Structuring complexity by mapping the possible in microbial ecosystems

Bajić, Djordje; van Oort, Marco; Gabriëls, Minke; Gojkovic, U.G.

DOI
10.1016/j.mib.2025.102658

Publication date
2025

Document Version
Final published version

Published in
Current Opinion in Microbiology

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Review

Structuring complexity by mapping the possible in microbial ecosystems

Djordje Bajić, Marco van Oort, Minke Gabriëls and Uroš Gojković

Microbial ecosystems consist of many interacting components that integrate through stochastic and highly dynamic processes across multiple scales. Yet, despite this complexity, microbial communities exhibit remarkably robust patterns and reproducible functions. This apparent paradox reflects the role of constraints, whether physical, physiological, or evolutionary, that channel stochasticity into structured outcomes. Due to the limited knowledge of the nature of these constraints, models in ecology have traditionally relied on stochastic exploration under minimal mechanistic assumptions. Now, advances in data availability and computational methods increasingly allow us to construct models that incorporate explicit mechanistic constraints. In this review, we synthesize emerging modeling approaches that explore the space of ecological possibility in microbial ecosystems under realistic constraints, such as those imposed by metabolic stoichiometry, thermodynamics, or the structure of ecological interaction networks. We argue that integrating such constraints can significantly improve the predictive resolution of models, helping us build a much needed bridge between theory and data. We further discuss how novel statistical approaches are revealing simple, low-dimensional patterns in microbial communities, offering empirical clues for identifying the underlying constraints. Together, these developments suggest a path toward a data-driven and mechanistically informed theory in microbial ecology.

Address
Section of Industrial Microbiology, Department of Biotechnology, Delft University of Technology, Delft, the Netherlands

Corresponding author: Bajić, Djordje (d.bajic@tudelft.nl)

Current Opinion in Microbiology 2025, **88**:102658

This review comes from a themed issue on **Microbiota**

Edited by **Akos T Kovacs**

Available online xxxx

https://doi.org/10.1016/j.mib.2025.102658

Introduction

Life is essentially shaped by stochastic processes. Evolution is, to a large extent, driven by random genetic drift [1], and adaptive forces are fueled by random mutations. As the movement of molecules inside cells is essentially Brownian, even the processes that biologists often treat as deterministic, such as enzyme catalysis or gene regulation, are fundamentally underpinned by chance. Transcription factors routinely bind nonspecific sites [2], and most enzymes can carry out side reactions in addition to the primary one, which can fuel the evolution of metabolic innovations [3]. At the same time, organisms are highly ordered, evolved systems, in which mechanism and structure impose strong constraints to randomness [4]. Constraints may be fundamental, arising from physical or chemical laws (e.g. stoichiometric balance or thermodynamic feasibility); physiological, stemming from organismal traits and trade-offs (e.g. limits on metabolic fluxes or maximum growth rates); or structural, emerging from from the topology of interactions between system components, or their spatial connectivity. Each of these constraints imposes bounds to the biological processes occurring at other scales, irrespective of the specific parameters governing each of these processes. For example, the architecture of regulatory networks 'canalizes' the development of multicellular organisms, turning a collection of purely stochastic processes into a beautiful sequence of events that appears as fully deterministic, and allowing organisms to reach consistent phenotypes despite variation in genetic or environmental parameters [5]. In other cases, the room left for more diverse outcomes seems much more ample [6]. A case in point is microbial metabolism, where the sampling of possible flux spaces bounded by stoichiometric constraints has helped understand interstrain variability [7], or explain the diversity of metabolites that might be secreted across species [8,9].

How do biological systems navigate this space of possibility? Fueled by adaptive considerations, many biological processes are often understood within an optimization framework. Commonly, the emphasis is put on a quantity being optimized, most typically fitness, and comparatively less attention is given to the space where this optimization process happens. However, a growing body of work recognizes that many biological phenomena appear rather suboptimal. For example, one might think that enzymes would evolve to be as efficient as possible. Yet, when looking at the data, we see that

most of them are rather far from optimal [10]. Microbial metabolism at large also appears to behave often suboptimally [11,12], and even conflicting optimization objectives can allow space for stochasticity to shape the diversification of metabolic strategies [13]. These varied examples show that it is useful to consider the complementary (rather than opposed) perspective that extends beyond optimal states to consider all possible ones. In order to explore these spaces of biological possibility, it becomes critical to chart specifically what constraints delimit them, and how.

A field in which this paradigm, relying on stochasticity-within-constraints, has been historically most successful is ecology. In contrast to single organisms, ecological communities typically do not evolve adaptively (unless structured by the environment in specific ways [14]). Freed from the yoke of adaptive explanations, and to an extent inspired by the dynamics of ecological communities themselves, some of the most influential models in ecological theory have historically embraced stochastic exploration methods. From May's seminal work exploring stability in randomly connected communities [15,16] to Hubbell's neutral theory [17], many theoretical approaches begin by assuming that species differ randomly, or not at all, in their traits or interactions. The success of these approaches in ecology lies in their ability to explain many observed patterns, from species abundance distributions to ecosystem stability, while relying on minimal constraints. But even in Hubbell's neutral framework, where all species are genetically and phenotypically equivalent, the predicted ecological patterns are shaped by boundary conditions, such as spatial structure and limited dispersal. In the neutral model, the restrictions in the spatial scale and connectivity over which species disperse effectively constrain the possible patterns that can be observed at larger scales, such as community assembly (e.g. species abundance distributions) [17]. Importantly, ecosystem structure and function can also emerge from internal constraints. For instance, in randomly assembled food webs, the mere requirement of a trophic hierarchy (e.g. who eats whom) imposes a minimal structure that channels stochastic dynamics, allowing us to move to more mechanistic and precise predictions about ecosystem stability, dynamics, and function [18].

In many ways, microbial communities offer a more complex landscape than macro-organism ecology. First, microbial communities are structured in typically much more diffuse trophic networks [19,20], with many interactions happening within the same trophic level, for example, through competition and exchange of resources [20,21]. Their interactions are also more dynamic: in contrast to the fox, who will always eat the hare, and never the other way around, microbial interactions can change their strength or even their sign depending on the environment or on the presence of other community members [22,23]. Microbial interaction networks are also extremely large and complex [24,25], suggesting they might be unstable [16,19]. Yet, despite this complexity, evidence is accumulating that microbial communities, from the lab to natural habitats, exhibit remarkably simple and reproducible emergent patterns. For example, the composition of many natural and synthetic microbial communities seems predictable at the functional level, despite a much larger taxonomic complexity [26,27]. The composition of microbial communities also follows macroecological patterns of abundance that can be described using simple statistical laws [28–30]. More generally, evidence is mounting that both the structure and the function of microbial communities reveals signatures of low-dimensional structures [31–33], reflected in simple patterns that can be leveraged for prediction [34,35]. For example, the addition of a species to a community very often relates in linear terms to the function of the native community [36]. This leads to a number of key questions: what mechanistic constraints structure the complexity and the randomness intrinsic to microbial communities? And how do we incorporate these constraints into models that can meaningfully inform theory, allowing us to potentially understand the origin of these patterns?

Recent years are seeing increased efforts to elucidate how specific mechanistic or evolutionary constraints drive the assembly, dynamics, and function of microbial communities. For example, phylogenetic [37], metabolic [38–40], and thermodynamic constraints [41–43] have been argued as essential ingredients to structure both the evolution of microbial traits and the resulting patterns and processes observed in microbial ecology. Here, we argue that models exploring the space of biological possibility within mechanistically explicit, data-driven constraints constitute a powerful tool for advancing toward a predictive theory in microbial ecology. We also review data-driven approaches that aim to identify which constraints govern the assembly of microbiomes, their function, and their response to environmental change.

From constraints to patterns

In the last decades, the advances in diverse 'omic' techniques (from genomics to proteomics) have produced an immense amount of data mapping the structure and function of microbial communities in the most diverse settings. Yet many models in ecology still lack an explicit connection to data. For example, we have extensive evidence that the intricate architecture of metabolic networks profoundly affects resource use strategies [44,45]. Yet when applied to microbes, MacArthur consumer-resource models only incorporate extremely coarse-grained mechanistic constraints in terms of metabolic structure [46,47]. In contrast, trait-based models

rely on large trait databases (often genome inferred) to build phenomenological models that predict ecosystem-level responses [48–50]. Although this approach has been instrumental in generating detailed insight into many ecosystems [48,49,51], trait-based models are by design system tailored, making their generalization challenging. Now, the accumulation of high-resolution microbiome data presents an opportunity to develop models that are mechanistically explicit yet retain the level of generality needed to uncover fundamental principles in microbial ecology and evolution.

The use of empirically calibrated models to explore the space of biological possibility is almost as old as our ability to acquire enough data to fuel it. One of the first precedents comes from the seminal work of Fontana and Schuster, who in the early 1990s systematically mapped the relationship between RNA sequences and their folded secondary structures [52]. The ability of these models to explicitly chart the structure of an empirical genotype-phenotype map revealed that, far from being evenly accessible, the space of possible RNA phenotypes is structured in a highly nonrandom manner. Some folds were vastly overrepresented, others nearly inaccessible, and genotype space was comprised of large 'neutral networks' containing many sequences coding for the same phenotype. This seminal work offered a powerful demonstration of how constraints imposed by physics and genetics can lead to predictable evolutionary outcomes [52,53].

In parallel to RNA folding models, the turn of the 21st century also saw the development of an altogether different class of empirically calibrated models, namely, genome-scale metabolic models (GEMs) [54]. These models connect the genome-derived network of biochemical reactions an organism can carry out to the physiological performance of the organism, for example, growth capabilities or metabolite production. GEMs are particularly attractive because they make relatively few assumptions: they do not require detailed knowledge of kinetic parameters and instead rely on mass balance constraints, thermodynamic feasibility, and optimization principles justified from the evolutionary standpoint. In their most typical use, GEMs are used to reconstruct the metabolic networks of existing species as accurately as possible in order to make inferences about the metabolic traits of these species [55]. However, as GEMs inherently encode constraints both from metabolic architecture and environmental inputs, they can serve naturally as an 'empirical template' for studying how biological function is shaped by these constraints.

Inspired by work on RNA structure, pioneering approaches started applying stochastic sampling methods on GEMs to explore the space of the metabolically possible [56–59]. The general approach consists of repeated random 'Monte-Carlo' sampling of a universal metabolic space, comprising all known reactions available to microbes under a given set of environmental constraints. Metabolic networks are usually sampled under a set of constraints, for example, any sampled metabolic network is required to be viable. Then, the properties of the resulting 'ensemble' of sampled networks can be systematically analyzed. This approach has helped characterize the architecture of epistatic interactions [57], uncover neutral networks in metabolism [56], evaluate how predictable is the evolution of minimal genomes [58], and learn that metabolic networks viable on one substrate are typically and almost automatically also viable on multiple others [40]. The application of statistical approaches on GEMs has also shed light on the long-term patterns of phenotypic evolution in bacteria [60] and even allowed us to evaluate early-life evolution scenarios [61–63], fields where experimentation is hardly accessible. Altogether, this approach showed that structural constraints imposed by metabolic architecture have profound effects on evolution at different timescales and even in the absence of direct selection and can inform us about past evolutionary patterns.

Can a similar approach also help us understand ecology? Because of the fundamental role of metabolic traits shaping microbial communities, an increasing number of studies are leveraging statistical approaches in combination with GEMs to explore how metabolic constraints shape microbial ecology [64,65]. For example, recent work has discovered that metabolism harbors a vast potential for costless secretion of metabolites, even under strong optimality assumptions [8,9,66]. Besides the substantial implications for our understanding of microbial interactions and community assembly [67], metabolic secretions can also impact eco-evolutionary interactions among microbes, affecting how microorganisms evolve in the long run [68]. Statistical approaches have also been developed to quantify biosynthetic capabilities under environmental and genomic uncertainty, allowing for probabilistic estimates of cross-feeding potential [69], as well as algorithms to reverse-infer the metabolic environments of microbiomes from metagenomic data [70]. The application of GEMs in microbial ecology also has important limitations. For example, some interaction mechanisms are by definition excluded, for example, toxin secretion, or extracellular enzyme production. However, opportunities exist to implement these mechanisms ad-hoc, for example, in dynamic metabolic simulations, thereby evaluating their potential effect on microbial community properties [71]. Overall, the use of GEMs in microbial ecology has the potential to highlight how mechanistic constraints at the individual level aggregate to shape community- and ecosystem-level properties.

Currently, the predictions of GEMs are being increasingly refined by the application of machine learning

techniques. For example, the use of sophisticated statistical approaches is enabling more accurate reconstruction of metabolic networks [72] or the prediction of kinetic constants from an enzyme sequence. This could bring kinetic GEMs within reach, potentially allowing us to extend our models from stoichiometric to kinetic constraints [73,74]. But the broader lesson extends well beyond metabolism: constraints, whether physical, ecological, or evolutionary, both delimit and enable biological function [75]. Models that explicitly incorporate these constraints will be critical to mechanistically connect the increasing wealth of biological data to theory in microbial ecology [65].

From patterns to constraints

While the approaches discussed above represent opportunities to incorporate data-driven constraints into theory, the question of how to identify those constraints in a rigorous way remains challenging. The reason is the complex entanglement of constraints of different mechanistic origin. For example, stoichiometric and thermodynamic constraints mutually influence each other [76]; these shape cross-feeding and thus the structure of the ecological interaction network [77] and are shaped by environmental change [22,23], genetic architecture [38,68], and evolution. In a complex system such as a microbial community, how do we identify which constraints are important to model?

Encouragingly, an increasing number of studies are revealing that, despite their high dimensionality, the structure and behavior of complex biological systems often fold onto a surprisingly low number of dimensions [32,36,78,79]. Identifying these dimensions and linking them to mechanistic variables present an opportunity to identify the constrains governing these systems. In microbial ecology, multiple independent studies have uncovered that diversity and complexity often exhibit simple emergent patterns [50]. For example, at a functional level, the outcome of microbial community assembly is reproducible [26,47], which can be traced to mechanistic constraints, for example, cellular resource allocation and rate-yield trade-offs [27], and even genomic features [80]. The outcome of microbial community assembly processes seems to be statistically predictable, suggesting it might depend only on a small number of degrees of freedom [79]. Moreover, it is becoming increasingly clear that many community functions are largely predictable using surprisingly simple statistical models [34,79,81], likely because they are governed by one or a few dominant constraints, for example, pH [81,82]. Recent findings also show that the functional effect of adding a species to a community often depends in a simple, linear manner on the function of the receiving community [36]. An analogous phenomenon can be found across many biological systems [32] and is usually attributed to the presence of an underlying latent constraint [83,84], but can also emerge from the statistical structure of underlying interactions [85,86]. Increasing efforts are being made to both find new quantitative frameworks to characterize these patterns [32] and map them to mechanistic constraints, such as metabolic budgets [87] or specific pathway architectures [86].

Altogether, these simple emergent patterns delineate a compelling path from data to theory [65]. By characterizing the low-dimensional manifolds that capture most of the variation in a system, we might be able to infer the key mechanistic constraints that structure it. Incorporating these constraints to models provides an opportunity to advance toward a general predictive theory in microbial ecology that is grounded in empirical data and biological function.

Conclusion

In recent years, there has been a growing call to build a new theoretical framework that can explain, unify, and make predictions in microbial ecology [4,88–91]. For instance, it has been argued that we need to develop a new 'statistical physics of life', established not around traditional symmetries and equilibrium assumptions that are applicable in physics, but around concepts such as 'typicality with constraints' [4]. In this view, sufficiently complex biological systems may display statistical regularities typical of random systems, provided that they are bounded by similar constraints. However, while powerful, this framework remains deliberately agnostic to the origin and nature of the constraints themselves [4]. Although mechanistic agnosticism can be a powerful advantage for theory, constraints in biological systems are generally more complex than in physics [92]. Thus, if our goal is to understand why we observe a pattern, how it will respond to a perturbation, or how we can engineer it toward a specific state, we must engage explicitly with the mechanisms and structures that hold this pattern in place.

Funding

DB and MG would like to thank the Zero Emission Biotechnology Program, TU Delft.

Author contributions

Conceptualization: DB, MvO, MG, UG; Investigation: DB; Visualization: DB; Writing – original draft: DB; Writing: review and editing: DB, MvO, MG, UG.

Data Availability

No data were used for the research described in the article.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Pankaj Mehta, Jean Vila, Juan Diaz-Colunga, and Matti Gralka, and two anonymous reviewers for helpful feedback, and members of the Industrial Microbiology Section at TU Delft for enriching discussions.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Lynch M, Walsh B: **The Origins of Genome Architecture**. Oxford University Press; 2007.

2. Samee MAH: **Noncanonical binding of transcription factors: time to revisit specificity?** *Mol Biol Cell* 2023, **34**:e4.

3. Notebaart RA, Szappanos B, Kintses B, Pál F, Györkei Á, Bogos B, Lázár V, Spohn R, Csörgő B, Wagner A, *et al.*: **Network-level architecture and the evolutionary potential of underground metabolism**. *Proc Natl Acad Sci USA* 2014, **111**:11762-11767.

4. •• Mehta P: **A twenty-first century statistical physics of life**. *arXiv [Phys bioph]* 2024,2410.20506.
This essay discusses promising directions for theory development in microbial ecology and how they can capitalize on quantitative methods developed in other disciplines such as physics.

5. Bergman A, Siegal ML: **Evolutionary capacitance as a general feature of complex gene networks**. *Nature* 2003, **424**:549-552.

6. Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O, Mayo A, Dekel E, Kavanagh K, Alon U: **Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space**. *Science* 2012, **336**:1157-1160.

7. Scott WT Jr, Smid EJ, Block DE, Notebaart RA: **Metabolic flux sampling predicts strain-dependent differences related to aroma production among commercial wine yeasts**. *Micro Cell Fact* 2021, **20**:204.

8. Pacheco AR, Moel M, Segrè D: **Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems**. *Nat Commun* 2019, **10**:103.

9. San Roman M, Wagner A: **An enormous potential for niche construction through bacterial cross-feeding in a homogeneous environment**. *PLoS Comput Biol* 2018, **14**:e1006340.

10. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R: **The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters**. *Biochemistry* 2011, **50**:4402-4410.

11. Balakrishnan R, de Silva RT, Hwa T, Cremer J: **Suboptimal resource allocation in changing environments constrains response and growth in bacteria**. *Mol Syst Biol* 2021, **17**:e10597.

12. Towbin BD, Korem Y, Bren A, Doron S, Sorek R, Alon U: **Optimality and sub-optimality in a bacterial growth law**. *Nat Commun* 2017, **8**:14123.

13. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U: **Multidimensional optimality of microbial metabolism**. *Science* 2012, **336**:601-604.

14. Doulcier G, Lambert A, De Monte S, Rainey PB: **Eco-evolutionary dynamics of nested Darwinian populations and the emergence of community-level heredity**. *Elife* 2020, **9**:e53433.

15. Ives AR, Carpenter SR: **Stability and diversity of ecosystems**. *Science* 2007, **317**:58-62.

16. May RM: **Stability and complexity in model ecosystems**. *Monogr Popul Biol* 1973, **6**:1-235.

17. Hubbell SP: **The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)**. Princeton University Press; 2001.

18. Allesina S, Tang S: **Stability criteria for complex ecosystems**. *Nature* 2012, **483**:205-208.

19. Butler S, O'Dwyer JP: **Stability criteria for complex microbial communities**. *Nat Commun* 2018, **9**:2970.

20. Gralka M, Szabo R, Stocker R, Cordero OX: **Trophic interactions and the drivers of microbial community assembly**. *Curr Biol* 2020, **30**:R1176-R1188.

21. D'Souza G, Shitut S, Preussger D, Yousif G, Waschina S, Kost C: **Ecology and evolution of metabolic cross-feeding interactions in bacteria**. *Nat Prod Rep* 2018, **35**:455-488.

22. Meacock OJ, Mitri S: **Environment-organism feedbacks drive changes in ecological interactions**. *Ecol Lett* 2025, **28**:e70027.

23. Di Martino R, Picot A, Mitri S: **Oxidative stress changes interactions between 2 bacterial species from competitive to facilitative**. *PLoS Biol* 2024, **22**:e3002482.

24. Faust K, Raes J: **Microbial interactions: from networks to models**. *Nat Rev Microbiol* 2012, **10**:538-550.

25. Röttjers L, Faust K: **From hairballs to hypotheses-biological insights from microbial networks**. *FEMS Microbiol Rev* 2018, **42**:761-780.

26. Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DS, Parfrey LW, Farjalla VF, Doebeli M: **High taxonomic variability despite stable functional structure across microbial communities**. *Nat Ecol Evol* 2016, **1**:15.

27. Estrela S, Vila JCC, Lu N, Bajić D, Rebolleda-Gómez M, Chang C-Y, Goldford JE, Sanchez-Gorostiaga A, Sánchez Á: **Functional attractors in microbial community assembly**. *Cell Syst* 2022, **13**:29-42 e7..

28. Grilli J: **Macroecological laws describe variation and diversity in microbial communities**. *Nat Commun* 2020, **11**:4743.

29. Shoemaker WR, Sánchez Á, Grilli J: **Macroecological patterns in experimental microbial communities**. *PLoS Comput Biol* 2025, **21**:e1013044.

30. George AB, O'Dwyer J: **Universal abundance fluctuations across microbial communities, tropical forests, and urban populations**. *Proc Natl Acad Sci USA* 2023, **120**:e2215832120.

31. Frentz Z, Kuehn S, Leibler S: **Strongly deterministic population dynamics in closed microbial communities**. *Phys Rev X* 2015, **5**:041014.

32. •• Russo CJ, Husain K, Murugan A: **Soft modes as a predictive framework for low-dimensional biological systems across scales**. *Annu Rev Biophys* 2025, **54**:401-426, https://doi.org/10.1146/annurev-biophys-081624-030543.
Review showcasing how a dynamical systems framework based on the identification of so-called 'soft modes ' can be extremely useful to characterize emergent properties of disparate biological systems, including microbial communities.

33. • Plata G, Srinivasan K, Krishnamurthy M, Herron L, Dixit P: **Designing host-associated microbiomes using the consumer/resource model**. *mSystems* 2025, **10**:e0106824.
Original contribution combining machine learning and consumer-resource modeling to uncover which host phenotypes constrain the composition of host-associated microbiomes.

34. Skwara A, Gowda K, Yousef M, Diaz-Colunga J, Raman AS, Sanchez A, Tikhonov M, Kuehn S: **Statistically learning the functional landscape of microbial communities**. *Nat Ecol Evol* 2023, **7**:1823-1833.

35. Clark RL, Connors BM, Stevenson DM, Hromada SE, Hamilton JJ, Amador-Noguez D, Venturelli OS: **Design of synthetic human gut microbiome assembly and butyrate production**. *Nat Commun* 2021, **12**:3254.

36. • Diaz-Colunga J, Skwara A, Vila JCC, Bajic D, Sanchez A: **Global epistasis and the emergence of function in microbial consortia**. *Cell* 2024, **187**:3108-3119 e30.

This paper shows how the phenomenon of global epistasis, initially discovered in genetic systems, also extends to microbial community-function landscapes, and explains this phenomenon in terms of the interactions among community members.

37. Martiny JBH, Jones SE, Lennon JT, Martiny AC: **Microbiomes in light of traits: a phylogenetic perspective**. *Science* 2015, **350**:aac9323.

38. Takano S, Vila JCC, Miyazaki R, Sánchez Á, Bajic D: **The architecture of metabolic networks constrains the evolution of microbial resource hierarchies**. *Mol Biol Evol* 2023, **40**:msad187.

39. Vila JCC, Goldford J, Estrela S, Bajic D, Sanchez-Gorostiaga A, Damian-Serrano A, Lu N, Marsland R 3rd, Rebolleda-Gomez M, Mehta P, *et al.*: **Metabolic similarity and the predictability of microbial community assembly**. *bioRxiv* 2023, 10.25.564019.

40. Barve A, Wagner A: **A latent capacity for evolutionary innovation through exaptation in metabolic systems**. *Nature* 2013, **500**:203-206.

41. •• Goyal A, Flamholz AI, Petroff AP, Murugan A: **Closed ecosystems extract energy through self-organized nutrient cycles**. *Proc Natl Acad Sci USA* 2023, **120**:e2309387120.

This paper pioneers the incorporation of bioenergetic and thermodynamic constraints in consumer-resource models, showing how these constraints shape the efficiency and stability of self-organized communities.

42. George AB, Wang T, Maslov S: **Functional convergence in slow-growing microbial communities arises from thermodynamic constraints**. *ISME J* 2023, **17**:1482-1494.

43. González-Cabaleiro R, Martinez-Rabert E, Argiz L, van Kessel MA, Smith CJ: **A framework based on fundamental biochemical principles to engineer microbial community dynamics**. *Curr Opin Biotechnol* 2021, **67**:111-118.

44. Wang X, Xia K, Yang X, Tang C: **Growth strategy of microbes on mixed carbon sources**. *Nat Commun* 2019, **10**:1279.

45. de Groot DH, van Boxtel C, Planqué R, Bruggeman FJ, Teusink B: **The number of active metabolic pathways is bounded by the number of cellular constraints at maximal metabolic rates**. *PLoS Comput Biol* 2019, **15**:e1006858.

46. Dal Bello M, Lee H, Goyal A, Gore J: **Resource–diversity relationships in bacterial communities reflect the network structure of microbial metabolism**. *Nat Ecol Evol* 2021, **5**:1424-1434.

47. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, Segrè D, Mehta P, Sanchez A: **Emergent simplicity in microbial community assembly**. *Science* 2018, **361**:469-474.

48. Lajoie G, Kembel SW: **Making the most of trait-based approaches for microbial ecology**. *Trends Microbiol* 2019, **27**:814-823.

49. Litchman E, Edwards KF, Klausmeier CA: **Microbial resource utilization traits and trade-offs: implications for community structure, functioning, and biogeochemical impacts at present and in the future**. *Front Microbiol* 2015, **6**:254.

50. van den Berg NI, Machado D, Santos S, Rocha I, Chacón J, Harcombe W, Mitri S, Patil KR: **Ecological modelling approaches for predicting emergent properties in microbial communities**. *Nat Ecol Evol* 2022, **6**:855-865.

51. Marschmann GL, Tang J, Zhalnina K, Karaoz U, Cho H, Le B, Pett-Ridge J, Brodie EL: **Predictions of rhizosphere microbiome dynamics with a genome-informed and trait-based energy budget model**. *Nat Microbiol* 2024, **9**:421-433.

52. Fontana W, Schuster P: **Shaping space: the possible and the attainable in RNA genotype-phenotype mapping**. *J Theor Biol* 1998, **194**:491-515.

53. Schluter D: **Adaptive radiation along genetic lines of least resistance**. *Evolution* 1996, **50**:1766-1774.

54. Edwards JS, Palsson BO: **Systems properties of the Haemophilus influenzae Rd metabolic genotype**. *J Biol Chem* 1999, **274**:17410-17416.

55. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdóttir HS, Wachowiak J, Keating SM, Vlasov V, *et al.*: **Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0**. *Nat Protoc* 2019, **14**:639-702.

56. Samal A, Matias Rodrigues JF, Jost J, Martin OC, Wagner A: **Genotype networks in metabolic reaction spaces**. *BMC Syst Biol* 2010, **4**:30.

57. Segrè D, Deluna A, Church GM, Kishony R: **Modular epistasis in yeast metabolism**. *Nat Genet* 2005, **37**:77-83.

58. Pál C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD: **Chance and necessity in the evolution of minimal metabolic networks**. *Nature* 2006, **440**:667-670.

59. Matias Rodrigues JF, Wagner A: **Evolutionary plasticity and innovations in complex metabolic reaction networks**. *PLoS Comput Biol* 2009, **5**:e1000613.

60. Plata G, Henry CS, Vitkup D: **Long-term phenotypic evolution of bacteria**. *Nature* 2015, **517**:369-372.

61. Goldford JE, Hartman H, Smith TF, Segrè D: **Remnants of an ancient metabolism without phosphate**. *Cell* 2017, **168**:1126-1134 e9..

62. Goldford JE, Hartman H, Marsland R 3rd, Segrè D: **Environmental boundary conditions for the origin of life converge to an organo-sulfur metabolism**. *Nat Ecol Evol* 2019, **3**:1715-1724.

63. Ralser M, Varma SJ, Notebaart RA: **The evolution of the metabolic network over long timelines**. *Curr Opin Syst Biol* 2021, **28**:100402.

64. Quinn-Bohmann N, Wilmanski T, Sarmiento KR, Levy L, Lampe JW, Gurry T, Rappaport N, Ostrem EM, Venturelli OS, Diener C, *et al.*: **Microbial community-scale metabolic modelling predicts personalized short-chain fatty acid production profiles in the human gut**. *Nat Microbiol* 2024, **9**:1700-1712.

65. Qian Y, Lan F, Venturelli OS: **Towards a deeper understanding of microbial communities: integrating experimental data with dynamic models**. *Curr Opin Microbiol* 2021, **62**:84-92.

66. Waschina S, D'Souza G, Kost C, Kaleta C: **Metabolic network architecture and carbon source determine metabolite production costs**. *FEBS J* 2016, **283**:2149-2163.

67. Kost C, Patil KR, Friedman J, Garcia SL, Ralser M: **Metabolic exchanges are ubiquitous in natural microbial communities**. *Nat Microbiol* 2023, **8**:2244-2252.

68. Bajić D, Vila JCC, Blount ZD, Sánchez A: **On the deformability of an empirical fitness landscape by microbial evolution**. *Proc Natl Acad Sci USA* 2018, **115**:11286-11291.

69. Bernstein DB, Dewhirst FE, Segrè D: **Metabolic network percolation quantifies biosynthetic capabilities across the human oral microbiome**. *Elife* 2019, **8**:e39733.

70. Garza DR, van Verk MC, Huynen MA, Dutilh BE: **Towards predicting the environmental metabolome from metagenomics with a mechanistic model**. *Nat Microbiol* 2018, **3**:456-460.

71. Quintin M, Dukovski I, Bhatnagar J, Segrè D: **Optimality of extracellular enzyme production and activity in dynamic flux balance modeling**. *bioRxiv* 2021, https://doi.org/10.1101/2021.11.01.466736 11.01.466736.

72. Boer MD, Melkonian C, Zafeiropoulos H, Haas AF, Garza DR, Dutilh BE: **Improving genome-scale metabolic models of incomplete genomes with deep learning**. *iScience* 2024, **27**:111349.

73. • Li F, Yuan L, Lu H, Li G, Chen Y, Engqvist MKM, Kerkhoven EJ, Nielsen J: **Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction**. *Nat Catal* 2022, **5**:662-672.

This article shows that it is possible to predict catalytic constants from protein sequence using machine learning, opening the path to including more realistic constraints in GEMs.

74. Choudhury S, Narayanan B, Moret M, Hatzimanikatis V, Miskovic L: **Generative machine learning produces kinetic models that**

accurately characterize intracellular metabolic states. *Nat Catal* 2024, **7**:1086-1098.

75. Price ND, Reed JL, Palsson BØ: **Genome-scale models of microbial cells: evaluating the consequences of constraints**. *Nat Rev Microbiol* 2004, **2**:886-897.

76. Großkopf T, Soyer OS: **Microbial diversity arising from thermodynamic constraints**. *ISME J* 2016, **10**:2725-2733.

77. Giri S, Oña L, Waschina S, Shitut S, Yousif G, Kaleta C, Kost C: **Metabolic dissimilarity determines the establishment of cross-feeding interactions in bacteria**. *Curr Biol* 2021, **31**:5547-5557 e6.

78. Goyal A, Rocks JW, Mehta P: **Universal niche geometry governs the response of ecosystems to environmental perturbations**. *PRX Life* 2025, **3**:013010.

79. Arya S, George AB, O'Dwyer JP: **Sparsity of higher-order landscape interactions enables learning and prediction for microbiomes**. *Proc Natl Acad Sci USA* 2023, **120**:e2307313120.

80. •• Gralka M, Pollak S, Cordero OX: **Genome content predicts the carbon catabolic preferences of heterotrophic bacteria**. *Nat Microbiol* 2023, **8**:1799-1808.
This paper shows that microbial catabolic traits, central to their ecology, groups along very few dimensions, defined by the architecture of metabolic networks, further constraining the evolution of carbon-nitrogen requirements.

81. Moran J, Tikhonov M: **Emergent predictability in microbial ecosystems**. *bioRxiv* 2024, https://doi.org/10.1101/2024.03.26.586882 03.26.586882.

82. Lee KK, Liu S, Crocker K, Huggins DR, Tikhonov M, Mani M, Kuehn S: **Functional phases encode the response of the soil microbiome to environmental change**. *bioRxiv* 2024, https://doi.org/10.1101/2024.03.15.584851 03.15.584851.

83. Otwinowski J, McCandlish DM, Plotkin JB: **Inferring the shape of global epistasis**. *Proc Natl Acad Sci USA* 2018, **115**:E7550-E7558.

84. Husain K, Murugan A: **Physical constraints on epistasis**. *Mol Biol Evol* 2020, **37**:2865-2874.

85. Diaz-Colunga J, Skwara A, Gowda K, Diaz-Uriarte R, Tikhonov M, Bajic D, Sanchez A: **Global epistasis on fitness landscapes**. *Philos Trans R Soc Lond B Biol Sci* 2023, **378**:20220053.

86. Reddy G, Desai MM: **Global epistasis emerges from a generic model of a complex trait**. *Elife* 2021, **10**:e64740.

87. • Venkataraman P, Saini S: **On the metabolic basis and predictability of global epistasis**. *bioRxiv* 2025, https://doi.org/10.1101/2025.02.04.636564 02.04.636564.
This paper characterizes the metabolic underpinnings of global epistasis, an emergent phenomenon that has also been recently extended to microbial communities.

88. Blaser MJ, Cardon ZG, Cho MK, Dangl JL, Donohue TJ, Green JL, Knight R, Maxon ME, Northen TR, Pollard KS, *et al*.: **Toward a predictive understanding of Earth's microbiomes to address 21st century challenges**. *MBio* 2016, **7**, https://doi.org/10.1128/mbio.00714-16

89. Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, Cordero OX, Brown SP, Momeni B, Shou W, *et al*.: **Challenges in microbial ecology: building predictive understanding of community function and dynamics**. *ISME J* 2016, **10**:2557-2568.

90. Gralka M: **Searching for principles of microbial ecology across levels of biological organization**. *Integr Comp Biol* 2023, **63**:1520-1531.

91. Arya S, George AB, O'Dwyer JP: **The architecture of theory and data in microbiome design: towards an S-matrix for microbiomes**. 2024,

92. Krakauer DC, Collins JP, Erwin D, Flack JC, Fontana W, Laubichler MD, Prohaska SJ, West GB, Stadler PF: **The challenges and scope of theoretical biology**. *J Theor Biol* 2011, **276**:269-276.