CellPress

Cell Systems

Review

The community-function landscape of microbial consortia

Alvaro Sanchez,[1,2,5,6,7,*] Djordje Bajic,[1,5] Juan Diaz-Colunga,[1,5] Abigail Skwara,[1,5] Jean C.C. Vila,[1,5] and Seppe Kuehn[3,4,5,6]

[1]Department of Ecology & Evolutionary Biology & Microbial Sciences Institute, Yale University, New Haven, CT, USA
[2]Department of Microbial Biotechnology, CNB-CSIC, Campus de Cantoblanco, Madrid, Spain
[3]Center for the Physics of Evolving Systems, The Unviersity of Chicago, Chicago, IL, USA
[4]Department of Ecology and Evolution, The University of Chicago, Chicago, IL, USA
[5]These authors contributed equally
[6]Senior author
[7]Lead contact
*Correspondence: alvaro.sanchez@cnb.csic.es
https://doi.org/10.1016/j.cels.2022.12.011

SUMMARY

Quantitatively linking the composition and function of microbial communities is a major aspiration of microbial ecology. Microbial community functions emerge from a complex web of molecular interactions between cells, which give rise to population-level interactions among strains and species. Incorporating this complexity into predictive models is highly challenging. Inspired by a similar problem in genetics of predicting quantitative phenotypes from genotypes, an ecological community-function (or structure-function) landscape could be defined that maps community composition and function. In this piece, we present an overview of our current understanding of these community landscapes, their uses, limitations, and open questions. We argue that exploiting the parallels between both landscapes could bring powerful predictive methodologies from evolution and genetics into ecology, providing a boost to our ability to engineer and optimize microbial consortia.

INTRODUCTION

Microorganisms have colonized every habitat on earth, forming complex and diverse ecosystems that play critical roles throughout the biosphere. Besides their many environmental roles, microbial communities have also been harnessed for biotechnological applications at least since the dawn of the neolithic revolution. The biotechnological applications of microbial consortia are growing from their traditional roles in food and drink[1–4] to contemporary uses in biofuel production,[5–8] the valorization of discarded plant materials,[9–11] bioremediation,[12–14] crop fertilization, [15,16] and many more.[17–20]

Relative to monocultures, microbial communities offer multiple advantages in biotechnology. Among these, they permit specialization and division of labor[17,21] avoiding physiological and cellular tradeoffs and other constraints that limit the efficiency of many biochemical processes. Communities may also contain much more genetic diversity than one would find in a single organism due to genome size limitations (e.g., Mizrahi et al.[22]). This diversity can enable communities to remain resilient to perturbations that single strains might not survive.[23] Finally, microbial consortia form spontaneously through evolutionary and ecological processes that are very difficult to avoid, even when a monoculture is started from a single isogenic population and propagated under otherwise sterile laboratory conditions.[24–26] Even in environments supplied with a single limiting resource, diversity and coexistence always seem to find a way, suggesting that a community is the natural endpoint of microbial systems both in natural and synthetic conditions.[27–31] Learning how to manipulate and engineer microbial consortia is therefore critical to realizing the biotechnological potential of microorganisms.

Despite growing interest, our ability to engineer microbial consortia lags behind bioengineering efforts in other biological systems at or below the organismal level, such as proteins[32,33] or metabolic and genetic networks.[34,35] One major reason is the nested hierarchical complexity present in a consortium. Specifically, the collective properties and services provided by microbial consortia (i.e., their "functions") emerge from the contributions of individual community members and their interactions with one another and their environment. The physiological traits of individual taxa dictate interactions, and these traits depend on genomic diversity, regulatory variation, and life history. Community functions then emerge from the collective action of these interactions, which are often non-linear and historically contingent. This means that parsing the mapping from structure to function from a detailed accounting of each process in the community is an immense task even for relatively simple consortia. Amidst this complexity, how are we to approach the problem of community design and control?

MAPPING COMMUNITY COMPOSITION TO FUNCTION CAN DRAW INSPIRATION FROM PROTEIN ENGINEERING

The field of molecular engineering has very similar goals and has encountered similar challenges. For instance, protein

engineers seek to design enzymes with desirable catalytic activities.[36–38] The catalytic rate of an enzyme is encoded in its sequence of amino acids, and it is also a collective property of the enzyme that arises from a large number of local and long-range biophysical interactions between its amino acids. These interactions give rise to the folded structure of the enzyme and govern its stability and intermolecular dynamics. Engineering every possible amino acid interaction to produce a desired enzymatic function is obviously daunting, but even the simpler task of predictively connecting sequence with function has been a major open challenge in biophysics. However, this has not precluded our ability to engineer and optimize enzymatic function.[33,36,38] In the process of understanding the connection between amino acid sequence and function, protein engineering has benefited greatly from insights provided by the theory of fitness landscapes.[39–41] Perhaps, the most successful example is the development of directed evolution, which has enabled the top-down engineering of different kinds of proteins.[33] Directed evolution involves an assisted exploration of the genotype-phenotype map in search for genotypes of desired or optimized functionality (this map is often referred to as the fitness landscape in the context of directed evolution where an objective function—i.e., "fitness" can be externally imposed). This assisted search is implemented through a process that mimics that of evolution by the iterative application of sequence randomization followed by selection on expressed phenotypes.[33]

In addition to the algorithmic explorations of fitness landscapes, a complementary approach has been to infer the principles of protein design by examining the statistics of sequence variation in naturally occurring proteins—effectively learning the landscape from extant variation. This approach has enabled the synthetic design of functional enzymes,[42] inferring folds[43] and insights into evolvability[44] and allostery.[45] One key insight from this body of work is that within the astronomically large space of possible protein sequences, natural functional proteins inhabit a much lower-dimensional subspace.[46] In proteins, this low dimensionality manifests as groups of co-evolving amino acids that are derived from principal components of a covariance matrix describing coevolution between sites in a multiple sequence alignment. In this case, low dimensionality arises because there are only a few independent groups of amino acids that co-evolve. Hence, although a protein has a huge number of degrees of freedom, there are emergent collective modes or "sectors" that faithfully describe protein function.[42] This result means that engineering proteins may not require an exhaustive search of sequence space (an impossible task) but instead a constrained search within a low-dimensional subspace. In communities, we speculate that similar emergent low-dimensional features might exist. In the community case, these features might emerge as groups of species or genotypes whose abundances covary across replicate communities or environments. For example, studies with synthetic communities[47,48] and human microbiota[49] have shown that variation in species abundances can be described by a few collective modes. Note again that a community has many degrees of freedom (species abundances, traits), but the low dimensionality might be an emergent property of the collective. The role of these collective modes in community function is yet to be elucidated, but exploring this emergent low dimensionality might dramatically reduce the size of the compositional space that must be explored to design functional consortia.

Can we extend the theory of fitness landscapes to study and engineer microbial community function? An important challenge is that, unlike molecular systems, microbial communities are made up of multiple self-replicating individual genotypes, each possessing its own fitness landscapes. It is therefore not immediately obvious how the idea of fitness landscapes may be extended to entire communities. In particular, any notion of fitness at the community level is not clearly defined, given the independent replication of genotypes rather than communities as a whole. Although this is true, community-level selection can be applied under artificial conditions, where an arbitrary fitness function can be applied.[50–52] More broadly, for a landscape to exist, it is not necessary that the scalar property that is being mapped to the composition of the community be defined in terms of fitness; it can instead be any collective function of the community.[53–55]

In recent years, a small but growing body of work has started to extend the theory of fitness landscapes to communities and suggested ways in which it may help us guide the design of microbial consortia.[52,56–59] Examples range from fruit-fly microbial consortia whose function is the host's lifespan and other life-history traits[60] to sugarcane biorefinery consortia whose function is the amount of ethanol produced during a single-batch fermentation.[8] These and other studies[56–58,61–67] have formally defined the structure-function (or composition-function or community-function) landscape as the empirical map between community composition and function in a given habitat and set of conditions. The structure of a microbial consortium is given by the list of all its genotypes $\boldsymbol{g} = \{g_1, g_2, \ldots, g_n\}$ and their respective abundances $\boldsymbol{x_g} = \{x_1, x_2, \ldots, x_n\}$. If a molecular fitness landscape is a map between a genotype g (where g represents the DNA sequence of the molecule) and a quantitative phenotype P (i.e., P(g)), a community structure-function landscape can be conceptualized as the map between the abundance vector $\boldsymbol{x_g}$ and a collective function F of the consortium $F(\boldsymbol{x_g})$.

To make this concept useful and productive, it is critical that we identify and understand the similarities and differences between structure-function landscapes and molecular fitness landscapes. The goal of this study is to synthesize our current understanding of the community structure-function landscape, highlighting promising directions and open questions. We start by drawing parallels between genetic interactions (epistasis) in simple genetic landscapes, and their ecological analogs in simple structure-function landscapes. We then discuss how various concepts from fitness landscape theory may be generalized to communities. Finally, we discuss under what conditions an ecological structure-function landscape is defined so that a collective property of interest can be said to depend uniquely on species composition. Our focus is eminently practical, and we focus on those ideas and methods from fitness landscape theory that, in addition to providing ecological insights, may help us guide our efforts to engineer and manage community services and functions. We also highlight how "landscape thinking"[68] may provide a helpful theoretical framework to help us

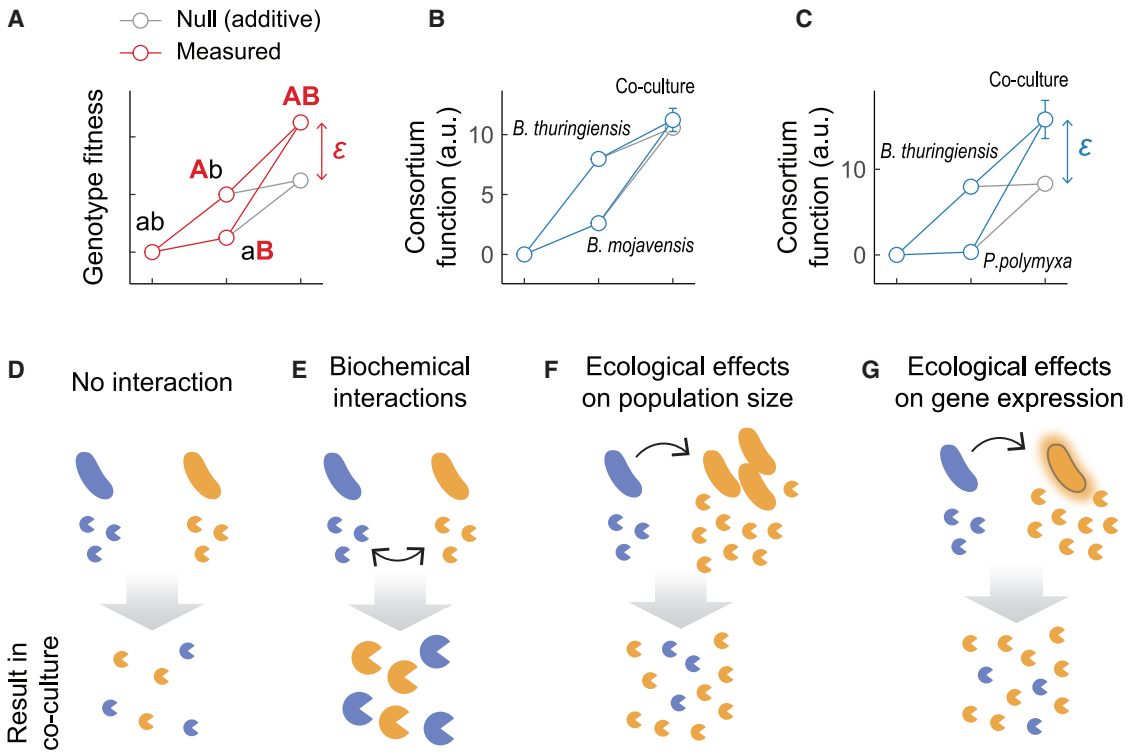

Figure 1. Species interactions create non-additive effects on community function

(A) In population genetics, two mutations A and B are said to interact when their phenotypic effects do not combine additively (or multiplicative, depending on the scale). This interaction is quantified by the deviation from additivity (referred to as the epistasis, ε).

(B–G) (B) Empirical measurements have found that the function of pairwise microbial co-cultures is often described by the sum of the functions in monoculture, as exemplified here by the amylolytic activities (in hr^{-1}) of monocultures and pairwise co-culture of *B. mojavensis* and *B. thuringiensis* (data from Sanchez-Gorostiaga et al.[56]). Other pairs, however, exhibit marked deviations. For instance, the pair formed by *B. thuringiensis* and *P. polymyxa* (C) has an amylolytic rate that far exceeds the expected value if both species acted independently. Three different types of interactions may cause this deviation from the situation where species functional contributions are additive (D). For instance, the enzymes and other molecules secreted by each species may interact with one another either enhancing or limiting their amylolytic activity (biochemical interactions, E). Alternatively, a species may promote (or suppress) the growth of its partner, limiting the size of its population and thus, potentially, its net expression of amylases (F). Finally, a population of one species may impact the per-capita expression of amylases by another, similarly impacting the net production of this function (G).

conceptualize the challenges associated with engineering microbial consortia.

A SIMPLE EXAMPLE OF LANDSCAPE THINKING IN COMMUNITY FUNCTION: AN ECOLOGICAL PARALLEL TO EPISTASIS

To develop our intuition of how fitness landscape theory may be extended to microbial communities, it is useful to start from the simplest scenario. The simplest genotype-phenotype map consists of two mutations, a → A and b → B, which define four possible genotypes: the "wild-type" (ab), the two single-mutants: (Ab and aB), and the double mutant (AB) (Figure 1A). One then needs a null model that describes how both mutations combine their effects when they act independently on the phenotype. Typically, the null model assumes that mutations act additively on the phenotype (or multiplicatively, depending on the scale). The deviation between the phenotype of the double mutant AB and its expected value under the null, interaction-free model, is known as the pairwise "epistasis" between those mutations (Figure 1A). Thus defined, epistasis gives us a metric of interactions between mutations.

Interactions can similarly be defined in other combinatorial systems that are not genetic, and in fact, the term epistasis has been used to describe systems as diverse as drug interactions[69,70] or combinations of stressors,[71] among others.[72] In recent years, we (and others) have extended it to ecological systems as well,[8,56–58,61,73,74] and the underlying idea was already present in earlier efforts to model the emergence of community function.[75,76]

In ecology, the simplest type of consortium is one containing just two different genotypes, g_1 and g_2. We could inoculate identical habitats with either cell from just one of those genotypes (g_1), the other (g_2), or both (g_1 and g_2), and measure a function of interest of each habitat after some defined incubation time. We could then establish a null model that would describe the function of the pairwise consortium if both species did not interact with one another in any way.[56] By analogy with the epistasis concept in genetics, the deviation between the function of the pairwise consortium and the expected value under the null model, which assumes no interactions, is defined as the functional interaction between both genotypes, an ecological equivalent of epistasis.

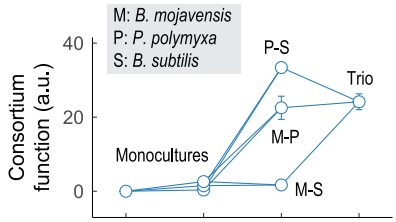

Figure 2. High-order functional interactions in microbial consortia
We show an example of a third-order interaction that shapes the function of microbial consortia, in this case leading to diminishing returns in the rate of starch degradation by the set of enzymes released by the cells in the community. The data was re-plotted from Sanchez-Gorostiaga et al.[56] Adding either *B. subtilis* (S) or *B. mojavensis* (M) to a monoculture of *P. polymyxa* (P) dramatically enhances its function through a pairwise functional interaction. However, when we add either *B. subtilis* or *B. mojavensis* to the co-culture of *P. polymyxa* with the other partner, their impact on function is either neutral or negative. This shows that the functional effect of adding a species to a consortium may be different when a second species is present, indicating the existence of a high-order functional interaction (HOFI).

To illustrate this idea, in Figures 1B and 1C, we present a recent empirical example of a simple structure-function landscape. In this example, drawn from Sanchez-Gorostiaga et al.,[56] the function of interest is the rate of starch degradation by extracellularly amylase enzymes secreted by different strains of the phylum Bacillota. Biochemical modeling tells us that these enzymes should combine additively, a point that was confirmed empirically.[56] Therefore, in the absence of any interactions, the amylolytic rate function of any consortium should be the sum of the functions of each genotype in monoculture. Indeed, many genotype pairs were very well described by this interaction-free model (e.g., as shown in Figure 1B, the one formed by *B. mojavensis* and *B. thuringiensis*). The (surprising) effectiveness of simple additive models has been reported in other systems, as a recent study showed similar success with an additive regression model for predicting fluxes of nitrate and nitrite through synthetic denitrifying communities.[77] Interestingly although, other genotype pairs in the starch degrading communities, deviated markedly from the additive model (Figure 1C), indicating the existence of strong, pairwise functional interactions between them. These interactions indicate the presence of epistasis-like interactions in these simple community-function landscapes.[56]

What is the mechanistic basis of these pairwise interactions? In general, functional interactions may arise from three different mechanisms (Figures 1D–1F).[56] First, the functional contributions of each community member may interact with each other. For instance, going back to the secreted enzyme example that is serving as an illustration, enzymes secreted by two species may act independently on the substrate, in which case their catalytic rates will be additive. However, some enzymes act synergistically on their substrate, as is the case of endo- and exo-cellulases: the former create new substrates for the latter, reaching an activity together that is higher than the sum of each of them separately.[78] The enzymes secreted by each species may also act antagonistically, for instance by aggregating (and therefore inhibiting) one another. These deviations from additivity may be called "abiotic" interactions, as they would occur even if no cells were present. The second type of interaction involves changes in the amount contributed by a given genotype to the community function. For instance, a genotype may either promote or inhibit the per-capita functional contribution by another genotype, altering its behavior. These "behavioral" interactions may include chemical signaling from one species that modifies the behavior of another.[79] Alternatively, a genotype may affect the growth (and therefore the total number of cells in the population) of another genotype. These "population" interactions can also alter the collective function of the ecosystem in a context-dependent manner. The three types of interactions summarized in Figures 1D–1F can be separated empirically.[56]

HIGH-ORDER FUNCTIONAL INTERACTIONS

In communities with more than two species, functional interactions may be more complex than pairwise.[8,56,74,79] Consider, for instance, the example provided in Figure 2, where the structure-function landscape comprising every combinatorial consortia of three amylolytic bacteria is given.[56] This landscape shows that co-culturing *P. polymyxa* with *B. mojavensis* or *B. subtilis* increases function beyond what we might expect from the additive model, indicating the presence of strong pairwise interactions. However, the beneficial effect of adding both *B. mojavensis* or *B. subtilis* to *P. polymyxa* is negligible, as there is no additional benefit of adding those strains. This "diminishing returns" effect indicates that the same genotype (e.g., *B. subtilis*) that is functionally "beneficial" when added to with *P. polymyxa* alone is functionally neutral when added to a consortium formed by *P. polymyxa* and *B. mojavensis*. The functional effect of adding a species to a consortium is thus different when two species, as opposed to one, are present. This would be the canonical definition of high-order epistasis if, instead of species and their functional effect, we were talking about mutations and their fitness effect.[73,80]

Besides the example discussed above, high-order functional interactions (HOFIs) have been observed in the production of ethanol by sugarcane biorefinery consortia,[8] the extension of a host lifespan by *Drosophila* gut microbiome consortia,[57] the metabolic activity of synthetic consortia,[74] and, more recently, gene expression in simple defined communities.[81] Just as they do in fitness landscapes, HOFIs could have profound implications for the topography of structure-function landscapes. For instance, in sugarcane biorefinery consortia, HOFIs have been found to tone down the predominantly negative effects of pairwise interactions between bacteria on the net ethanol yield.[8] Based on pairwise interactions alone, we would have expected that as bacterial biodiversity increases in our bioreactors the ethanol yield would have collapsed. However, the opposite was true and although most pairs of bacteria had negative effects on the ethanol yield, this detrimental effect vanished as communities increased in richness, reaching average levels that were comparable with those of pure yeast monocultures.[8] Despite this and other recent attempts to characterize HOFIs,[56,57,61] our understanding of the effect and implications of HOFIs is still very incomplete. When do they complicate and when do they simplify the navigability of structure-function landscapes? How do they affect the number and stability of functionally stable equilibria? These are still open questions, representing an open frontier in functional microbial ecology.

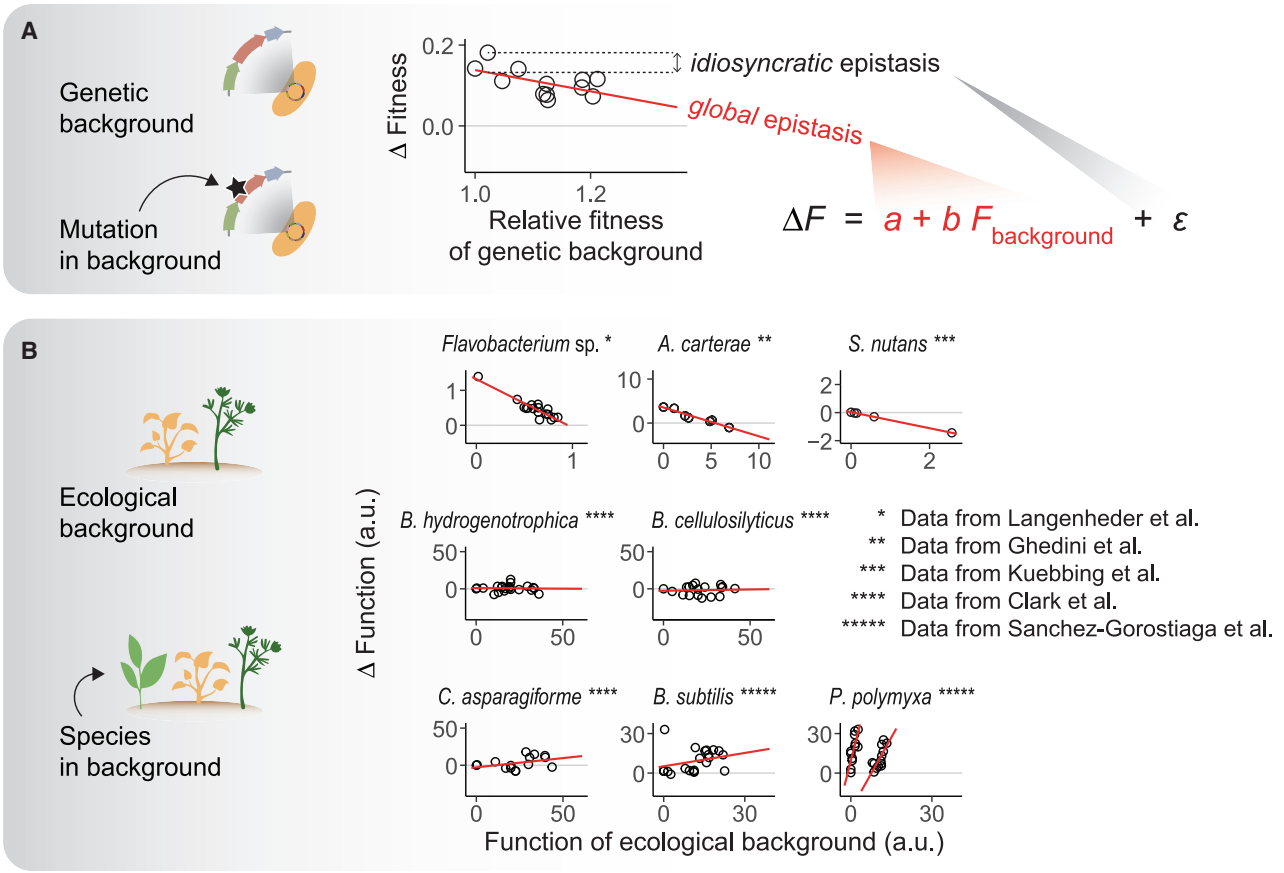

Figure 3. An analog to global epistasis explains the functional effect of adding new species to microbial consortia

(A) Research in quantitative genetics has shown that the fitness effect of a mutation is often well predicted by a simple linear regression on the fitness of the genetic background where it arises. Epistasis can thus be partitioned as the sum of a *global* component captured by such a linear fit (red), and an *idiosyncratic* component, not predictable from the fitness of the genetic background alone, represented by the residuals of that fit (green). Data from Khan et al.[83] Note that the success of a linear regression does not necessarily reflect that an underlying linear relationship exists, but rather that a linear model can explain a substantial fraction of the variance (see e.g., Reddy and Desai[82]).

(B) An ecological parallel to global epistasis can be formulated: the effect on ecosystem function resulting from the addition of a species to a community (an *ecological background*) can be well predicted via a linear regression by the function of the ecological background itself. Species can have less beneficial (or more deleterious) functional effects in backgrounds with higher functions (red lines), or vice versa (blue lines). These regressions that capture the functional effect of adding a species to a gamut of different consortia have been termed functional effect equations (FEEs).[84] In some cases, the functional effect of a species may be dominated by an idiosyncratic component rather than a global one (black lines). Data correspond to butyrate production by synthetic gut microbial communities,[59] biomass in plankton communities,[85] above-ground biomass in multi-species plant communities,[86] xylose oxidation by soil bacterial communities,[87] and amylase secretion in bacterial consortia.[56]

AN ECOLOGICAL PARALLEL TO GLOBAL EPISTASIS AND THE EMERGENCE OF SIMPLE FUNCTIONAL EFFECT EQUATIONS

Building predictive models of the structure-function landscape from the bottom up (by combining additive, pairwise, third-order interactions, etc.) is generally challenging. There is no guarantee that the complexity of interactions ends at the second or third order;[56] hence, the number of interactions that one would need to measure in order to build a predictive model of the landscape can blow up. An alternative is provided by defining global functional interactions in a way that is inspired by recent developments in quantitative genetics. Genetic interactions can be partitioned as the sum of a "global epistasis" effect, where the fitness effect of a mutation is predicted by the fitness of the genetic background and an "idiosyncratic epistasis" effect, which captures the part of the fitness effect of a mutation that depends on the genetic background while being independent of the background fitness[82] (Figure 3A).

Can we extend this way of partitioning interactions to microbial consortia? In recent work, we have found that the functional effects of adding a species to a consortium are indeed well predicted by linear regression on the function of the background consortium, similar to what has been observed in genetic systems.[84] The existence of these global functional interaction patterns appears to be rather general in ecosystems as we also found them in plant and algal communities.[84] Importantly, different species within a consortium tend to exhibit different quantitative relationships between their "functional effect" (i.e., the change in community function when they are present relative to when they are not) and the function of the background community to which we add them. We call these relationships the "functional effect equations"

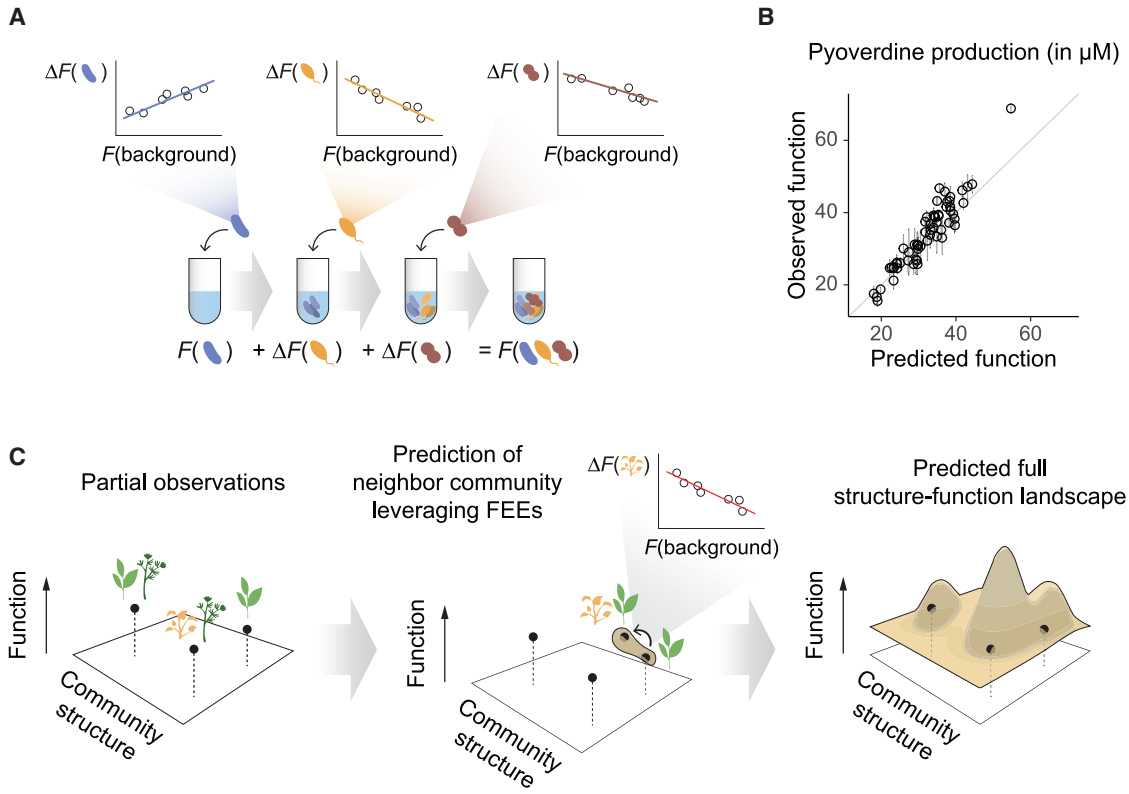

Figure 4. Global epistasis—like patterns in ecology—may help us infer the quantitative relationship between community composition and function

(A) It is possible to predict the function of any combinatorial consortia assembled from a set of species, as long as we know their functional effect equations (FEEs). For instance, a simple procedure consists of stitching together their respective FEEs, by iteratively applying them as depicted.[84]

(B) For instance, we have recently shown that this approach is capable of accurately predicting the amount of pyoverdine produced by newly assembled consortia if we had previously characterized the FEEs associated with this function for all of its constituent members.

(C) This is just a particular example of a broader research agenda, depicted here in cartoon fashion: bringing to ecology methods that have been developed in evolutionary genetics, and which have proven to be successful at predicting quantitative phenotypes from a smaller subset of measured genotypes.[92]

of a species.[84] Examples include diminishing returns, as well as increasing costs, accelerated returns, and other patterns (Figure 3B). How the particular global functional patterns exhibited by a species depend on its traits, as well as the traits of the species it interacts with, is still not well understood. In addition, it will be important to understand how this simple "global" epistasis emerges from the pairwise and potentially higher-order interactions in the consortia, extending and complementing the work that is currently being done to understand the origins of global epistasis in genetic fitness landscapes.[82,88–90]

THE USEFULNESS OF THE STRUCTURE-FUNCTION LANDSCAPE CONCEPT

An important consequence of the existence of these predictive functional effect equations is that they make it possible to predict with reasonable accuracy how adding a given species to a consortium will change its function. This illustrates what may be one of the most important benefits of bringing the concept of a structure-function landscape from genetics to ecological research: we could apply the arsenal of analytical and statistical tools that have been developed in genetics to infer and navigate these landscapes. For instance, several machine learning methodologies have been developed in recent years to infer a full genotype-phenotype landscape from a small subset of measured genotype-phenotype relationships. These methods have found impressive success in predicting biological function from DNA sequence under constant environmental conditions.[91–93] Adapting and applying these methodologies to microbial consortia is an exciting prospect,[59] and its feasibility is encouraged by the success of simpler inference approaches. For instance, we have recently tested the predictive power of a simple model consisting of "stitching together" the functional effect equations of all community members.[84] This very simple approach, summarized in Figure 4, does an excellent job at predicting various community functions for the full set of all possible consortia one may form with a defined set of taxa. Importantly, the ability to predict the full structure-function landscape makes it possible to identify the community compositions that will maximize and minimize these functions, paving the way to engineering community functions from the bottom up. The application of machine learning and neural networks to reconstruct community-function landscapes from a limited set of observations is still in its infancy. However, promising results are being published,[59] and the success of earlier regression-based approaches to predict the landscape of small consortia[75] is also an encouraging sign.

The landscape perspective allows one to approach the problem of community design from a statistical point of view. We propose that from this perspective, the complex hierarchy of processes discussed above that influence the structure-function landscape might yield simple descriptions. Indeed, our recent work suggests that taking this perspective can uncover simple rules for mapping genomes to phenotypes[77] and community composition to emergent function.[84] Despite these advances, we do not yet have a clear picture of the topography of these structure-function landscapes and this will be important if what we wish is to optimize communities using evolutionary engineering approaches.

THE TOPOGRAPHY AND NAVIGABILITY OF AN ECOLOGICAL STRUCTURE-FUNCTION LANDSCAPE

The topography of a fitness landscape gives us a measure of its navigability by either evolution or other assisted search processes. Smooth single-peak landscapes are navigated more easily than rugged ones since there are a larger number of adaptive paths connecting a given genotype to the global fitness peak.[94,95] Smoothness is high when different mutations act independently, whereas ruggedness increases in the presence of interactions between mutations (epistasis). In particular, strongly positive interactions between deleterious mutations (reciprocal sign epistasis) play a key role in determining landscape navigability as they are necessary for the presence of multiple fitness peaks.[96] In multi-peaked fitness landscapes, evolutionary algorithms can become trapped on local optima and fail to find the global fitness peak.

The simplest evolutionary algorithms used to navigate fitness landscapes involve an iterative two-step process consisting of a selection of the mutants of the highest fitness, followed by sequence randomization. These belong to the class of "hill-climbing" search algorithms, which work particularly well for smooth landscapes. Rugged fitness landscapes with many distinct peaks, on the other hand, are more challenging to search through a hill-climbing approach[63,97] because local information is not informative globally. By the same logic, the ruggedness of the ecological community-function landscape will also determine its navigability using analogous hill-climbing search algorithms, such as the directed evolution approaches reviewed in Sánchez et al.[52] For example, consider one configuration of a community that gives rise to a function that is locally a maximum, meaning that any small change in composition reduces function. In a rugged landscape, there will be many such optima, and understanding the structure (community composition) to function map at one peak will not in general be informative as to the structure-function map at another peak. This means that those genotypes whose changing relative abundances have the greatest impact on function can and will be distinct from one local optimum to another. In principle, any directed evolution algorithm may thus get stuck on a sub-optimal community and fail to find the optimal configuration of genotypes.

LEARNING THE LANDSCAPE

A complementary approach to directed evolution for exploring the structure-function landscape is to attempt to learn the landscape via either regression or more sophisticated machine learning methods. In this approach, one collects data on a large number of communities comprised of diverse genotypes and measures the function of interest. Learning the landscape then amounts to performing a regression with the following form: $y^i = F(\mathbf{x}_g^i)$ where F is a proposed functional form stipulated by the regression being used (e.g., linear model, random forest) and y^i is the measured function (degradation rate, pathogen inhibition, etc.) for the community with composition $\mathbf{x}_g^i$. Such an approach differs from a directed evolution approach because it posits a specific functional form for the structure-function landscape. This statistical approach faces the challenges of any inference problem, including overfitting and model misspecification.

Just as with the directed evolution approach, in a situation where the landscape is exceedingly rugged, the regression approach will face challenges because the contributions of each genotype to the function may depend strongly on the community composition. In this scenario, any local optimum may be well approximated by a model, but this model may dramatically fail to predict function[98] in the neighborhood of a different local optimum where the impact of adding or removing a given genotype may be very different and where the model has not been trained. Consider as an example—a set of species with a modest number of 50 genotypes. The full space of all possible communities comprising these genotypes is 2^{50} or 10^{15} possible communities. If a space of this size is truly rugged and contains many local optima, learning the structure-function map would require enumerating each optimum and the genotypes that impact function around it, one by one. Even for 50 genotypes, this is a daunting task that may be feasible in theory, but in practice, it is prohibitive, even computationally. It is therefore crucial to ask what controls the ruggedness of these landscapes and what is known about how rugged they might be.

THE NAVIGABILITY OF STRUCTURE-FUNCTION LANDSCAPES MAY BE CONNECTED WITH GLOBAL FUNCTIONAL EFFECTS

In simple models of landscapes, such as the well-known Kaufmann NK-mode,[99,100] the frequency of random epistatic (non-additive) interactions determines the ruggedness, with increasing epistasis driving more rugged landscapes. Critically, epistasis in the NK model is random, with any site in a genotype equally likely to have an epistatic interaction with any other site. In the community structure-function context, high levels of epistasis would be analogous to many random, strong interactions between genotypes that impact function non-additively. Given the small handful of cases where a structure-function landscape has been enumerated, we simply cannot say yet if this type of epistasis is prevalent in community structure-function landscapes. This remains an important open question that should be addressed in future work.

However, recent studies on landscapes in proteins have revealed that ruggedness is not a necessary outcome of many strong epistatic interactions. Instead, some proteins have strong epistasis and smooth landscapes. How can this be? In proteins, this occurs when a single "soft mode" dominates the physical dynamics of the system.[88] To understand what this means consider the normal modes of a protein, i.e., the coherent

motions of all atoms in the protein in response to a perturbation. These modes, or oscillations, have different stiffness that dictates how they respond when the system is perturbed. We can think of a soft mode as a specific set of coherent motions of all atoms in a protein that are soft—in this case, *any* perturbation to the protein causes the system to excite that mode.

Experimental studies of proteins with soft mechanical modes have shown that mutations cause physical deformations along that soft mode.[88,101] In essence, the protein can respond to any perturbation, be it physical or mutational, in only one way—along the soft mode. In the limit of small perturbations, any two perturbations simply add up to nudge the system along the soft mode. Thus, mutations are roughly additive in their impact on the physical locations of atoms in the protein. Epistasis is defined not in terms of the physical deformation of the protein, but instead as the impact of pairs of mutations on a function such as the catalytic activity or thermal stability. Both of these are complex functions of the physical locations of all atoms, so although the impact of each mutation on *physical* locations is roughly additive, their impacts on thermal stability or catalysis are epistatic.[89] However—and this is crucial—when a system possesses a soft mode, this strongly constrains the epistatic interactions between mutations in the system because the impacts of mutations are highly correlated.[88] Remarkably, the very same logic applies to gene regulatory networks. In this case, a network with a soft mode responds to diverse perturbations with a common change in the pattern of gene expression. In essence, the response of the regulatory network is constrained to be low-dimensional. Low-dimensional landscapes present in systems with soft modes are less rugged and facilitate more rapid evolution that does not get trapped in local optima.

Returning to community-function landscapes in microbial communities, if the functional interactions between genotypes are random, then we expect that the landscape will be hard to navigate and directed evolution or landscape learning methods will face challenges. However, what if the community structure-function landscape possesses a soft mode as described above? In the community context, what would this entail? One analogy to the protein example above could be to consider the abundance of genotypes as analogs to the physical locations of atoms in the protein. In this case, a soft mode would manifest as a coherent variation in abundance along, for example, a single dominant principle component. Perturbations to the community would then be constrained to drive abundance dynamics primarily along that mode. We note that such modes of variation have been observed in simple communities of a few species[47,48] and more recently also in host associate microbiomes.[49] In analogy to protein function, community function can and often is a non-linear function of abundances. In this case, the pattern of epistatic interactions between genotypes will be non-random and constrained. In this situation, we could expect a structure-function landscape that is not rugged but instead smooth, potentially learnable via regression and navigable by directed evolution.

We stress that the above sketch of how the theory of fitness landscapes in proteins or gene regulatory networks might map to communities is at present speculative. Our goal here is to propose plausible scenarios for what might control the ruggedness of these landscapes given the many insights provided by fitness landscape theory applied to proteins, gene circuits, and other biological systems defined at lower levels of biological organization. Finally, we wish to emphasize that the validity of the community landscape concept does not require that a community is a unit of selection, just as a protein genotype-phenotype map does not require that either. This is important because communities are not, in general, units of selection; hence, the idea that they may possess a fitness value may be confusing. Although this is true in most natural settings, there is mounting evidence that communities can respond to artificial community-level selection (e.g., Swenson et al.,[50] Blouin et al.,[51] and Sánchez et al.[52]), where any community-level function may be assigned as a fitness value at the community level. The limits of community-level selection are being investigated at the moment, but there is solid theoretical and empirical evidence that communities can be considered units of selection at least in the context of directed evolution or artificial selection.[52]

DOES COMMUNITY COMPOSITION UNIQUELY DETERMINE COMMUNITY FUNCTION?

Before we end, we would like to address what may appear to be the proverbial elephant in the room. Although we hope that we have convinced the reader that learning the map between community composition and function may have a transformative impact on our ability to understand and engineer microbial consortia, it may not be immediately obvious that such a map will necessarily always exist. To what extent does ecological function measured at a given time depend on the composition of a community at that same time? This question is more nuanced than it might appear at first sight. For instance, an important function of microbial consortia is the production of extracellular molecules from metabolites to secreted enzymes. The change in concentration of these secreted molecules depends on the rate at which they are produced, which indeed depends on the abundances of different members of the consortium as well as on their respective per-capita production rates (Figure 1). However, the concentration of secreted molecules also depends on the rates of molecular degradation, biochemical inactivation, diffusion out of the volume or area of interest, and other degradative processes that eliminate the target molecule and that do not necessarily depend on the current state of the community. This creates conditions for which the current state of the function of a community depends not just on its current composition but rather on the history of assembly. This idea is perhaps best illustrated through a simple mathematical model.

We can formally model the rate of accumulation of an extracellular molecule (say, an enzyme E) in a volume of interest as:

$$dE/dt = h(\mathbf{z}, \mathbf{x}_g) - \lambda(\mathbf{z}, E)$$

Where $h(.)$ represents the rate of enzyme secretion as a function of the collection of environmental parameters $\boldsymbol{z}$ and the present species abundance vector $\boldsymbol{x_g}$, and $\lambda(.)$ represents the net rate of enzyme loss through all possible pathways. The latter should depend on the enzyme concentration E as well as on the environmental parameters captured in $\mathbf{z}$ (which may include

the concentration of proteases, enzyme inhibitors, or other environmental parameters affecting the stability of the enzyme such as the pH). Of course, the environment and genotypes obey their own equations, which are an extension to higher dimensions of those introduced by Lewontin:

$$d\mathbf{x}_g / dt = \boldsymbol{r}(\mathbf{z}, \mathbf{x}_g)$$

$$d\mathbf{z} / dt = \boldsymbol{k}(\mathbf{z}, \mathbf{x}_g)$$

Where $\boldsymbol{r}$(.) and $\boldsymbol{k}$(.) denote the dynamical equations governing the temporal evolution of $\mathbf{x_g}$ and $\mathbf{z}$, respectively. In general, there is no reason to expect that if we integrate those equations, we should find that E is an explicit or even implicit mathematical function of $\mathbf{x_g}$ alone or even a function of $\mathbf{x_g}$ and $\mathbf{z}$. This reflects the fact, which should be true for many community-level traits, that the function of the consortium at a given time is not, in general, uniquely defined by its composition at that time. Rather, it should be a result of the particular dynamical process of community assembly (i.e., the assembly history) that has led the community to its current compositional state and, similarly, of the dynamical history of the environmental parameters captured in $\mathbf{z}$.

Does that mean that a function that uniquely maps community structure to function does, in general, not exist? It seems to follow from the above argument that, in general, it does not. However, there exist many important limits and cases of practical utility for which the function of a community at a given point in time can indeed be uniquely defined by its composition at that time. To illustrate these important scenarios, let us go back to the example given above, where the function of interest is the concentration of a target extracellular enzyme. For a structure-function landscape to be well-defined in this case, there should exist a function $E(\mathbf{x_g})$ that provides a 1:1 map between the concentration of secreted enzyme at a given time and the community composition at that time. One limit where this function exists occurs when the dynamics of E and $\mathbf{z}$ are very fast compared with the (population) dynamics of $\mathbf{x_g}$. In this limit, $\mathbf{x_g}$ is approximately constant in the timescale required for E and $\mathbf{z}$ to equilibrate, and therefore, E (and $\mathbf{z}$) will find a local equilibrium for every value of $\mathbf{x_g}$ before this changes significantly. Without loss of generality, let us consider the simple case where $(E,\mathbf{z}) = \lambda(z)E$. In the separation of timescales limit, we find that the form of the structure-function landscape is $E(\mathbf{x_g}) = h(\mathbf{z}^*,\mathbf{x_g})/\lambda(\mathbf{z}^*)$, where the relevant environmental variables captured in $\mathbf{z}$ also equilibrate rapidly, generally (though not necessarily) reaching a unique value ($\mathbf{z}^*$) for each $\mathbf{x_g}$. In this case, and save for special circumstances such as when there exist memory effects or hysteresis in the per-cell contribution to function, causing non-linearities in $k(\mathbf{x_g},\mathbf{z}^*)$, every $\mathbf{x_g}$ may be associated with a unique value of E.

Although the separation of timescales is a rather stringent limit that applies only to a narrow range of real-life scenarios, it brings up a larger point that although the structure-function landscape is not defined in general, it may exist when communities are in a steady state. For many biotechnological applications, communities may be maintained in (or close to) a steady state by either placing them in a continuous culture device or through serial passaging (Figure 5). In chemostats, both species composition and all environmental parameters should reach a steady state. Going back to the example discussed above, the concentration of our target enzyme E should be independent of assembly history and uniquely linked to the equilibrium concentration of $\mathbf{x_g}$ (save for the hysteretic situation discussed in the previous paragraph). In the case of serially passaged consortia, empirical communities have been generally found to converge to a state of "generational stability,"[67,102] at least when the passaging is done under constant conditions.[27,28,52,103,104]

Another situation of interest in biotechnology is single-batch synthetic communities. These can be formed by co-inoculating multiple community members at defined initial abundances in a bioreactor. This consortium is then incubated for a given time period, at the end of which the function of interest is measured. Here, the requirement for having a well-defined structure-function landscape is that the population dynamics of the consortium within the batch are highly reproducible and converge deterministically to the same final community state at the time of harvest. In this case, the entire within-batch dynamics including both environmental and species abundance variables are uniquely determined by the starting abundances of the members of the consortium. Thus, each initial community state $\boldsymbol{x_g}$ will be characterized by a single value of the function (i.e., E) at the time of harvest, which defines a 1:1 map between both (Figure 5). Beyond any specific assumptions regarding the model above, there is also empirical evidence for this last scenario. The primary evidence for this is the remarkable reproducibility and determinism of community structure and dynamics during community assembly. For example, a reproducible succession of three functional guilds reliably occurs on polysaccharide particles in marine communities.[105] This suggests that given a specific niche to colonize and a sufficiently diverse regional species pool, the structure of the assembled community is reproducible. This empirical observation suggests, but does not prove, that there are convergent ecological solutions to well-defined functional problems—degrading polysaccharides in this case. Similar results are observed in glucose and other small molecule enrichments[27,28] and detailed more broadly in surveys of the functional classes of bacteria in the marine microbiome.[106,107] Likewise, host-associated communities also exhibit highly conserved metagenomic structure from host to host,[106,107] suggesting that the functional landscape is a well-defined object with the structure being tightly and reliably linked to function. What remains is to learn this mapping quantitatively and leverage that knowledge to design and predict community behavior.

Conclusion

We have presented an overview of a concept that is gaining momentum in the systems biology of microbial communities: the community-function landscape as an ecological extension of genotype-phenotype maps and the fitness landscape concept in genetics. To fully exploit the landscape concept at the level of communities, we must be careful to consider the important differences that exist between communities of self-replicating organisms and biological systems like an enzyme, whose building blocks are non-replicating molecules. For instance, the amino acid sequence of an enzyme is (generally) stable once a mutation is introduced. By contrast, introducing a new species in an ecosystem will (also in general) lead to alterations in the population size of all other species that interact with it, either directly or

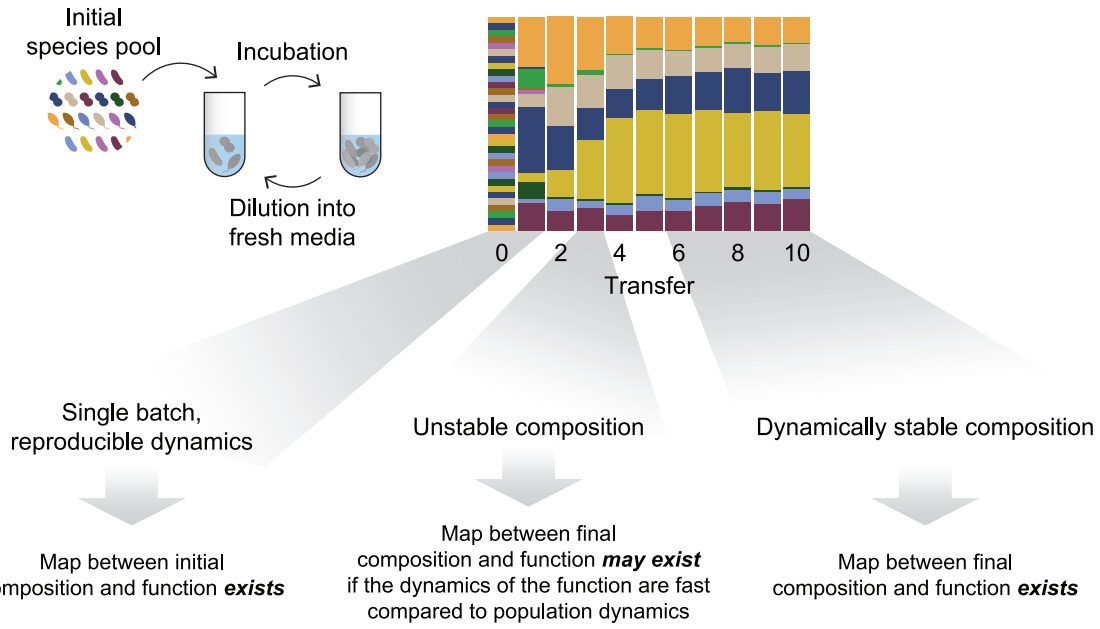

Figure 5. Conditions for the existence of an ecological structure-function map

For illustrative purposes, we use as an example a hypothetical case where we stabilize a community from a diverse initial pool of species through periodic transfers in the laboratory, and we measure an enzymatic function akin to the amylolytic activity discussed as an example in the main text (see also Figure 1). The barplot shows the composition of the community at each transfer. The structure-function map will exist in three scenarios: (1) if we map the initial composition to function, assuming that the ecological dynamics are reproducible; (2) if we map the final composition to function in a dynamically stable community, e.g., transfers 6–10; and (3) if we map the final composition to function in an unstable community, but only if the functional dynamics (enzyme concentration E and environmental parameters affecting its activity z, see main text) are fast compared with population dynamics.

indirectly. In other words: community composition is a dynamical process and we cannot easily freeze it in one particular state. Adding a community member is therefore not the same as adding a mutation to an enzyme, and care must be exercised to avoid overextending and overinterpreting the analogy. By the same token, it is important to be mindful of the fact that many community functions are not directly determined by their community composition. Rather, community functions may have their own dynamics, and their value at a given time can be a product of the history of the assembly of that community as well as past events. Therefore, a community function is not necessarily encoded by the particular composition of the community at the time when it was measured. Finally, one must be careful to recognize that, just as fitness landscapes map genotype to fitness under a fixed set of environmental conditions, any relationship between community composition and function will also be sensitive to environmental changes. How exactly changing the environment will alter community functions is a subject that remains poorly understood and that requires further investigation. Despite these caveats, there are often-realized limits when the association between community composition and function is indeed strong. Indeed, recent efforts to quantitatively predict the latter from the former have been successful, making us optimistic about the usefulness of the community landscape concept.

It should be obvious to the reader that we are merely scratching the surface of a very rich and we believe potentially rather fruitful line of inquiry. Parallelisms between the exploration of fitness landscapes in evolutionary engineering and the exploration of structure-function landscapes may provide important insights into our understanding of the mapping between community composition and function, and our ability to engineer microbial consortia. The field of quantitative genetics has built powerful methodologies to reconstruct and navigate genotype-phenotype maps, and it also has developed a strong conceptual and theoretical framework to understand the origins of these genetic landscapes. Extending these methods and ideas from quantitative genetics and computer science into microbial ecology could radically improve our ability to understand and engineer the function of microbial communities. We shall be most satisfied if this review contributes to stimulating some of these efforts.

ACKNOWLEDGMENTS

A.S. was supported by the Spanish Ministry of Science and Innovation under project PID2021-125478NA-100. S.K. received support from NSF grants MCB 2117477, EF 2025293, and MCB 1921439.

DECLARATION OF INTERESTS

A.S. is an editorial board member at *Cell Systems*.

REFERENCES

1. May, A., Narayanan, S., Alcock, J., Varsani, A., Maley, C., and Aktipis, A. (2019). Kombucha: a novel model system for cooperation and conflict in a complex multi-species microbial ecosystem. PeerJ *7*, e7565.

2. Wolfe, B.E., Button, J.E., Santarelli, M., and Dutton, R.J. (2014). Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. Cell *158*, 422–433.

3. Belda, I., Zarraonaindia, I., Perisin, M., Palacios, A., and Acedo, A. (2017). From vineyard soil to wine fermentation: microbiome approximations to explain the "terroir" concept. Front. Microbiol. *8*, 821.

4. Blasche, S., Kim, Y., Mars, R.A.T., Machado, D., Maansson, M., Kafkia, E., Milanese, A., Zeller, G., Teusink, B., Nielsen, J., et al. (2021). Metabolic cooperation and spatiotemporal niche partitioning in a kefir microbial community. Nat. Microbiol. *6*, 196–208.

5. Minty, J.J., Singer, M.E., Scholz, S.A., Bae, C.H., Ahn, J.H., Foster, C.E., Liao, J.C., and Lin, X.N. (2013). Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. Proc. Natl. Acad. Sci. USA *110*, 14592–14597.

6. Alper, H., and Stephanopoulos, G. (2009). Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? Nat. Rev. Microbiol. *7*, 715–723.

7. Jiang, Y., Dong, W., Xin, F., and Jiang, M. (2020). Designing synthetic microbial consortia for biofuel production. Trends Biotechnol. *38*, 828–831.

8. Senne de Oliveira Lino, F., Bajic, D., Vila, J.C.C., Sánchez, A., and Sommer, M.O.A. (2021). Complex yeast–bacteria interactions affect the yield of industrial ethanol fermentation. Nat. Commun. *12*, 1498.

9. Weng, J.K., Li, X., Bonawitz, N.D., and Chapple, C. (2008). Emerging strategies of lignin engineering and degradation for cellulosic biofuel production. Curr. Opin. Biotechnol. *19*, 166–172.

10. Maleki, N., Safari, M., and Eiteman, M.A. (2018). Conversion of glucose-xylose mixtures to pyruvate using a consortium of metabolically engineered *Escherichia coli*. Eng. Life Sci. *18*, 40–47.

11. Hu, J., Xue, Y., Guo, H., Gao, M.T., Li, J., Zhang, S., and Tsang, Y.F. (2017). Design and composition of synthetic fungal-bacterial microbial consortia that improve lignocellulolytic enzyme activity. Bioresour. Technol. *227*, 247–255.

12. Piccardi, P., Vessman, B., and Mitri, S. (2019). Toxicity drives facilitation between 4 bacterial species. Proc. Natl. Acad. Sci. USA *116*, 15979–15984.

13. Swenson, W., Arendt, J., and Wilson, D.S. (2000). Artificial selection of microbial ecosystems for 3-chloroaniline biodegradation. Environ. Microbiol. *2*, 564–571.

14. Zanaroli, G., Di Toro, S., Todaro, D., Varese, G.C., Bertolotto, A., and Fava, F. (2010). Characterization of two diesel fuel degrading microbial consortia enriched from a non acclimated, complex source of microorganisms. Microb. Cell Factories *9*, 10.

15. Baas, P., Bell, C., Mancini, L.M., Lee, M.N., Conant, R.T., and Wallenstein, M.D. (2016). Phosphorus mobilizing consortium Mammoth P(TM) enhances plant growth. PeerJ *4*, e2121.

16. Baas, P., Bell, C., Mancini, L., Lee, M., Wallenstein, M.D., and Conant, R.T. (2020). *In vitro* selection of a microbial consortium predictive of synergistic functioning along multiple ecosystem scales. Preprint at bioRxiv. https://doi.org/10.1101/2020.11.12.379529.

17. Roell, G.W., Zha, J., Carr, R.R., Koffas, M.A., Fong, S.S., and Tang, Y.J. (2019). Engineering microbial consortia by division of labor. Microb. Cell Factories *18*, 35.

18. Ergal, İ., Gräf, O., Hasibar, B., Steiner, M., Vukotić, S., Bochmann, G., Fuchs, W., and Rittmann, S.K.R. (2020). Biohydrogen production beyond the Thauer limit by precision design of artificial microbial consortia. Commun. Biol. *3*, 443.

19. Sgobba, E., and Wendisch, V.F. (2020). Synthetic microbial consortia for small molecule production. Curr. Opin. Biotechnol. *62*, 72–79.

20. Macia, J., Manzoni, R., Conde, N., Urrios, A., de Nadal, E., Solé, R., and Posas, F. (2016). Implementation of complex biological logic circuits using spatially distributed multicellular consortia. PLoS Comput. Biol. *12*, e1004685.

21. Thommes, M., Wang, T., Zhao, Q., Paschalidis, I.C., and Segrè, D. (2019). Designing metabolic division of labor in microbial communities. mSystems *4*, e00263-18.

22. Mizrahi, I., Wallace, R.J., and Moraïs, S. (2021). The rumen microbiome: balancing food security and environmental impacts. Nat. Rev. Microbiol. *19*, 553–566.

23. Erkus, O., de Jager, V.C.L., Spus, M., van Alen-Boerrigter, I.J., van Rijswijck, I.M.H., Hazelwood, L., Janssen, P.W., van Hijum, S.A., Kleerebezem, M., and Smid, E.J. (2013). Multifactorial diversity sustains microbial community stability. ISME J. *7*, 2126–2136.

24. Rozen, D.E., and Lenski, R.E. (2000). Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. Am. Nat. *155*, 24–35.

25. Good, B.H., McDonald, M.J., Barrick, J.E., Lenski, R.E., and Desai, M.M. (2017). The dynamics of molecular evolution over 60,000 generations. Nature *551*, 45–50.

26. Kinnersley, M.A., Holben, W.E., and Rosenzweig, F. (2009). *E Unibus Plurum*: genomic analysis of an experimentally evolved polymorphism in *Escherichia coli*. PLoS Genet. *5*, e1000713.

27. Goldford, J.E., Lu, N., Bajić, D., Estrela, S., Tikhonov, M., Sanchez-Gorostiaga, A., Segrè, D., Mehta, P., and Sanchez, A. (2018). Emergent simplicity in microbial community assembly. Science *361*, 469–474.

28. Estrela, S., Vila, J.C.C., Lu, N., Bajić, D., Rebolleda-Gómez, M., Chang, C.Y., Goldford, J.E., Sanchez-Gorostiaga, A., and Sánchez, Á. (2022). Functional attractors in microbial community assembly. Cell Syst. *13* 29–42.e7.

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

30. Estrela, S., Sanchez-Gorostiaga, A., Vila, J.C., and Sanchez, A. (2021). Nutrient dominance governs the assembly of microbial communities in mixed nutrient environments. eLife *10*, e65948.

31. Mancuso, C.P., Lee, H., Abreu, C.I., Gore, J., and Khalil, A.S. (2021). Environmental fluctuations reshape an unexpected diversity-disturbance relationship in a microbial community. eLife *10*, e67175.

32. Lu, H., Diaz, D.J., Czarnecki, N.J., Zhu, C., Kim, W., Shroff, R., Acosta, D.J., Alexander, B.R., Cole, H.O., Zhang, Y., et al. (2022). Machine learning-aided engineering of hydrolases for PET depolymerization. Nature *604*, 662–667.

33. Arnold, F.H. (2019). Innovation by evolution: bringing new chemistry to life (Nobel lecture). Angew. Chem. Int. Ed. Engl. *58*, 14420–14426.

34. Wendisch, V.F., Bott, M., and Eikmanns, B.J. (2006). Metabolic engineering of *Escherichia coli* and *Corynebacterium glutamicum* for biotechnological production of organic acids and amino acids. Curr. Opin. Microbiol. *9*, 268–274.

35. Yang, D., Park, S.Y., and Lee, S.Y. (2021). Production of rainbow colorants by metabolically engineered *Escherichia coli*. Adv. Sci. (Weinh) *8*, e2100743.

36. Chica, R.A., Doucet, N., and Pelletier, J.N. (2005). Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Curr. Opin. Biotechnol. *16*, 378–384.

37. Kuchner, O., and Arnold, F.H. (1997). Directed evolution of enzyme catalysts. Trends Biotechnol. *15*, 523–530.

38. Bornscheuer, U.T., Huisman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C., and Robins, K. (2012). Engineering the third wave of biocatalysis. Nature *485*, 185–194.

39. Bloom, J.D., and Arnold, F.H. (2009). In the light of directed evolution: pathways of adaptive protein evolution. Proc. Natl. Acad. Sci. USA *106* (*Suppl 1*), 9995–10000.

40. Romero, P.A., and Arnold, F.H. (2009). Exploring protein fitness landscapes by directed evolution. Nat. Rev. Mol. Cell Biol. *10*, 866–876.

41. Tracewell, C.A., and Arnold, F.H. (2009). Directed enzyme evolution: climbing fitness peaks one amino acid at a time. Curr. Opin. Chem. Biol. *13*, 3–9.

42. Russ, W.P., Figliuzzi, M., Stocker, C., Barrat-Charlaix, P., Socolich, M., Kast, P., Hilvert, D., Monasson, R., Cocco, S., Weigt, M., and Ranganathan, R. (2020). An evolution-based model for designing chorismate mutase enzymes. Science *369*, 440–445.

43. Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D.S., Sander, C., Zecchina, R., Onuchic, J.N., Hwa, T., and Weigt, M. (2011). Direct-

coupling analysis of residue coevolution captures native contacts across many protein families. Proc. Natl. Acad. Sci. USA *108*, E1293–E1301.

44. Stiffler, M.A., Hekstra, D.R., and Ranganathan, R. (2015). Evolvability as a function of purifying selection in TEM-1 β-lactamase. Cell *160*, 882–892.

45. Raman, A.S., White, K.I., and Ranganathan, R. (2016). Origins of allostery and evolvability in proteins: a case study. Cell *166*, 468–480.

46. Halabi, N., Rivoire, O., Leibler, S., and Ranganathan, R. (2009). Protein sectors: evolutionary units of three-dimensional structure. Cell *138*, 774–786.

47. Hekstra, D.R., and Leibler, S. (2012). Contingency and statistical laws in replicate microbial closed ecosystems. Cell *149*, 1164–1173.

48. Frentz, Z., Kuehn, S., and Leibler, S. (2015). Strongly deterministic population dynamics in closed microbial communities. Phys. Rev. X *5*, 041014.

49. Raman, A.S., Gehrig, J.L., Venkatesh, S., Chang, H.W., Hibberd, M.C., Subramanian, S., Kang, G., Bessong, P.O., Lima, A.A.M., Kosek, M.N., et al. (2019). A sparse covarying unit that describes healthy and impaired human gut microbiota development. Science *365*.

50. Swenson, W., Wilson, D.S., and Elias, R. (2000). Artificial ecosystem selection. Proc. Natl. Acad. Sci. USA *97*, 9110–9114.

51. Blouin, M., Karimi, B., Mathieu, J., and Lerch, T.Z. (2015). Levels and limits in artificial selection of communities. Ecol. Lett. *18*, 1040–1048.

52. Sánchez, Á., Vila, J.C.C., Chang, C.Y., Diaz-Colunga, J., Estrela, S., and Rebolleda-Gomez, M. (2021). Directed evolution of microbial communities. Annu. Rev. Biophys. *50*, 323–341.

53. Chang, C.Y., Vila, J.C.C., Bender, M., Li, R., Mankowski, M.C., Bassette, M., Borden, J., Golfier, S., Sanchez, P.G.L., Waymack, R., et al. (2021). Engineering complex communities by directed evolution. Nat. Ecol. Evol. *5*, 1011–1023.

54. Mueller, U.G., Juenger, T.E., Kardish, M.R., Carlson, A.L., Burns, K.M., Edwards, J.A., Smith, C.C., Fang, C.C., and Des Marais, D.L. (2021). Artificial selection on microbiomes to breed microbiomes that confer salt tolerance to plants. mSystems *6*, e0112521.

55. Wright, R.J., Gibson, M.I., and Christie-Oleza, J.A. (2019). Understanding microbial community dynamics to improve optimal microbiome selection. Microbiome *7*, 85.

56. Sanchez-Gorostiaga, A., Bajić, D., Osborne, M.L., Poyatos, J.F., and Sanchez, A. (2019). High-order interactions distort the functional landscape of microbial consortia. PLoS Biol. *17*, e3000550.

57. Gould, A.L., Zhang, V., Lamberti, L., Jones, E.W., Obadia, B., Korasidis, N., Gavryushkin, A., Carlson, J.M., Beerenwinkel, N., and Ludington, W.B. (2018). Microbiome interactions shape host fitness. Proc. Natl. Acad. Sci. USA *115*, E11951–E11960.

58. Senay, Y., John, G., Knutie, S.A., and Brandon Ogbunugafor, C. (2019). Deconstructing higher-order interactions in the microbiota: a theoretical examination. Preprint at bioRxiv. https://doi.org/10.1101/647156.

59. Baranwal, M., Clark, R.L., Thompson, J., Sun, Z., Hero, A.O., and Venturelli, O.S. (2022). Recurrent neural networks enable design of multifunctional synthetic human gut microbiome dynamics. eLife *11*, e73870.

60. Arora, J., Mars Brisbin, M.A., and Mikheyev, A.S. (2020). Effects of microbial evolution dominate those of experimental host-mediated indirect selection. PeerJ *8*, e9350.

61. Eble, H., Joswig, M., Lamberti, L., and Ludington, W.B. (2021). High dimensional geometry of fitness landscapes identifies master regulators of evolution and the microbiome. Preprint at bioRxiv. https://doi.org/10.1101/2021.09.11.459926.

62. Xie, L., and Shou, W. (2021). Steering ecological-evolutionary dynamics to improve artificial selection of microbial communities. Nat. Commun. *12*, 6799.

63. George, A.B., and Korolev, K.S. (2021). Ecological landscapes guide the assembly of optimal microbial communities. Preprint at bioRxiv. https://doi.org/10.1101/2021.12.14.472701.

64. Clark, R.L., Connors, B.M., Stevenson, D.M., Hromada, S.E., Hamilton, J.J., Amador-Noguez, D., and Venturelli, O.S. (2021). Design of synthetic human gut microbiome assembly and butyrate production. Nat. Commun. *12*, 3254.

65. Bittleston, L.S., Gralka, M., Leventhal, G.E., Mizrahi, I., and Cordero, O.X. (2020). Context-dependent dynamics lead to the assembly of functionally distinct microbial communities. Nat. Commun. *11*, 1440.

66. Gopalakrishnappa, C., Gowda, K., Prabhakara, K.H., and Kuehn, S. (2022). An ensemble approach to the structure-function problem in microbial communities. iScience *25*, 103761.

67. Xie, L., Yuan, A.E., and Shou, W. (2019). Simulations reveal challenges to artificial community selection and possible strategies for success. PLoS Biol. *17*, e3000295.

68. Wagner, A. (2019). Life Finds a Way: What Evolution Teaches Us About Creativity (Basic Books).

69. Wood, K., Nishida, S., Sontag, E.D., and Cluzel, P. (2012). Mechanism-independent method for predicting response to multidrug combinations in bacteria. Proc. Natl. Acad. Sci. USA *109*, 12254–12259.

70. Tekin, E., Beppler, C., White, C., Mao, Z., Savage, V.M., and Yeh, P.J. (2016). Enhanced identification of synergistic and antagonistic emergent interactions among three or more drugs. J. R. Soc. Interface *13*, 20160332.

71. Beppler, C., Tekin, E., Mao, Z., White, C., McDiarmid, C., Vargas, E., Miller, J.H., Savage, V.M., and Yeh, P.J. (2016). Uncovering emergent interactions in three-way combinations of stressors. J. R. Soc. Interface *13*, 20160800.

72. Tekin, E., Yeh, P.J., and Savage, V.M. (2018). General form for interaction measures and framework for deriving higher-order emergent effects. Front. Ecol. Evol. *6*, 166.

73. Sanchez, A. (2019). Defining higher-order interactions in synthetic ecology: lessons from physics and quantitative genetics. Cell Syst. *9*, 519–520.

74. Guo, X., and Boedicker, J. (2016). High-order interactions between species strongly influence the activity of microbial communities. Biophys. J. *110*, 143a.

75. Chen, Y., Lin, C.J., Jones, G., Fu, S., and Zhan, H. (2009). Enhancing biodegradation of wastewater by microbial consortia with fractional factorial design. J. Hazard. Mater. *171*, 948–953.

76. Eng, A., and Borenstein, E. (2019). Microbial community design: methods, applications, and opportunities. Curr. Opin. Biotechnol. *58*, 117–128.

77. Gowda, K., Ping, D., Mani, M., and Kuehn, S. (2022). Genomic structure predicts metabolite dynamics in microbial communities. Cell *185*. 530–546.e25.

78. Kim, I.J., Lee, H.J., Choi, I.G., and Kim, K.H. (2014). Synergistic proteins for the enhanced enzymatic hydrolysis of cellulose by cellulase. Appl. Microbiol. Biotechnol. *98*, 8469–8480.

79. Mickalide, H., and Kuehn, S. (2019). Higher-order interaction between species inhibits bacterial invasion of a phototroph-predator microbial community. Cell Syst. *9* 521–533.e10.

80. Poelwijk, F.J., Socolich, M., and Ranganathan, R. (2019). Learning the pattern of epistasis linking genotype and phenotype in a protein. Nat. Commun. *10*, 4213.

81. Morin, M.A., Morrison, A.J., Harms, M.J., and Dutton, R.J. (2022). Higher-order interactions shape microbial interactions as microbial community complexity increases. Preprint at bioRxiv. https://doi.org/10.1101/2022.05.19.492721.

82. Reddy, G., and Desai, M.M. (2021). Global epistasis emerges from a generic model of a complex trait. eLife *10*, e64740.

83. Khan, A.I., Dinh, D.M., Schneider, D., Lenski, R.E., and Cooper, T.F. (2011). Negative epistasis between beneficial mutations in an evolving bacterial population. Science *332*, 1193–1196.

84. Díaz-Colunga, J., Skwara, A., Vila, J.C.C., Bajić, D., and Sánchez, Á. (2022). Emergent ecosystem functions follow simple quantitative rules. Preprint at bioRxiv. https://doi.org/10.1101/2022.06.21.496987.

85. Ghedini, G., Marshall, D.J., and Loreau, M. (2022). Phytoplankton diversity affects biomass and energy production differently during community development. Funct. Ecol. *36*, 446–457.

86. Kuebbing, S.E., Classen, A.T., Sanders, N.J., and Simberloff, D. (2015). Above- and below-ground effects of plant diversity depend on species origin: an experimental test with multiple invaders. New Phytol. *208*, 727–735.

87. Langenheder, S., Bulling, M.T., Solan, M., and Prosser, J.I. (2010). Bacterial biodiversity-ecosystem functioning relations are modified by environmental complexity. PLoS One *5*, e10834.

88. Husain, K., and Murugan, A. (2020). Physical constraints on epistasis. Mol. Biol. Evol. *37*, 2865–2874.

89. Otwinowski, J., McCandlish, D.M., and Plotkin, J.B. (2018). Inferring the shape of global epistasis. Proc. Natl. Acad. Sci. USA *115*, E7550–E7558.

90. Wei, X., and Zhang, J. (2019). Patterns and mechanisms of diminishing returns from beneficial mutations. Mol. Biol. Evol. *36*, 1008–1021.

91. Romero, P.A., Krause, A., and Arnold, F.H. (2013). Navigating the protein fitness landscape with Gaussian processes. Proc. Natl. Acad. Sci. USA *110*, E193–E201.

92. Tonner, P.D., Pressman, A., and Ross, D. (2022). Interpretable modeling of genotype-phenotype landscapes with state-of-the-art predictive power. Proc. Natl. Acad. Sci. USA *119*. e2114021119.

93. Tareen, A., Kooshkbaghi, M., Posfai, A., Ireland, W.T., McCandlish, D.M., and Kinney, J.B. (2022). MAVE-NN: learning genotype-phenotype maps from multiplex assays of variant effect. Genome Biol. *23*, 98.

94. Nahum, J.R., Godfrey-Smith, P., Harding, B.N., Marcus, J.H., Carlson-Stevermer, J., and Kerr, B. (2015). A tortoise-hare pattern seen in adapting structured and unstructured populations suggests a rugged fitness landscape in bacteria. Proc. Natl. Acad. Sci. USA *112*, 7530–7535.

95. Aguilar-Rodríguez, J., Payne, J.L., and Wagner, A. (2017). A thousand empirical adaptive landscapes and their navigability. Nat. Ecol. Evol. *1*, 45.

96. Poelwijk, F.J., Tănase-Nicola, S., Kiviet, D.J., and Tans, S.J. (2011). Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. J. Theor. Biol. *272*, 141–144.

97. Wittmann, B.J., Yue, Y., and Arnold, F.H. (2021). Informed training set design enables efficient machine learning-assisted directed protein evolution. Cell Syst. *12* 1026–1045.e7.

98. Otwinowski, J., and Plotkin, J.B. (2014). Inferring fitness landscapes by regression produces biased estimates of epistasis. Proc. Natl. Acad. Sci. USA *111*, E2301–E2309.

99. Kauffman, S.A., and Weinberger, E.D. (1989). The NK model of rugged fitness landscapes and its application to maturation of the immune response. J. Theor. Biol. *141*, 211–245.

100. du Plessis, L., Leventhal, G.E., and Bonhoeffer, S. (2016). How good are statistical models at approximating complex fitness landscapes? Mol. Biol. Evol. *33*, 2454–2468.

101. Leo-Macias, A., Lopez-Romero, P., Lupyan, D., Zerbino, D., and Ortiz, A.R. (2005). An analysis of core deformations in protein superfamilies. Biophys. J. *88*, 1291–1299.

102. Doulcier, G., Lambert, A., De Monte, S., and Rainey, P.B. (2020). Eco-evolutionary dynamics of nested Darwinian populations and the emergence of community-level heredity. eLife *9*, e53433.

103. Chang, C.Y., Osborne, M.L., Bajic, D., and Sanchez, A. (2020). Artificially selecting bacterial communities using propagule strategies. Evolution *74*, 2392–2403.

104. Estrela, S., Sánchez, Á., and Rebolleda-Gómez, M. (2021). Multi-replicated enrichment communities as a model system in microbial ecology. Front. Microbiol. *12*, 657467.

105. Enke, T.N., Datta, M.S., Schwartzman, J., Cermak, N., Schmitz, D., Barrere, J., Pascual-García, A., and Cordero, O.X. (2019). Modular assembly of polysaccharide-degrading marine microbial communities. Curr. Biol. *29* 1528–1535.e6.

106. Louca, S., Jacques, S.M.S., Pires, A.P.F., Leal, J.S., Srivastava, D.S., Parfrey, L.W., Farjalla, V.F., and Doebeli, M. (2016). High taxonomic variability despite stable functional structure across microbial communities. Nat. Ecol. Evol. *1*, 15.

107. Louca, S., Polz, M.F., Mazel, F., Albright, M.B.N., Huber, J.A., O'Connor, M.I., Ackermann, M., Hahn, A.S., Srivastava, D.S., Crowe, S.A., et al. (2018). Function and functional redundancy in microbial systems. Nat. Ecol. Evol. *2*, 936–943.