The community-function landscape of microbial consortia

Alvaro Sanchez[1,2], Djordje Bajic[1], Juan Diaz-Colunga[1], Abigail Skwara[1], Jean Vila[1] & Seppe Kuehn[3,4]

[1]*Department of Ecology & Evolutionary Biology & Microbial Sciences Institute, Yale University, New Haven, Connecticut, 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, Illinois, USA.* [4]*Department of Ecology and Evolution, The University of Chicago, Chicago, Illinois, USA.*

Correspondence: A.S: alvaro.sanchez@cnb.csic.es, JDC: juan.diazcolunga@yale.edu , DB: djordje.bajic@yale.edu , A.Sk: abby.skwara@gmail.com, JV: jean.vila@yale.edu SK: seppekuehn@uchicago.edu ,

Abstract

Quantitatively linking the composition and function of microbial communities is a major aspiration of microbial ecology. It is also a critical step in the path towards engineering synthetic consortia and manipulating natural microbiomes. The functions of microbial communities are collective properties that emerge from a complex web of molecular interactions between individual cells, which in turn lead to population-level interactions among strains and species. Incorporating this complexity into predictive models has been highly challenging. A similar problem of predicting phenotype from genotype has been addressed for decades in the field of quantitative genetics, leading to advances in the fields of protein and molecular engineering. By analogy to the genotype-phenotype landscape, 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 (Belda et al., 2017; Blasche et al., 2021; May et al., 2019; Wolfe et al., 2014), to contemporary uses in biofuel production (Alper and Stephanopoulos, 2009; Jiang et al., 2020; Minty et al., 2013; Senne de Oliveira Lino et al., 2021), the valorization of discarded plant materials (Hu et al., 2017; Maleki et al., 2018; Weng et al., 2008), bioremediation (Piccardi et al., 2019; Swenson et al., 2000a; Zanaroli et al., 2010), crop fertilization (Baas et al., 2016, 2020), and many more (Ergal et al., 2020; Macia et al., 2016; Roell et al., 2019; Sgobba and Wendisch, 2020).

Relative to monocultures, microbial communities offer multiple advantages in biotechnology. Among these, they permit specialization and division of labor (Roell et al., 2019; Thommes et al., 2019) 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. This diversity can enable communities to remain resilient to perturbations that single strains might not survive (Erkus et al., 2013). 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 (Good et al., 2017; Kinnersley et al., 2009; Rozen and Lenski, 2000). 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 (Dal Bello et al., 2021; Estrela et al., 2021a, 2021b; Goldford et al., 2018; Mancuso et al., 2021). 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 (Arnold, 2019; Lu et al., 2022) or metabolic and genetic networks (Wendisch et al., 2006; Yang et al., 2021). 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 (Bornscheuer et al., 2012; Chica et al., 2005; Kuchner and Arnold, 1997). 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 (Arnold, 2019; Bornscheuer et al., 2012; Chica et al., 2005). 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 (Bloom and Arnold, 2009; Romero and Arnold, 2009; Tracewell and Arnold, 2009). Perhaps the most successful example is the development of directed evolution, which has enabled the top-down engineering of different kinds of proteins (Arnold, 2019). 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 (Arnold, 2019).

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 (Russ et al., 2020), inferring folds (Morcos et al., 2011), and insights into evolvability (Stiffler et al., 2015) and allostery (Raman et al., 2016). 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 (Halabi et al., 2009). This result means that engineering proteins does not require an exhaustive search of sequence space (an impossible task) but instead a constrained search within a low-dimensional subspace. It is intriguing to note that this inherent low-dimensionality is also observed in biological systems at higher scales of organization, such as behavior (Berman et al., 2013; Jordan et al., 2013), developmental programs (Alba et al., 2021), and even microbial communities (Raman et al., 2019).

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 their 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. While this is true, community-level selection can be applied under artificial conditions, where an arbitrary fitness function can be applied (Blouin et al., 2015; Sánchez et al., 2021; Swenson et al., 2000b). 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 (Chang et al., 2021a; Mueller Ulrich G. et al.; Wright et al., 2019).

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 (Baranwal et al., 2022; Gould et al., 2018; Sánchez et al., 2021; Sanchez-Gorostiaga et al., 2019; Senay et al., 2019). Examples range from fruit-fly microbial consortia whose function is the host's lifespan and other life-history traits (Arora et al., 2020), to sugarcane biorefinery consortia whose function is the amount of ethanol produced during a single-batch fermentation (Senne de Oliveira Lino et al., 2021). These and other studies (Bittleston et al., 2020; Clark et al., 2021; Eble et al., 2021; George and Korolev, 2021; Gopalakrishnappa et al., 2022; Gould et al., 2018; Sanchez-Gorostiaga et al., 2019; Senay et al., 2019; Xie and Shou, 2021; Xie et al., 2019) 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 it genotypes $\boldsymbol{g} = \{g_1, g_2, ..., g_n\}$ and their respective abundances $\boldsymbol{x_g} = \{x_1, x_2, ..., x_n\}$. 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 paper 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” (Wagner, 2019) may provide a helpful theoretical framework to help us 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) (Fig. 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 (Fig. 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 (Tekin Elif et al., 2016; Wood et al., 2012) or combinations of stressors (Beppler Casey et al., 2016), among others (Tekin et al., 2018). In recent years, we (and others) have extended it to ecological systems as well (Eble et al., 2021; Gould et al., 2018; Guo and Boedicker, 2016; Sanchez, 2019; Sanchez-Gorostiaga et al., 2019; Senay

et al., 2019; Senne de Oliveira Lino et al., 2021), and the underlying idea was already present in earlier efforts to model the emergence of community function (Chen et al., 2009; Eng and Borenstein, 2019).

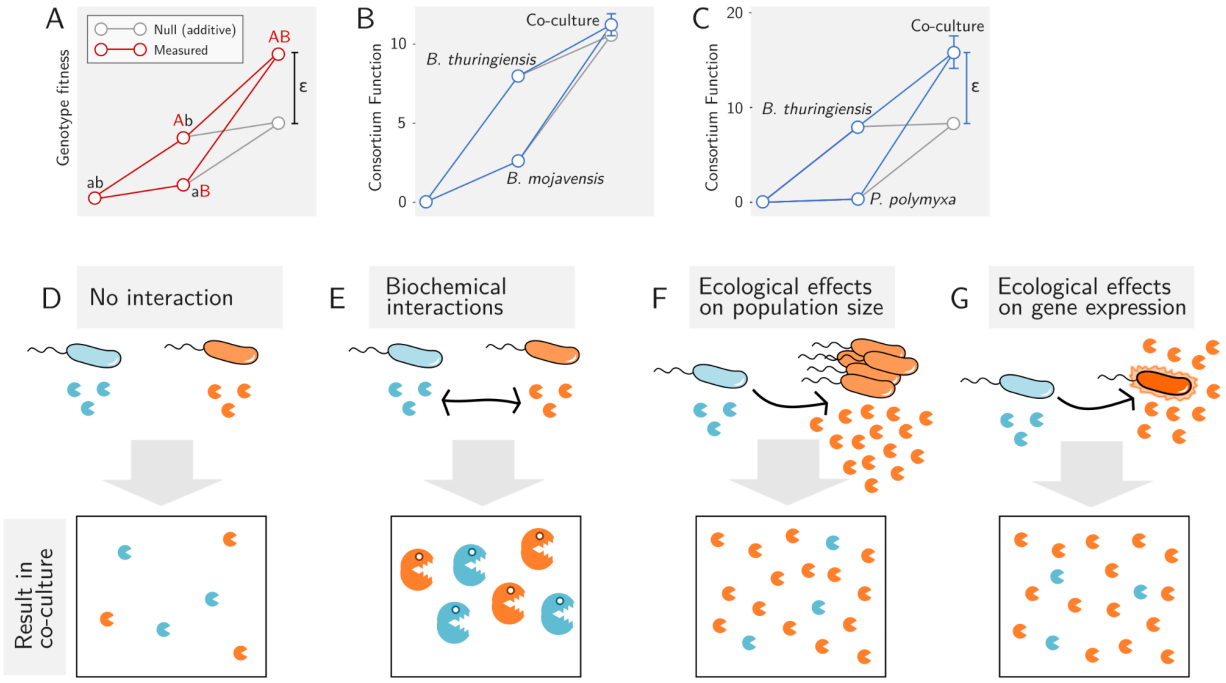

Fig. 1. Species interactions create non additive effects on community function. (A) In population genetics, two mutations A and B are said to interact when their phenotypic effects do not combine additively (or multiplicative, depending on the scale). This interaction is quantified by the deviation from additivity (referred to as the epistasis, ε). (B) 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., 2019)). 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 instances, 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).

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 cells 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 (Sanchez-Gorostiaga et al., 2019). 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.

To illustrate this idea, in Fig. 1B-C, we present a recent empirical example of a simple structure-function landscape. In this example, drawn from ref. (Sanchez-Gorostiaga et al., 2019), 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 (Sanchez-Gorostiaga et al., 2019). 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 Fig. 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 (Gowda et al., 2022). Interestingly though, other genotype pairs in the starch degrading communities, deviated markedly from the additive model (Fig. 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 (Sanchez-Gorostiaga et al., 2019).

What is the mechanistic basis of these pairwise interactions? In general, functional interactions may arise from three different mechanisms (Fig. 1D-F) (Sanchez-Gorostiaga et al., 2019). 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 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 (Kim et al., 2014). 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 (Mickalide and Kuehn, 2019). 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 Fig. 1 D-F can be separated empirically (Sanchez-Gorostiaga et al., 2019).

High-Order Functional Interactions. In communities with more than two species, functional interactions may be more complex than pairwise (Guo and Boedicker, 2016; Mickalide and Kuehn, 2019; Sanchez-Gorostiaga et al., 2019; Senne de Oliveira Lino et al., 2021). Consider, for instance, the example provided in Fig. 2, where the structure-function landscape comprising every combinatorial consortia of three amylolytic bacteria is given (Sanchez-Gorostiaga et al., 2019). 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. Yet, 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 (Poelwijk et al., 2019; Sanchez, 2019).

Besides the example discussed above, high-order functional interactions (HOFIs) have been observed in the production of ethanol by sugarcane biorefinery consortia (Senne de Oliveira Lino et al., 2021), the extension of a host lifespan by Drosophila gut microbiome consortia (Gould et al., 2018), the metabolic activity of synthetic consortia (Guo and Boedicker, 2016), and, more recently, gene expression in simple defined communities (Morin et al., 2022). Just as they do in fitness landscapes, high-order functional interactions 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 (Senne de Oliveira Lino et al., 2021). Based on pairwise interactions alone, we would have expected that as bacterial biodiversity increases in our bioreactors the ethanol yield would have collapsed. Yet, the opposite was true, and while most pairs of bacteria had negative effects in the ethanol yield, this detrimental effect vanished as communities increased in richness, reaching average levels that were comparable to those of pure yeast monocultures (Senne de Oliveira Lino et al., 2021). Despite this and other recent attempts to characterize high-order functional interactions (Eble et al., 2021; Gould et al., 2018; Sanchez-Gorostiaga et al., 2019), 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.

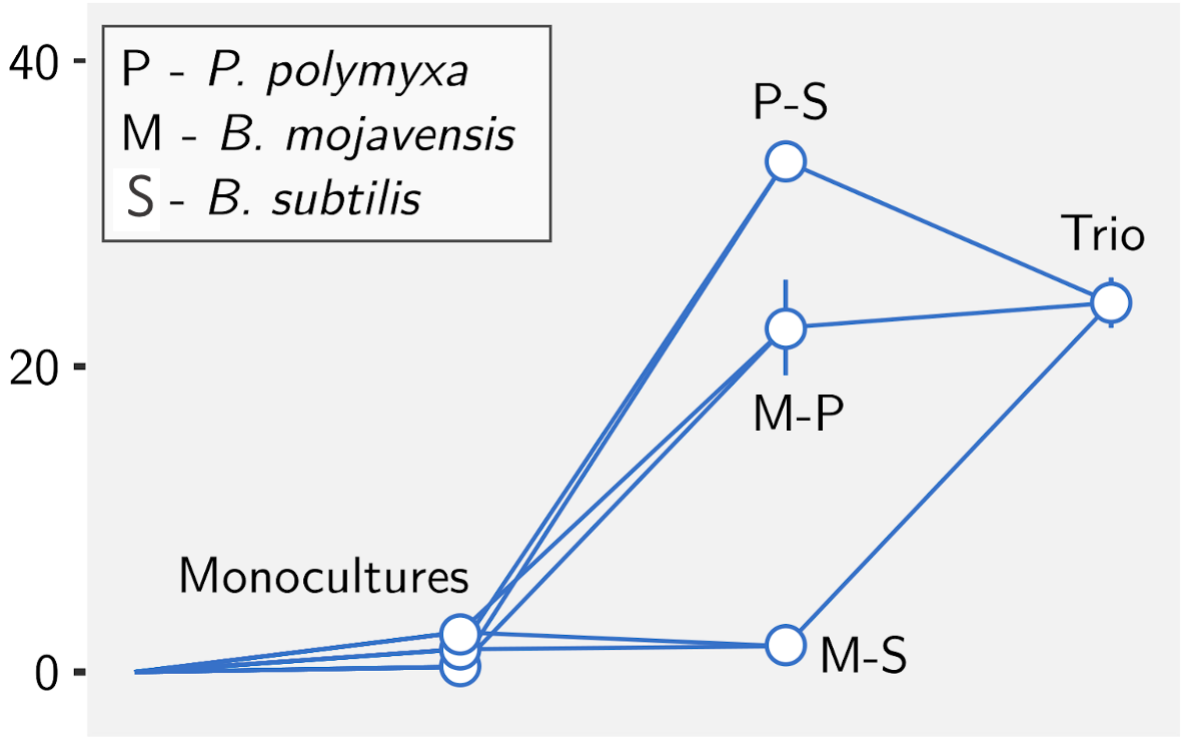

Fig. 2. High-order Functional Interactions in microbial consortia. We show an example of a third-order interaction that shapes the function of microbial consortia, in this case leading to diminishing returns. Adding either *B. subtilis* (S) or *B. mojavensis* (M) to a monoculture of *P. polymyxa* (P) dramatically enhances its function through a pairwise functional interaction. Yet, 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).

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 (Sanchez-Gorostiaga et al., 2019), so 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 (Reddy and Desai, 2021) (Fig. 3A).

Can we extend this way of partitioning interactions to microbial consortia? In recent work, we have found that, indeed, the functional effects of adding a species to a consortium does often scale linearly with the function of the background consortium, in a way that is very similar to what has been observed in genetic systems (Díaz-Colunga et al., 2022). Examples include diminishing returns, as well as increasing costs, accelerated returns, and others (Fig. 3B). 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 (Díaz-Colunga et al., 2022). Importantly, different species within a consortium tend to have different "Functional Effect Equations" describing their unique, global functional interaction patterns (Díaz-Colunga et al., 2022). 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 interesting 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 (Husain and Murugan, 2020; Otwinowski et al., 2018; Reddy and Desai, 2021; Wei and Zhang, 2019).

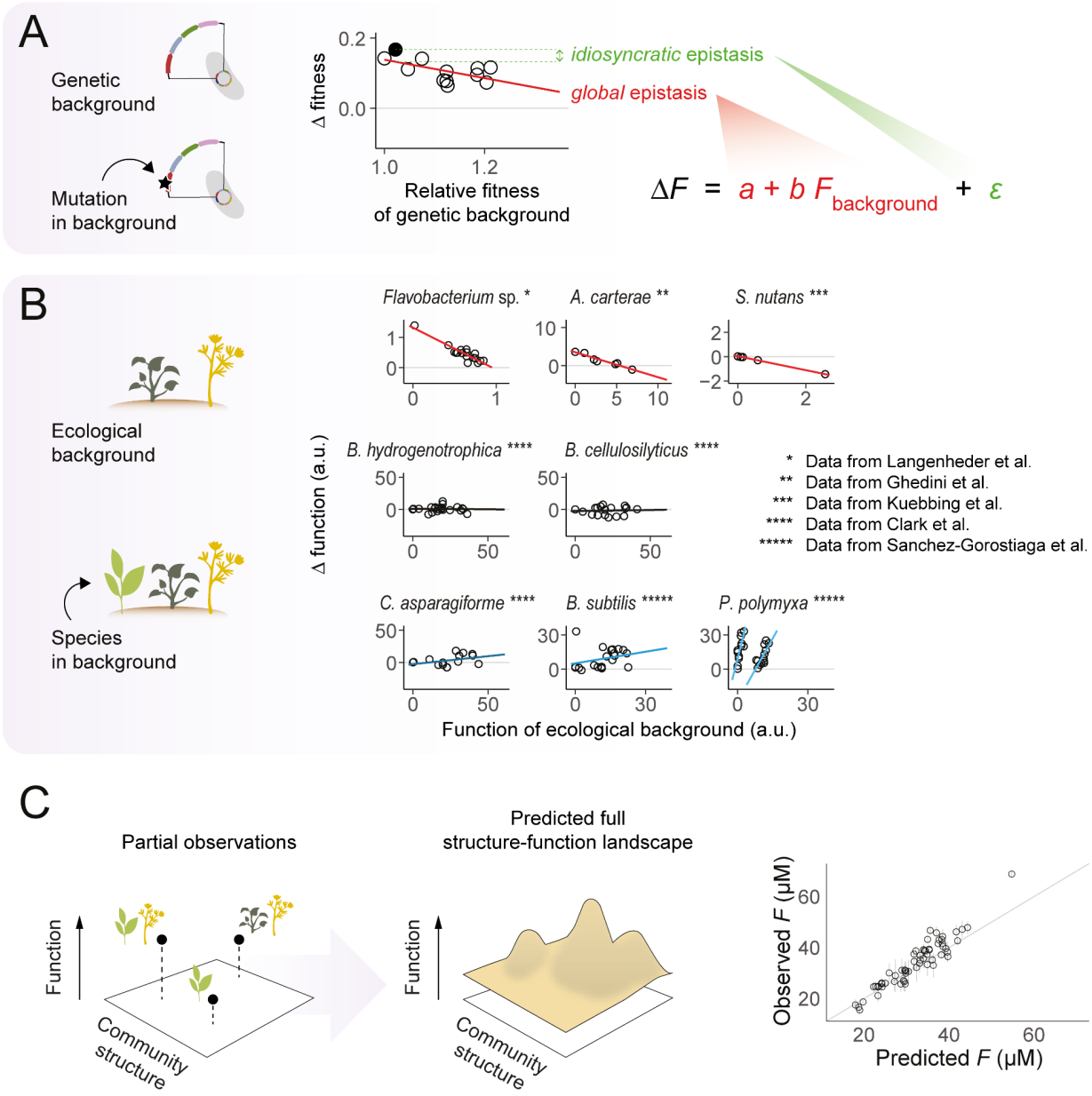

Fig. 3. An analog to global epistasis explains the functional effect of adding new species to microbial consortia. **(A)** Research in quantitative genetics has shown that the fitness effect of a mutation often depends linearly on the fitness of the genetic background where it arises. Epistasis can thus 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 ref. (Khan et al., 2011). **(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*) often scales linearly with 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) (Díaz-Colunga et al., 2022). In some cases, the functional effect of a species may be dominated by an idiosyncratic component rather than a global one (black lines). Data corresponds to butyrate production by synthetic gut microbial communities (Baranwal et al., 2022), biomass in plankton communities (Ghedini et al., 2022), above ground biomass in multi-species plant communities (Kuebbing et al., 2015), xylose oxidation by soil bacterial communities (Langenheder et al., 2010) and amylase secretion in bacterial consortia (Sanchez-Gorostiaga et al., 2019) **(C)** These functional trends can be exploited to predict the function of any combinatorial assemblage of species, and thus reconstruct entire ecological structure-function landscapes from a small subset of empirical observations. Data corresponds to the amount of pyoverdine secreted from newly assembled microbial consortia, whose function was predicted from a simple model that concatenates the Functional Effect Equations (FEEs) of each species as discussed in (Díaz-Colunga et al., 2022).

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: that 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 (Romero et al., 2013; Tareen et al., 2022; Tonner et al., 2021). Adapting and applying these methodologies to microbial consortia is an exciting prospect (Baranwal et al., 2022), 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 (Díaz-Colunga et al., 2022). This very simple approach 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 (Baranwal et al., 2022), and the success of earlier regression-based approaches to predict the landscape of small consortia (Chen et al., 2009) 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 to simple descriptions. Indeed, our recent work suggests that taking this perspective can uncover simple rules for mapping genomes to phenotypes (Gowda et al., 2022) and community composition to emergent function (Díaz-Colunga et al., 2022). 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 (Aguilar-Rodríguez et al., 2017; Nahum et al., 2015). 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 (Poelwijk et al., 2011). 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 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 (George and Korolev, 2021; Wittmann et al., 2021), 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., 2021). For example, consider one configuration of a community that gives rise to a function which 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 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 (Otwinowski and Plotkin, 2014) 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 which 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 (Kauffman and Weinberger, 1989; du Plessis et al., 2016), 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 (Husain and Murugan, 2020). 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 which 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 (Husain and Murugan, 2020; Leo-Macias et al., 2005). 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 even though the impact of each mutation on *physical* locations is roughly additive, their impacts on thermal stability or catalysis are epistatic (Otwinowski et al., 2018). 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 (Husain and Murugan, 2020). 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 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 abundances of genotypes as analogs to physical locations of atoms in the protein. In this case, a soft mode would manifest as a coherent variation in abundances 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 (Frentz et al., 2015; Hekstra and Leibler, 2012), and more recently also in host associate microbiomes (Raman et al., 2019). 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.

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. While we hope we have convinced the reader that learning the map between community composition and function may have a transformative impact in 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 (Fig. 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 which eliminate the target molecule and which 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 ***r***(.) and ***k***(.) denote the dynamical equations governing the temporal evolution of $\mathbf{x_g}$ and **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 **z**. This reflects the fact, which should be true for many community-level traits, that the function of 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 **z**.

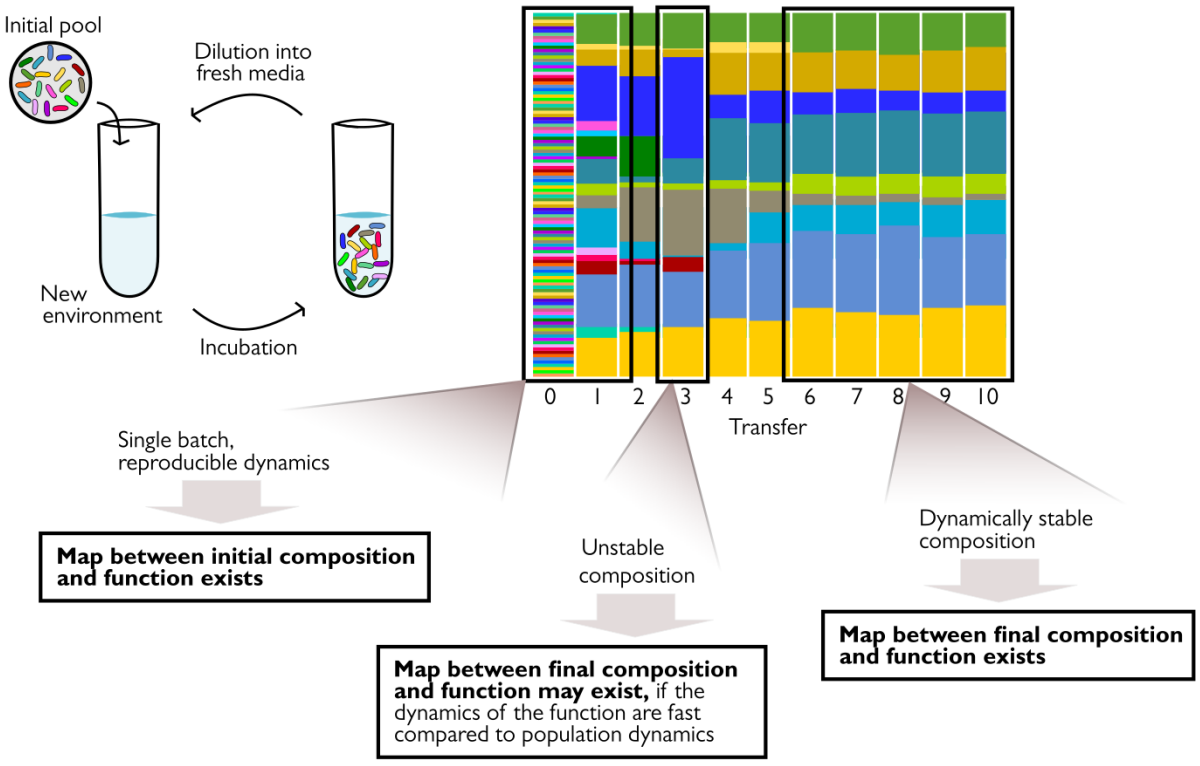

Fig. 4. Conditions for the existence of an ecological structure-function map. For illustrative purposes, we use as an example a hypothetical case where we stabilize a community from a diverse initial pool of species through periodic transfers in the laboratory, and we measure an enzymatic function akin to the amylolytic activity discussed as an example in the main text (see also Fig. 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 initial composition to function, assuming that the ecological dynamics are reproducible; 2) if we map final composition to function in a dynamically stable community, e.g. transfers 6-10; and 3) if we map

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 to population dynamics.

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. Yet, 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 $\lambda(E,\mathbf{z})=\lambda(\mathbf{z})E$. 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 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 steady-state. For many biotechnological applications, communities may be maintained in (or close to) steady-state by either placing them in a continuous culture device or through serial passaging (Fig. 4). In chemostats, both species composition and all environmental

parameters should reach 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" (Doulcier et al., 2020; Xie et al., 2019), at least when the passaging is done under constant conditions (Chang et al., 2021b; Estrela et al., 2021a, 2021c; Goldford et al., 2018; Sánchez et al., 2021).

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 (Fig. 4). 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 (Enke et al., 2019). 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 (Estrela et al., 2021a; Goldford et al., 2018) and detailed more broadly in surveys of the functional classes of bacteria in the marine

microbiome (Louca et al., 2016, 2018). Likewise, host associated communities also exhibit highly conserved metagenomic structure from host to host (Louca et al., 2016, 2018), suggesting that the functional landscape is a well-defined object with structure being tightly and reliably linked to function. What remains is to learn this mapping quantitatively and to leverage that knowledge to design and predict community behavior.

Conclusion. 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 exploration of fitness landscapes in evolutionary engineering and the exploration of structure-function landscapes may provide important insights to 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.

References

Aguilar-Rodríguez, J., Payne, J.L., and Wagner, A. (2017). A thousand empirical adaptive landscapes and their navigability. Nature Ecology & Evolution *1*, 0045. .

Alba, V., Carthew, J.E., Carthew, R.W., and Mani, M. (2021). Global constraints within the developmental program of the Drosophila wing. Elife *10*. https://doi.org/10.7554/eLife.66750.

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

Arnold, F.H. (2019). Innovation by Evolution: Bringing New Chemistry to Life (Nobel Lecture). Angew. Chem. Int. Ed Engl. *58*, 14420–14426. .

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. .

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. .

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.

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*. https://doi.org/10.7554/eLife.73870.

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. .

Beppler Casey, Tekin Elif, Mao Zhiyuan, White Cynthia, McDiarmid Cassandra, Vargas Emily, Miller Jeffrey H., Savage Van M., and Yeh Pamela J. (2016). Uncovering emergent interactions in three-way combinations of stressors. J. R. Soc. Interface *13*, 20160800. .

Berman, G.J., Choi, D.M., Bialek, W., and Shaevitz, J.W. (2013). Mapping the stereotyped behaviour of freely-moving fruit flies.

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. .

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. .

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

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

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. .

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. (2021a). Engineering complex communities by directed evolution. Nat Ecol Evol https://doi.org/10.1038/s41559-021-01457-5.

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. (2021b). Engineering complex communities by directed evolution. Nature Ecology & Evolution 1–13. .

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. .

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. .

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. .

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. .

Díaz-Colunga, J., Skwara, A., Vila, J.C.C., Bajić, D., and Sánchez, Á. (2022). Emergent ecosystem functions follow simple quantitative rules.

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 53433. .

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.

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

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. .

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

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

Estrela, S., Vila, J.C.C., Lu, N., Bajic, D., Rebolleda-Gomez, M., Chang, C.-Y., Goldford, J.E., Sanchez-Gorostiaga, A., and Sanchez, A. (2021a). Functional attractors in

microbial community assembly. Cell Systems *In Press*. .

Estrela, S., Sanchez-Gorostiaga, A., Vila, J.C., and Sanchez, A. (2021b). Nutrient dominance governs the assembly of microbial communities in mixed nutrient environments. Elife *10*, 2020.08.06.239897. .

Estrela, S., Sánchez, Á., and Rebolleda-Gómez, M. (2021c). Multi-Replicated Enrichment Communities as a Model System in Microbial Ecology. Front. Microbiol. *12*, 760. .

Frentz, Z., Kuehn, S., and Leibler, S. (2015). Strongly Deterministic Population Dynamics in Closed Microbial Communities. Phys. Rev. X *5*, 041014. .

George, A.B., and Korolev, K.S. (2021). Ecological landscapes guide the assembly of optimal microbial communities.

Ghedini, G., Marshall, D.J., and Loreau, M. (2022). Phytoplankton diversity affects biomass and energy production differently during community development. Functional Ecology *36*, 446–457. https://doi.org/10.1111/1365-2435.13955.

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. .

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 https://doi.org/10.1038/nature24287.

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. .

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. U. S. A. *115*, E11951–E11960. .

Gowda, K., Ping, D., Mani, M., and Kuehn, S. (2022). Genomic structure predicts metabolite dynamics in microbial communities. Cell *0*. https://doi.org/10.1016/j.cell.2021.12.036.

Guo, X., and Boedicker, J. (2016). High-Order Interactions between Species Strongly Influence the Activity of Microbial Communities. Biophys. J. *110*, 143a. .

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

Hekstra, D.R., and Leibler, S. (2012). Contingency and statistical laws in replicate

microbial closed ecosystems. Cell *149*, 1164–1173. .

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. .

Husain, K., and Murugan, A. (2020). Physical constraints on epistasis. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msaa124.

Jiang, Y., Dong, W., Xin, F., and Jiang, M. (2020). Designing Synthetic Microbial Consortia for Biofuel Production. Trends Biotechnol. *0*. https://doi.org/10.1016/j.tibtech.2020.02.002.

Jordan, D., Kuehn, S., Katifori, E., and Leibler, S. (2013). Behavioral diversity in microbes and low-dimensional phenotypic spaces. Proc. Natl. Acad. Sci. U. S. A. *110*, 14018–14023. .

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. .

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. https://doi.org/10.1126/science.1203801.

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. .

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. .

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

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. .

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. .

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. .

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. .

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. Nature Ecology & Evolution https://doi.org/10.1038/s41559-018-0519-1.

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. .

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. .

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. .

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*, 2020.07.28.225987. .

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. .

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. .

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. U. S. A. *110*, 14592–14597. .

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. U. S. A. *108*, E1293–E1301. .

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.

Mueller Ulrich G., Juenger Thomas E., Kardish Melissa R., Carlson Alexis L., Burns Kathleen M., Edwards Joseph A., Smith Chad C., Fang Chi-Chun, Des Marais David L., and Shade Ashley Artificial Selection on Microbiomes To Breed Microbiomes That

Confer Salt Tolerance to Plants. mSystems *0*, e01125–21. .

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. U. S. A. *112*, 7530–7535. .

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

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

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

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. .

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. .

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. .

Raman, A.S., White, K.I., and Ranganathan, R. (2016). Origins of Allostery and Evolvability in Proteins: A Case Study. Cell *166*, 468–480. .

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*. https://doi.org/10.1126/science.aau4735.

Reddy, G., and Desai, M.M. (2021). Global epistasis emerges from a generic model of a complex trait. Elife *10*, 2020.06.14.150946. .

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 Fact. *18*, 35. .

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

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

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. .

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

Sanchez, A. (2019). Defining Higher-Order Interactions in Synthetic Ecology: Lessons from Physics and Quantitative Genetics. Cell Syst *9*, 519–520. .

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. .

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. .

Senay, Y., John, G., Knutie, S.A., and Brandon Ogbunugafor, C. (2019). Deconstructing higher-order interactions in the microbiota: A theoretical examination.

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. .

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

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. .

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

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

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. .

Tekin, E., Yeh, P.J., and Savage, V.M. (2018). General Form for Interaction Measures and Framework for Deriving Higher-Order Emergent Effects. Frontiers in Ecology and Evolution *6*, 166. .

Tekin Elif, Beppler Casey, White Cynthia, Mao Zhiyuan, Savage Van M., and Yeh Pamela J. (2016). Enhanced identification of synergistic and antagonistic emergent interactions among three or more drugs. J. R. Soc. Interface *13*, 20160332. .

Thommes, M., Wang, T., Zhao, Q., Paschalidis, I.C., and Segrè, D. (2019). Designing Metabolic Division of Labor in Microbial Communities. mSystems *4*. https://doi.org/10.1128/mSystems.00263-18.

Tonner, P.D., Pressman, A., and Ross, D. (2021). Interpretable modeling of genotype-phenotype landscapes with state-of-the-art predictive power.

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. .

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

Wei, X., and Zhang, J. (2019). Patterns and Mechanisms of Diminishing Returns from Beneficial Mutations. Mol. Biol. Evol. *36*, 1008–1021. .

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. .

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. .

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. .

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. .

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. U. S. A. *109*, 12254–12259. .

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

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

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. .

Yang, D., Park, S.Y., and Lee, S.Y. (2021). Production of Rainbow Colorants by Metabolically Engineered Escherichia coli. Adv. Sci. *8*, e2100743. .

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 Fact. *9*, 10. .