series analyses. *Genome biology*, 17:1–17, 2016.
34. Travis E Gibson, Younhan Kim, Sawal Acharya, David E Kaplan, Nicholas DiBenedetto, Richard Lavin, Bonnie Berger, Jessica R Allegretti, Lynn Bry, and Georg K Gerber. Intrinsic instability of the dysbiotic microbiome revealed through dynamical systems inference at scale. *BioRxiv*, page 469105, 2021.

35. Charles K. Fisher and Pankaj Mehta. Identifying Keystone Species in the Human Gut Microbiome from Metagenomic Timeseries Using Sparse Linear Regression. *PLoS ONE*, 9(7):e102451, July 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0102451.
36. Ophelia S Venturelli, Alex V Carr, Garth Fisher, Ryan H Hsu, Rebecca Lau, Benjamin P Bowen, Susan Hromada, Trent Northen, and Adam P Arkin. Deciphering microbial interactions in synthetic human gut microbiome communities. *Molecular systems biology*, 14(6):e8157, 2018.
37. Sudeep Ghimire, Chayan Roy, Supapit Wongkuna, Linto Antony, Abhijit Maji, Mitchel Chan Keena, Andrew Foley, and Joy Scaria. Identification of clostridioides difficile-inhibiting gut commensals using culturomics, phenotyping, and combinatorial community assembly. *Msystems*, 5(1):10–1128, 2020.
38. Richard R Stein, Vanni Bucci, Nora C Toussaint, Charlie G Buffie, Gunnar Rättsch, Eric G Pamer, Chris Sander, and Joao B Xavier. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS computational biology*, 9(12):e1003388, 2013.
39. Babak Momeni, Li Xie, and Wenying Shou. Lotka-volterra pairwise modeling fails to capture diverse pairwise microbial interactions. *Elife*, 6:e25051, 2017.
40. James D Brunner and Nicholas Chia. Metabolite-mediated modelling of microbial community dynamics captures emergent behaviour more effectively than species–species modelling. *Journal of the Royal Society Interface*, 16(159):20190423, 2019.
41. Robert MacArthur. Species packing and competitive equilibrium for many species. *Theoretical population biology*, 1(1):1–11, 1970.
42. Robert Marsland III, Wenping Cui, Joshua Goldford, Alvaro Sanchez, Kirill Korolev, and Pankaj Mehta. Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities. *PLoS computational biology*, 15(2):e1006793, 2019.
43. Joshua E. Goldford, Nanxi Lu, Djordje Bajic, Sylvie Estrela, Mikhail Tikhonov, Alicia Sanchez-Gorostiaga, Daniel Segrè, Pankaj Mehta, and Alvaro Sanchez. Emergent simplicity in microbial community assembly. *Science*, 361(6401):469–474, August 2018. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aat1168.
44. Lori Niehaus, Ian Boland, Minghao Liu, Kevin Chen, David Fu, Catherine Henckel, Kaitlin Chung, Suyen Espinoza Miranda, Samantha Dyckman, Matthew Crum, et al. Microbial coexistence through chemical-mediated interactions. *Nature communications*, 10(1):2052, 2019.
45. Stacey Butler and James P O'Dwyer. Cooperation and stability for complex systems in resource-limited environments. *Theoretical ecology*, 13:239–250, 2020.
46. Robert Marsland III, Wenping Cui, and Pankaj Mehta. A minimal model for microbial biodiversity can reproduce experimentally observed ecological patterns. *Scientific reports*, 10(1):3308, 2020.
47. Robert Marsland, Wenping Cui, Joshua Goldford, and Pankaj Mehta. The Community Simulator: A Python package for microbial ecology. *PLoS ONE*, 15(3):e0230430, March 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0230430.
48. Karna Gowda, Derek Ping, Madhav Mani, and Seppe Kuehn. Genomic structure predicts metabolite dynamics in microbial communities. *Cell*, 185(3):530–546, 2022.
49. Mario E Muscarella and James P O'Dwyer. Species dynamics and interactions via metabolically informed consumer-resource models. *Theoretical Ecology*, 13(4):503–518, 2020.
50. Chen Liao, Tong Wang, Sergei Maslov, and Joao B Xavier. Modeling microbial cross-feeding at intermediate scale portrays community dynamics and species coexistence. *PLoS computational biology*, 16(8):e1008135, 2020.
51. Tim N Enke, Manoshi S Datta, Julia Schwartzman, Nathan Cermak, Désirée Schmitz, Julien Barrere, Alberto Pascual-García, and Otto X Cordero. Modular assembly of polysaccharide-degrading marine microbial communities. *Current Biology*, 29(9):1528–1535, 2019.
52. Jean CC Vila, Joshua Goldford, Sylvie Estrela, Djordje Bajic, Alicia Sanchez-Gorostiaga, Alejandro Damian-Serrano, Nanxi Lu, Robert Marsland III, Maria Rebolleda-Gomez, Pankaj Mehta, et al. Metabolic similarity and the predictability of microbial community assembly. *bioRxiv*, 2023.
53. James D Brunner and Nicholas Chia. Metabolic model-based ecological modeling for probiotic design. *Elife*, 13:e83690, 2024.
54. Germán Plata, Madan Krishnamurthy, Lukas Herron, and Purushottam Dixit. Designing host-associated microbiomes using the consumer/resource model, January 2024. Pages: 2023.04.28.538625 Section: New Results.
55. Jaron C. Thompson, Victor M. Zavala, and Ophelia S. Venturelli. Integrating a tailored recurrent neural network with Bayesian experimental design to optimize microbial community functions. *PLOS Computational Biology*, 19(9):e1011436, September 2023. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1011436. Publisher: Public Library of Science.
56. Yasa Baig, Helena R Ma, Helen Xu, and Lingchong You. Autoencoder neural networks enable low dimensional structure analysis of microbial growth dynamics. *Nature Communications*, 14(1):7937, 2023.
57. Albane Ruaud, Cansu Sancaktar, Marco Bagatella, Christoph Ratzke, and Georg Martius. Modelling microbial communities with graph neural networks. In *Forty-first International Conference on Machine Learning*.
58. Sebastian Michel-Mata, Xu-Wen Wang, Yang-Yu Liu, and Marco Tulio Angulo. Predicting microbiome compositions from species assemblages through deep learning. *iMeta*, 1(1):e3, 2022. ISSN 2770-596X. doi: 10.1002/imt2.3. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/imt2.3>.
59. Eitan E. Asher and Amir Bashan. Model-free prediction of microbiome compositions. *Microbiome*, 12(1):17, February 2024. ISSN 2049-2618. doi: 10.1186/s40168-023-01721-9.
60. Tong Wang, Xu-Wen Wang, Kathleen A. Lee-Sarwar, Augusto A. Litorjua, Scott T. Weiss, Yizhou Sun, Sergei Maslov, and Yang-Yu Liu. Predicting metabolomic profiles from microbial composition through neural ordinary differential equations. *Nat Mach Intell*, 5(3):284–293, March 2023. ISSN 2522-5839. doi: 10.1038/s42256-023-00627-3. Number: 3 Publisher: Nature Publishing Group.
61. Sebastian Michel-Mata, Xu-Wen Wang, Yang-Yu Liu, and Marco Tulio Angulo. Predicting microbiome compositions from species assemblages through deep learning. *imeta*, 1(1):e3, 2022.
62. Kateryna Melnyk, Kuba Weimann, and Tim OF Conrad. Understanding microbiome dynamics via interpretable graph representation learning. *Scientific Reports*, 13(1):2058, 2023.
63. Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Why should I trust you? explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144. ACM, 2016. doi: 10.1145/2939672.2939778.
64. Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, pages 4768–4777. Curran Associates Inc., 2017. doi: 10.48550/arXiv.1705.07874.
65. Martina Dal Bello, Hyunseok Lee, Akshith Goyal, and Jeff Gore. Resource–diversity relationships in bacterial communities reflect the network structure of microbial metabolism. *Nature Ecology & Evolution*, 5(10):1424–1434, 2021.
66. Ashish B George and Kirill S Korolev. Ecological landscapes guide the assembly of optimal microbial communities. *PLoS Computational Biology*, 19(1):e1010570, 2023.
67. Alvaro Sanchez, Djordje Bajic, Juan Diaz-Colunga, Abigail Skwara, Jean C. C. Vila, and Seppe Kuehn. The community-function landscape of microbial consortia. *Cell Systems*, 14(2):122–134, February 2023. ISSN 2405-4712. doi: 10.1016/j.cels.2022.12.011.
68. Chandana Gopalakrishnappa, Karna Gowda, Kaumudi Prabhakara, and Seppe Kuehn. An ensemble approach to the structure-function problem in microbial communities. *arXiv:2111.06279 [q-bio]*, November 2021.
69. Abigail Skwara, Karna Gowda, Mahmoud Yousef, Juan Diaz-Colunga, Arjun S Raman, Alvaro Sanchez, Mikhail Tikhonov, and Seppe Kuehn. Statistically learning the functional landscape of microbial communities. *Nature Ecology & Evolution*, 7(11):1823–1833, 2023.
70. Shreya Arya, Ashish B George, and James P O'Dwyer. Sparsity of higher-order landscape interactions enables learning and prediction for microbiomes. *Proceedings of the National Academy of Sciences*, 120(48):e2307313120, 2023.
71. David H Brookes, Amirali Aghazadeh, and Jennifer Listgarten. On the sparsity of fitness functions and implications for learning. *Proceedings of the National Academy of Sciences*, 119(1):e2109649118, 2022.
72. Ashish B George, Tong Wang, and Sergei Maslov. Functional universality in slow-growing microbial communities arises from thermodynamic constraints. *arXiv preprint arXiv:2203.06128*, 2022.
73. Geoffrey F Chew and Steven C Frautschi. Principle of equivalence for all strongly interacting particles within the s-matrix framework. *Physical Review Letters*, 7(10):394, 1961.
74. Steven Weinberg. Conceptual foundations of the unified theory of weak and electromagnetic interactions. *Reviews of Modern Physics*, 52(3):515, 1980.
75. Sebastian Mizera. Physics of the analytic s-matrix. *Physics Reports*, 1047:1–92, 2024.

Annotations

- 16 * This study found that resource competition shapes ecological outcomes in a synthetic community of 15 gut strains derived from humanized mice. They demonstrate how a coarse-grained consumer-resource model can be developed from supernatant screening and untargeted metabolomics data. This model was then successfully employed to predict abundances of strains in various assemblages.
- 20 ** This study highlights how genome content information can be applied to predict coarse-grained descriptors of a community, in this case, the preference for glycolytic versus gluconeogenic substrates in a community of 186 bacteria on 140 carbon sources. This description of the community in terms of preference for either sugars or acid may help provide insights into interactions within this bacterial community.

- 52 ** This study shows that patterns in community compositions can be discerned by linking shared metabolic traits and substrate similarity. This leads to successful predictions of community compositions in novel environments.
- 66 ** This study develops upon, in detail, the idea of structure-function landscapes for microbiomes, inspired from fitness landscapes in evolutionary biology. It links features of a community, like niche overlap between its members, with the ruggedness of the corresponding structure-function landscape.
- 69 ** This study provides a middle-ground approach to study community functions and the structure-function landscape. They demonstrate for various synthetic microcosms that the structure-function landscape is smooth, which makes function prediction possible from relatively low number of observed data-points.
- 70 ** This study provides a middle-ground approach to study community abundances and functions and the structure-function and community- composition landscape. They show that the low ruggedness of these landscapes translates to dominance of effective, low-order interactions amongst the community members. This observation was consequently used to predict abundances and functions using a method that down-weighs higher-order interactions.