Title:
Directed evolution of microbial communities

Author's names:
Álvaro Sánchez*, Jean C.C. Vila, Chang-Yu Chang, Juan Diaz-Colunga, Sylvie Estrela and María Rebolleda-Gomez

Affiliations:
Department of Ecology & Evolutionary Biology and Microbial Sciences Institute, Yale University, New Haven, CT, USA. (All Authors)

Email Addresses & ORCID numbers:
AS: alvaro.sanchez@yale.edu. orcid.org/0000-0002-2292-5608
JCCV: jean.vila@yale.edu. orcid.org/0000-0003-0499-0200
C-YC: chang-yu.chang@yale.edu orcid.org/0000-0003-3659-5256
JD-C: juan.diazcolunga@gmail.com orcid.org/0000-0001-8995-4529
SE: sylvie.estrela@yale.edu orcid.org/0000-0003-0499-0200
MR-G: maria.rebolleda-gomez@yale.edu. orcid.org/0000-0002-3592-4479

Corresponding author contact information:
*Alvaro Sanchez. Applied Biosciences Building, Room 215. 840 West Campus Drive. Yale West Campus. West Haven, CT, 06516. alvaro.sanchez@yale.edu

(Short) Running title:
Directed evolution of microbial communities

ABSTRACT

Directed evolution is a form of artificial selection that has been used for decades to find biomolecules and organisms with new or enhanced functional traits. Directed evolution can be conceptualized as a guided exploration of the genotype-phenotype map, where genetic variants with desirable phenotypes are first selected and then mutagenized to search the genotype space for an even better mutant. In recent years, the idea of applying artificial selection to microbial communities has gained momentum. Here, we review the main limitations of artificial selection when applied to large and diverse collectives of asexually dividing microbes, and discuss how the tools of directed evolution may be deployed to engineer communities from the top-down. We conceptualize directed evolution of microbial communities as a guided exploration of an ecological structure-function landscape, and propose practical guidelines for navigating these ecological landscapes.

KEYWORDS

Directed evolution. Microbial communities. Artificial ecosystem-selection. Structure-function landscape. Collective community functions.

DIRECTED EVOLUTION OF ORGANISMS AND BIOMOLECULES

Evolution has given shape to all forms of life, and humans have harnessed its power for millennia. Our ancestors learned to domesticate animals, plants, and a wide range of microorganisms by artificial selection long before they were aware of evolution itself (27, 92). The revolution in our understanding of evolutionary biology, genetics, and molecular biology in the 19th and 20th Centuries, together with the development of novel genomic and molecular technologies (1, 31, 100, 102, 115), has allowed us to extend artificial selection beyond domestication and learn how to direct the evolution of biomolecules (12, 48), genetic circuits (126), microorganisms (80), and viruses (16), to improve their phenotypes and even to invent new ones (48, 81). A major advantage of using directed evolution to engineer biological systems from the top-down, as opposed to engineering them from the bottom-up, is that the latter works with already known parts and traits, whereas the former does not require a priori knowledge and therefore it may allow us to find entirely new pathways and novel functions, encoded in hitherto unexplored regions of the genotype space.

Beyond its practical utility, directed evolution can also lead to profound insights to fundamental biological principles. For instance, directed evolution in vitro and in silico has revealed principles of organization of genetic and metabolic networks (11, 34, 125, 126), and it has been instrumental to our growing understanding of the mapping between genotype and phenotype in biomolecules (14, 89, 108), metabolic pathways (79), and other cellular phenotypes (115). In turn, as our understanding of the genetic

basis of adaptation has improved, it has enabled us to design more efficient strategies and methods for directed evolution (1, 6, 7, 64, 89, 102, 128).

The genotype-phenotype map (also referred to as the fitness landscape in the context of directed evolution) is defined as the relationship between the DNA sequence of a gene (or a higher-level functional unit, such as a pathway or genetic network) and the magnitude of the quantitative trait(s) it codes for in a given environment. Directed evolution can be conceptualized as a guided exploration of this genotype-phenotype map, in search for genotypes of high or novel functions (89, 108). Traditionally, directed evolution starts by first generating a library of genetic variants (74). All variants are then scored for the phenotype under selection, and those that are closer to the desired value (hereafter referred to as "the fittest") are chosen for reproduction. From this selected group, a new generation of genotypic variants is created through random mutagenesis or recombination (74). This two-step process is generally applied iteratively for as many rounds as needed. In recent years, several techniques have been developed that increase the throughput of the process and reduce human intervention (31, 115), but the fundamental evolutionary process remains the same (64, 90).

ARTIFICIAL SELECTION ABOVE THE ORGANISM.

Given the growing appreciation of the many important roles that groups of individuals (populations, communities, and ecosystems) play in natural and technological processes (13, 49, 57, 63, 69, 72, 84, 97, 110, 120), it has been proposed that artificial selection may

be also applied above the level of the organism to engineer community-level functions from the top-down (25, 72, 104) (Fig. 1A). An additional method to engineer communities from the top down that has a long tradition in microbiology is the Enrichment Approach (Fig. 1B) (For some recent applications see (22, 38, 56, 60, 62, 77, 127)). This method has been reviewed elsewhere, and although we will refer to it in this piece, our focus is on artificial selection, which has been less-well studied despite its substantial promise.

Conditions for artificial selection: Heritability and variation. Artificial selection can be in principle applied to any level of biological organization, provided that the evolving units exhibit phenotypic variation along the axis of selection, and that a substantial fraction of this variation can be reliably passed from parents to offspring (66) (Fig. 1A). Selection can be very efficient at the organismal level, as organisms fulfill both criteria: the phenotypes of an organism are (at least in part) determined by its genotype, which is either partially or entirely passed from parent to offspring. This ensures that many phenotypes have a heritable component upon which selection can act. Whether the conditions that are required for natural selection (which, in addition to the two mentioned above, includes that phenotypic differences must be associated with fitness differences between the replication units (66)) are met by any supra-organismal entities in nature has remained controversial. However, there is solid empirical and theoretical evidence that those conditions can be met under artificial selection conditions (15, 26, 40, 43, 65, 111, 113, 114, 118, 119, 123).

Artificial selection of populations and small synthetic communities. The idea that groups of organisms could respond to group-level selection was first tested in small animal populations (111, 112), and two-species communities (40, 43). In these experiments the selection units were populations containing N~10 genetically diverse, sexually reproducing animals (belonging to the flour beetle *Tribolium* genus), which interacted exclusively with one another but not with individuals from other populations. These populations were scored for an emergent trait that was a property of the entire group, such as the total number of adult animals in the population after ~40 days of incubation. The best performing communities (i.e. the "parental" groups) were then selected and used as the genetic stock to "breed" a new generation of groups (i.e., the "offspring" groups; Fig. 1C-D). As controls, some of these studies established random selection lines (where the populations selected for reproduction were chosen randomly, without regard for their phenotype), as well as no-selection lines (where all parent populations were selected for reproduction and each seeded exactly one offspring population) (Fig. 1E-F).

In order to generate an offspring group, these studies employed two different strategies. One reproduction strategy was called the "propagule" method (43). In this method, a small (N~10) random subset of individuals from the selected parental population was introduced in the new habitat, acting as the inoculum for the offspring population (Fig. 1C). The second reproduction method is referred to as the "migrant pool" method (43).

It consists of pooling together all of the animals from the selected parental populations, and then selecting a small random subset (also N~10) from that pool to seed an offspring population in a new habitat (Fig. 1D).

All of these experiments found a robust directional change in the mean phenotype of the "population of populations" (hereafter the metapopulation) in response to selection at the population level. Indeed, the metapopulations in these experiments fulfilled the two conditions that are required for artificial selection to work. First, the authors found a significant between-population variation in the phenotype under selection, due to the combination of small, genetically diverse populations and sexual recombination. Follow-up studies also demonstrated that this variation had a heritable component, which stemmed from interactions between specific combinations of genotypes, which were directly responsible for the population level trait under selection (e.g. the number of adult individuals) (41, 43).

Artificial selection of microbial communities and ecosystems. In the early 2000s, artificial selection above the organismal level was extended from populations and small pairwise communities to entire microbial ecosystems. In a landmark set of studies (103, 104), Swenson and co-workers adapted the propagule and migrant pool strategies to select for microbial ecosystems with high scores in three emergent community-level traits: (i) the pH of the aquatic medium on which the ecosystems were growing, (ii) the collective degradation of 3-chloroaniline, a water contaminant, and (iii) an indirect

microbiome phenotype, such as the above-earth biomass of the plants on which those communities had been inoculated. Although these experiments were promising, the effect of selection was modest compared to the robust and large responses observed in animal populations (40, 43, 111, 112).

These studies were followed by a handful of additional artificial microbiome selection experiments, all of which adopted similar protocols and selection strategies. Using a migrant pool method, Panke-Buisse et al artificially selected for soil microbiomes that induced either early or late flowering in various genotypes of *Arabidopsis thaliana* and *Brassica rapa* (75, 76). This experiment found a strong and statistically significant relative difference between the mean flowering time of microbiomes that were selected for early vs. late flowering. However, both lines drifted over time and flowered later than the starting (non-selected) microbiomes. In a later study (15), Blouin et al used an experimental design with multiple artificial selection (as well as random selection) lines, and selected for low CO_2 emission in aquatic ecosystems. The amount of respiration was lower in the artificially selected lines than in the random controls. In both, however, the amount of CO_2 produced declined over time. More recent studies have attempted to select microbiomes that degrade extracellular polymers (18, 121), protect plants against drought (55, 71), alter the development of animal embryos (8), and facilitate the growth of a species that could not grow on its own (18). We believe it is fair to say that success has been mixed (some experiments succeeded while others failed or were inconclusive), and generally modest.

What limits the success of artificial selection at the community level? As we discuss above, artificial selection at the community level requires that communities exhibit variation on the selected trait, and that this trait is reliably passed from "parent" to "offspring" communities. With regard to the heritability of community-level traits, the method used to generate "offspring" communities from their "parents" is therefore critical (83). Due to the success of the propagule and migrant pool strategies in animal populations, both methods have been universally adopted in all microbial community-level selection studies we are aware of. There are, however, important quantitative differences between animal populations and microbial communities. First, microbial communities are generally several orders of magnitude larger. A conservative estimate of the number of bacterial cells that were used to inoculate each generation in previous experiments is $N \sim 10^6$, and the actual number is likely to have been several orders of magnitude higher (19). This large inoculum size could lower the amount of between-population variation, which is critical for selection to act on (19). In animal populations, sexual reproduction and a genetically diverse starting pool of animals also ensured a high between-population variation (Fig. 1C-D). By contrast, most microbes reproduce asexually (despite the potential for horizontal gene transfer) and recombination is generally rare, diminishing the potential to generate novel genotypes using standing genetic variation alone. Stochasticity and selection are both needed for an efficient exploration of the adaptive landscape (122).

With regard to between-ecosystem variation, Blouin et al have discussed the conflict that exists between this variation and selection (15). As discussed above, selection runs on phenotypic variance. Yet, as selection proceeds it will exhaust this variance, as it will inevitably eliminate alleles and species from the metapopulation (Fig. 2A). In the absence of mechanisms that regenerate between-community variation (more on this below), we should expect diminishing returns in artificial selection: the amount of heritable variation should decrease with every selection round, leading to an ever weakening response to selection. Re-generating this variation is thus critical if we want artificial selection to be successful beyond the first few rounds. In later sections of this paper we address how this variation may be replenished. (Fig. 2A)

A second important limitation of artificial selection at the level of communities or ecosystems is their inherently dynamic nature. In most artificial selection experiments, selection is applied to communities that are grown in serial batch mode (Fig. 2B). In the absence of selection, serial batch culture starts by seeding a habitat with individuals coming from a previous batch (the "parental" batch), and continues by letting them grow in an environment that is, in principle, identical or at least as similar as possible to the one in the previous generation. At the end of the batch-incubation time *t*, cells are again randomly drawn from the offspring batch to inoculate yet another habitat and continue the process (Fig. 2B). Within each batch incubation, all species grow and undergo an ecological succession (24, 29, 33) (Fig. 2C). These successions are not necessarily identical between parent and offspring batches, and neither are their

compositions at the end of their respective batch incubations (Fig. 2C). As we discuss below and have demonstrated elsewhere (19), this "generational instability" can have detrimental effects on heritability and severely limit the success of ecosystem-level selection in multispecies consortia (Box I).

Artificial selection as top-down engineering. In the very first paper on artificial ecosystem selection, Swenson et al already advanced the idea that artificial selection could be used to engineer microbiomes from the top-down (104). However, the main goal of this and other previous studies was not so much to engineer ecosystems, but to demonstrate the feasibility of ecosystem-level selection and to study its fundamental limits. Perhaps for that reason, most studies have focused on the directional response of the mean function to selection, generally by comparing it to a random selection control, and none of the microbiome selection experiments we are aware of has included a no-selection control. A no-selection control is in essence an exercise on ecological prospecting (18): one sets up a diverse set of enrichment communities, let them stabilize without mixing, and in the end choose whichever one has the most desirable trait. In the absence of sexual recombination, the advantages of selection over ecological prospecting are not obvious and therefore eco-prospecting represents a benchmark against the success of a selection strategy. Importantly, the ultimate goal of artificial selection as a means of top-down microbiome engineering is not to improve the mean function in the metacommunity, but to find a fitter microbiome than the best one we started with (19). As we advanced in the introduction, directed evolution at or below the organismal level seeks to find optima in the genotype-phenotype space. In what

follows, we argue that the directed evolution of microbial communities can similarly be conceptualized as a guided exploration of an ecological structure-function landscape: the map between community composition and community function (Box I).

DIRECTED EVOLUTION AS A GUIDED EXPLORATION OF THE ECOLOGICAL STRUCTURE-FUNCTION LANDSCAPE

Before we describe how directed evolution may help us explore the ecological structure-function landscape in search for communities with optimal traits, it is important to clarify what we mean by the ecological structure-function landscape and in which ways it differs from the genotype-phenotype map. This will help us better appreciate the differences that exist between directed evolution above and below the organismal level.

The dynamical ecological structure-function landscape of microbial communities. Much like a fitness landscape maps genotypes and phenotypes, the ecological structure-function landscape is a map between community composition (i.e. the vector of abundances of all taxa in the community) and the traits (or functions) of the community (see Box I). The idea that community composition impacts emergent or collective community functions is an old one in ecology (67, 85, 106, 107, 117). In recent years, the structure-function landscape of microbial communities has been explicitly formalized (45–47, 94, 95) and combinatorially explored (10, 28, 32, 45–47, 52, 54, 58, 94,

101), by mapping numerous different combinations of bacteria with one of more of their quantitative collective-level properties.

Perhaps the biggest practical difference between fitness landscapes and ecological structure-function landscapes is that, as described above, community composition changes within a batch, from the moment of inoculation to the point of harvesting. Moreover, the successional dynamics within a batch are not necessarily the same in the parent as in the offspring community, even in the absence of group-level selection (19). This means that the state of the community, which is defined by the vector of species abundances at the end of the batch incubation, will change over generations even when no artificial selection is applied (Fig. 2C). Eventually, the communities may converge to a state of "generational stability", which can be represented as a fixed point in their dynamical landscape (Fig. 3A).

It is pertinent to ask at this point whether a generationally stable state of reproducible successions is ever to be expected. Early work proposed that community assembly might be chaotic, so that communities that are seeded with slightly different initial compositions (due to unavoidable random sampling) would diverge in both composition and function over time (104). Enrichment experiments with multiple replicates (which are the equivalent of a "no-selection" control in artificial selection experiments (19, 111)) have found, however, that community assembly is not chaotic (30): replicate habitats that were seeded from the same inoculum generally adopted a

discrete set of alternative (generationally) stable states (30, 39). These experiments have mapped the basin of attraction of stable states in self-assembled communities, and even found stochastic transitions among them (30, 109). Other studies with synthetic or bottom-up communities have similarly found evidence of dynamical multi-stability in microbial communities (4, 20, 23, 36). Because different stable community states contain different species, they may also differ in a range of community-level properties and functions (37). These studies, along with related theoretical work (26), support that the structure-function landscape is more than a convenient metaphor, and that despite its limitations (see Box I and the section above) it is a practical and useful tool to help us think through the process of directed evolution of biological systems above the organism.

Directed Evolution of microbial communities: Methods to explore the ecological structure-function landscape. At the community level, directed evolution would start by creating a library of "generationally stable" communities that differ from each other in the collective, community-level trait under selection. The fittest community is then selected, and used to generate a new library of proximal "compositional variants". Those variants are propagated by serial batch culture until they are generationally stable, and the fittest amongst those is again selected so the process can be iterated as many times as needed (Fig. 3A). A key step in the development of directed evolution for protein and network engineering has been the invention of methods for gene diversification, which enable the exploration of the fitness landscape of the system

under selection (74). If we wish to apply directed evolution to microbial communities, we should similarly ask how exactly we can generate a library of compositional variants of a selected community? Many ideas and methods have been already tested empirically in enrichment-based approaches for the top-down engineering of microbial communities (22, 56, 60, 62, 77). Below, we discuss those and a few other possibilities (19).

Horizontal gene transfer and mobile elements (Fig. 3A)*:* Mobile elements can create new strains, whose contribution to community-level functions may differ from their ancestor. This suggests that adding mobile elements (e.g. bacteriophages and plasmids) to different communities in a metacommunity may stochastically lead to the appearance of new strains in the community and, therefore, to between-community variation in function. In a recent study, Quistad et al transplanted mobile elements from one community to another, and this process led to genetic changes (e.g. amplification of genes involved in Nitrogen metabolism) in the recipient species (78). These genetic changes were associated with functional changes at the community-level (e.g. biochemical rates of ammonification), demonstrating that stimulated species-level evolution can be a means to create compositional and functional variants of a successful community. More targeted tools to deliver plasmids to a stable microbiome by horizontal gene transfer have been developed in recent years (91).

Coalescence (Fig. 3B): A library of variants of the selected community may also be created by coalescing the selected, stable community with each of the non-selected ones (68, 86, 87, 99). Multiple theoretical and empirical lines of evidence show that mixing two communities together produces a new "offspring" community that resembles both of its parents both in composition and function (99, 105).

Horizontal migration (Fig. 3C): Randomly sampled species from one or more natural species pools could be added to the community to generate proximal variants. These invasive species may displace some of the resident taxa, or augment the community by fixing without driving others extinct (61). Even those species that do not fix may in principle push a community to an alternative state (4).

Selective knock-ins (Fig. 3D): Species that are deemed to have a beneficial effect on the selected function can be selectively added one at a time to the community, thus generating a library of compositional variants. To ensure that this added species will not be outcompeted by the resident species, an exclusive metabolic niche (i.e. a nutrient that the added species may utilize but most members in the invaded community do not) could be supplied together with the added species (96).

Bottlenecking (Fig. 3E): Variants of a selected community can be generated by subjecting it to multiple harsh bottlenecks, an approach known as dilution-to-extinction (22, 35, 53, 56, 62). If a sufficiently low number of cells is sampled from the community during each

bottleneck, the inherent sampling stochasticity will ensure that each of the variants will have a different community composition (56). The specific bottleneck size that will maximize functional variation between compositional variants can be determined empirically, as it is a function of the population size (19, 53).

Selective knockouts (Fig. 3F): Individual species from the resident community may also be selectively targeted for elimination, for instance by narrow spectrum antibiotics or bacteriophages (17, 70, 73, 124). This elimination will create community variants with different composition. In addition, bacteriophages and antibiotics also represent selective pressures that can alter the genotypic composition of communities (3, 116), presenting an additional mechanism to generate compositional variants.

Environmental pulse perturbations (Fig. 3G): In addition to adding and removing species, one may push communities into random directions of the ecological landscape by transiently changing the environment, for instance by altering nutrient composition, increasing the temperature, salinity or pH, or through other means. This will alter the fitness of all members of the community, allowing some of the rarer members to increase in abundance, and some of the more abundant members to decline. When the environment is returned to its normal state, the perturbed communities may converge to a different fixed point due to either species loss or to multistability. Environmental pulse perturbations can also result in phenotypic switching in microbial populations

(9). We speculate that cellular memory may allow us to push microbial communities to new functional states even in the absence of changes in their genotypic composition.

Environmental press perturbations (Fig. 3H): Finally, whereas all of the above strategies to create compositional variants involve the communities jumping to a new stable state, another possibility is to change the environment in a small but permanent way, thereby finding new stable states that did not exist before (60). We speculate that this may allow a finer control over the community composition, as the press-perturbations may be made, in principle, as small as desired, thus potentially allowing for very small changes in species abundances (60).

The above methods are not exhaustive. The reader will have noticed that the two methods used in artificial selection, the propagule and mixed-pool, are also means to explore the ecological structure-function landscape. The bottleneck method is essentially an extreme form of a propagule method, where a very small number of cells is chosen to seed the offspring generation. As for the migrant-pool method, it is also an example of coalescence, where more than two communities are mixed in equal ratios. Both of these methods can successfully create a new library of variants if the number of cells that is sampled is small enough (18).

Directed evolution of microbial communities: Selection and heritability at the community level. After a sufficiently heterogeneous library of variants has been

generated, we must reckon with the fact that not all of the variation will be heritable. For instance, some variation may come from measurement error in determining the function of each community. How does this non-heritable component affect selection? The community-level heritability quantifies the degree to which community functions are passed from a "parent"community to its "offspring". If heritability is very low, this means that there exists a low correlation between the parent and offspring community function. Therefore, a high community-level heritability is critical in order for directed evolution to work. To understand how this heritability is affected by different experimental and ecological processes let us first consider a metacommunity that is being passaged in absence of artificial community-level selection (Fig. 2B), until all communities are successionally stable.

Let us denote by f_x and f_y the experimentally measured functions of a "parent" community and its "offspring". By assumption, both communities are in the same generational equilibrium state X^*. The functions of both communities can be written as:

$$f_x = F(X^*) + \xi_x$$

$$f_y = F(X^*) + \xi_y ,$$

where $F(X^*)$ denotes the function associated to the equilibrium state X^* in the structure-function landscape, and ξ_x and ξ_y are uncorrelated random variables of zero mean and equal variance (σ_ξ^2). These two variables capture the effect of small stochastic

deviations in community composition from the true equilibrium due to drift (e.g. due to the stochastic sampling introduced by pipetting), as well as measurement error, environmental fluctuations, and other stochastic factors. If we regress the function of the offspring on the parent function across the entire metacommunity, the regression slope will be given by b=Cov(f_x,f_y)/σ_x^2, where σ_x^2 is the experimentally measured variance in the parent metacommunity. This will in turn be given by $\sigma_x^2 = \sigma_F^2 + \sigma_\xi^2$, where σ_F^2 represents the fraction of variance that is due to different communities in the parent metacommunity being in different equilibria, i.e. the variance in $F(X^*)$ over the metacommunity.

By assumption, each offspring community is fluctuating around the same dynamical equilibrium state as its parent. Therefore, the component of the variance that derives from different communities being in different steady states (σ_F^2) will be passed intact from the parent to the offspring metacommunity. By contrast, σ_ξ^2 is non-heritable, as it includes all of the sources of variation that are stochastic and are uncorrelated between parent and offspring communities, from drift in population dynamics to environmental fluctuations or measurement error. Because by assumption ξ_x and ξ_y are uncorrelated, Cov(f_x,f_y) = σ_F^2. The slope of the regression between parent and offspring function will thus be equal to b=σ_F^2/σ_x^2. It is straightforward to see that the slope b is equal to the fraction of the total variation in function across the parent metacommunity that is heritable, i.e. the community heritability h^2 (15, 42):

$h^2 = \sigma_F^2/\sigma_x^2 = \sigma_F^2/(\sigma_F^2 + \sigma_\xi^2) = (1 + \sigma_\xi^2/\sigma_F^2)^{-1}$.

The larger is σ_ξ^2 relative to σ_F^2, the weaker will be the response to selection. This highlights the detrimental role of all non-heritable components for community-level selection. The effect of pipetting errors has been recently examined (123), and other factors such as the importance of precise and accurate measurements of community-level functions, and of working with genetically diverse communities in equilibrium are also highlighted. Uneven spatial and temporal environmental conditions among the populations should be avoided.

SUMMARY AND OUTLOOK

Our motivation for writing this piece was to synthesize the differences and similarities that exist between directed evolution of biological systems above and below the organismal level. Just as directed evolution has been used to engineer proteins and strains, it may be used to engineer communities and organisms as well. The underlying idea is similar, but important differences exist. Unlike the genotype of an organism (which is stable throughout its lifespan) or of a molecule, the composition of a community changes not only during the successions within each batch, but also across successive batches. Only after the communities stabilize, their composition and collective properties can be reliably passed from parent to offspring communities. At that point, communities become a valid unit of selection and can be subject to directed evolution. In this paper, we have discussed how one may carry out such an experiment,

and provided ideas that we hope will be useful for other researchers as they design their own approaches.

Box I. The structure-function landscape.

In the text we argue that the functions of a community are determined by their structure. But can we be more precise? In order for the structure-function landscape to be meaningful, we must specify what exactly we refer to as the function and the structure. To address this question, it is helpful to think of a concrete example, which is inspired by recent experiments in our laboratory (94). Let us consider a community of species that is seeded into a habitat at time $\tau = 0$ and allowed to grow in it. Species abundances (given by abundance vector $X(\tau)$) will change over time according to $X'(\tau) = dX/d\tau = g(X(\tau);\Theta(\tau))$, where Θ represents the set of all environmental parameters. As the species grow in this habitat, some of these environmental parameters change as well. For instance, species may be secreting an enzyme to the environment. The concentration of this enzyme in the habitat ($C(\tau)$) will increase as a result of secretion, but it may also decline as the enzymes become degraded by proteases, or inactivated through other means. The concentration of enzyme $C(\tau)$ will thus be governed by a differential equation just as the species abundance does, where:

$$C'(\tau) = h(X(\tau), X'(\tau), C(\tau); \Theta(\tau)) - C(\tau)m(X(\tau), C(\tau); \Theta(\tau))$$

Here, we have introduced a function $h(.)$, which captures the instantaneous rate of enzyme production in the community, and a second function $m(.)$ which reflects the instantaneous rate of degradation/dilution per enzyme. The total secretion rate will be a function of the abundances of all secreting species $X(\tau)$ (94) and, potentially, also of the growth rates of all species $X'(\tau)$, as the expression of most genes is regulated by growth rate (59). Finally, it may also be a function of the total concentration of the enzyme in the environment, as the byproducts of an enzyme are often inhibitors of its expression (44, 94). This rate may also depend on additional factors, such as the previous growth history of all cells, as gene expression can exhibit hysteresis (9, 88).

In this example, the enzyme concentration at the end of the batch-incubation (i.e. the "harvesting time") t is a community trait that one may want to select for, and therefore it is a potential function (F) of the community: $F=C(t)$. The abundance of all species at the time of harvest, $X(t)$, could be thought of as the structure of the community. Yet, the function F is in this case a cumulative property of the community over the incubation time t. In principle, F will causally depend not only on the species abundances at time t, but also on their entire ecological (i.e. succession) dynamics over the incubation time t, i.e. on $\{X(\tau),X'(\tau),\Theta(\tau)\}$, where $\tau \in (0,t)$ (Fig. 2B-C). Therefore, it is unclear that a structure-function landscape $F(X(t))$ that uniquely maps every structure (i.e. the abundance vector $X(t)$) with a function F even exists in this case.

This situation is resolved if we think of a population in equilibrium. It may sound strange to speak of communities being in equilibrium given that they are being serially passaged, and therefore engaging in a dynamical ecological succession during each incubation. A community is generationally stable when the ecological successions within subsequent incubations are identical (26), and as a result the abundance vectors at the end of the i-th and all subsequent incubation periods are also identical, i.e. if $X_i(t)=X_{i+j}(t)$ ($j > 1$). If two consecutive successions are identical, then $X_i(\tau)=X_{i+1}(\tau)$ for all $\tau \epsilon (0,t)$, and we should expect that the same equality will hold for $X'(\tau)$ and $\Theta(\tau)$, and therefore for $C(\tau)$. We thus argue that the structure-function landscape $F(X(t))$ is well-defined for communities that are generationally stable. It is important to note, however, that F is not causally determined by $X(t)$ but, rather, that $X(t)$ and $F(t)$ are both linked together through the same underlying dynamical process. Though there exists an association between both, this association does not immediately imply causation.

Finally, we should also note that the function F can (and often does) feed back to population dynamics. For instance, the concentration of extracellular enzymes will affect the fitness of different microbial strains (20, 93), and the per-capita contribution of each strain to this function may also be costly at the individual level (98). This can lead to high-function states to be fragile to invasion by "cheater" strains that have high fitness when the function is high, but which do not contribute to its production, thus avoiding the cost. Understanding how multi-level selection can contribute to avoid cheater invasion is an area of intense theoretical and experimental interest (2, 21, 50, 51,

82), which can help us design efficient methods of artificial community-level selection (5, 123)

DISCLOSURE STATEMENT

The authors declare no affiliations, memberships, funding, or financial holdings that may affect the objectivity of this review.

ACKNOWLEDGMENTS

The authors wish to thank Maddie Bender, Richard Li, Madeleine C. Mankowski, Brian Von Herzen, Molly Bassette, Julia Borden, Stefan Golfier, Paul G. Sanchez, Rachel Waymack, Xinwen Zhu, and Alicia Sanchez-Gorostiaga, Djordje Bajic, and other members of our research group for stimulating discussions about artificial ecosystem selection. We also wish to thank Wenying Shou and Paul Rainey for sharing their results with us, some of it before it was published, and for illuminating discussions on the topic of this review. This work was supported by the National Institutes of Health through grant 1R35 GM133467-01, and by a Packard Foundation Fellowship to AS. C-YC was supported by a graduate fellowship Government Scholarship to Study Abroad by the Government of Taiwan. MR-G was supported by a Gaylord Donnelley postdoctoral fellowship through the Yale Institute for Biospheric Studies.

LITERATURE CITED

1. Acevedo-Rocha CG, Hoebenreich S, Reetz MT. 2014. Iterative Saturation

Mutagenesis: A Powerful Approach to Engineer Proteins by Systematically Simulating Darwinian Evolution. In *Directed Evolution Library Creation: Methods and Protocols*, ed EMJ Gillam, JN Copp, D Ackerley, pp. 103–28. New York, NY: Springer New York

2. Ackermann M, Stecher B, Freed NE, Songhet P, Hardt W-D, Doebeli M. 2008. Self-destructive cooperation mediated by phenotypic noise. *Nature*. 454(7207):987–90
3. Alseth EO, Pursey E, Luján AM, McLeod I, Rollie C, Westra ER. 2019. Bacterial biodiversity drives the evolution of CRISPR-based phage resistance. *Nature*. 574(7779):549–52
4. Amor DR, Ratzke C, Gore J. 2020. Transient invaders can induce shifts between alternative stable states of microbial communities. *Sci Adv*. 6(8):eaay8676
5. Arias-Sánchez FI, Vessman B, Mitri S. 2019. Artificially selecting microbial communities: If we can breed dogs, why not microbiomes? *PLoS Biol.* 17(8):e3000356
6. Arkin AP, Youvan DC. 1992. An algorithm for protein engineering: simulations of recursive ensemble mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* 89(16):7811–15
7. Arnold FH, Georgiou G. 2003. *Directed enzyme evolution : screening and selection methods*. Totowa, N.J.: Humana Press
8. Arora J, Brisbin MM, Mikheyev AS. 2019. *The microbiome wants what it wants: microbial evolution overtakes experimental host-mediated indirect selection*. Work. Pap.
9. Axelrod K, Sanchez A, Gore J. 2015. Phenotypic states become increasingly

sensitive to perturbations near a bifurcation in a synthetic gene network. *Elife*. 4:

10. Ayed L, Achour S, Khelifi E, Cheref A, Bakhrouf A. 2010. Use of active consortia of constructed ternary bacterial cultures via mixture design for Congo Red decolorization enhancement. *Chem. Eng. J.* 162(2):495–502
11. Barve A, Wagner A. 2013. A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature*. 500(7461):203–6
12. Beaudry AA, Joyce GF. 1992. Directed evolution of an RNA enzyme. *Science*. 257(5070):635–41
13. Ben-David Y, Moraïs S, Bayer EA, Mizrahi I. 2020. Rapid adaptation for fibre degradation by changes in plasmid stoichiometry within Lactobacillus plantarum at the synthetic community level. *Microb. Biotechnol.*
14. Bloom JD, Arnold FH. 2009. In the light of directed evolution: pathways of adaptive protein evolution. *Proc. Natl. Acad. Sci. U. S. A.* 106 Suppl 1:9995–10000
15. Blouin M, Karimi B, Mathieu J, Lerch TZ. 2015. Levels and limits in artificial selection of communities. *Ecol. Lett.* 18(10):1040–48
16. Burrowes BH, Molineux IJ, Fralick JA. 2019. Directed in Vitro Evolution of Therapeutic Bacteriophages: The Appelmans Protocol. *Viruses*. 11(3):
17. Chan BK, Turner PE, Kim S, Mojibian HR, Elefteriades JA, Narayan D. 2018. Phage treatment of an aortic graft infected with Pseudomonas aeruginosa. *Evol Med Public Health*. 2018(1):60–66
18. Chang C-Y, Osborne ML, Bajic D, Sanchez A. 2020. *Artificially selecting microbial communities using propagule strategies*. Work. Pap.

19. Chang CY, Vila JCC, Bender M, Mankowski MC, Li R, et al. 2020. Top-down engineering of complex communities by directed evolution. *BioRxiv*
20. Chen A, Sanchez A, Dai L, Gore J. 2014. Dynamics of a producer-freeloader ecosystem on the brink of collapse. *Nat. Commun.* 5:3713
21. Chuang JS, Rivoire O, Leibler S. 2009. Simpson's paradox in a synthetic microbial system. *Science*. 323(5911):272–75
22. Cortes-Tolalpa L, Jiménez DJ, de Lima Brossi MJ, Salles JF, van Elsas JD. 2016. Different inocula produce distinctive microbial consortia with similar lignocellulose degradation capacity. *Appl. Microbiol. Biotechnol.*
23. Dai L, Vorselen D, Korolev KS, Gore J. 2012. Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science*. 336(6085):1175–77
24. Datta MS, Sliwerska E, Gore J, Polz MF, Cordero OX. 2016. Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nat. Commun.* 7:11965
25. Day MD, Beck D, Foster JA. 2011. Microbial Communities as Experimental Units. *Bioscience*. 61(5):398–406
26. Doulcier G, Lambert A, De Monte S, Rainey PB. 2020. Eco-evolutionary dynamics of nested Darwinian populations and the emergence of community-level heredity. *Elife*, p. 53433
27. Driscoll CA, Macdonald DW, O'Brien SJ. 2009. From wild animals to domestic pets, an evolutionary view of domestication. *Proc. Natl. Acad. Sci. U. S. A.* 106 Suppl 1:9971–78

28. Eng A, Borenstein E. 2019. Microbial community design: methods, applications, and opportunities. *Curr. Opin. Biotechnol.* 58:117–28
29. Enke TN, Datta MS, Schwartzman J, Cermak N, Schmitz D, et al. 2019. Modular Assembly of Polysaccharide-Degrading Marine Microbial Communities. *Curr. Biol.* 29(9):1528–35.e6
30. Estrela S, Vila JCC, Lu N, Bajic D, Rebolleda-Gomez M, et al. 2020. *Metabolic rules of microbial community assembly*. Work. Pap.
31. Esvelt KM, Carlson JC, Liu DR. 2011. A system for the continuous directed evolution of biomolecules. *Nature*. 472(7344):499–503
32. Faith JJ, Ahern PP, Ridaura VK, Cheng J, Gordon JI. 2014. Identifying gut microbe-host phenotype relationships using combinatorial communities in gnotobiotic mice. *Sci. Transl. Med.* 6(220):220ra11
33. Fernández A, Huang S, Seston S, Xing J, Hickey R, et al. 1999. How stable is stable? Function versus community composition. *Appl. Environ. Microbiol.* 65(8):3697–3704
34. François P, Hakim V. 2004. Design of genetic networks with specified functions by evolution in silico. *Proc. Natl. Acad. Sci. U. S. A.* 101(2):580–85
35. Franklin RB, Mills AL. 2006. Structural and functional responses of a sewage microbial community to dilution-induced reductions in diversity. *Microb. Ecol.* 52(2):280–88
36. Friedman J, Higgins LM, Gore J. 2017. Community structure follows simple assembly rules in microbial microcosms. *Nat Ecol Evol*. 1(5):109
37. Fukami T, Dickie IA, Paula Wilkie J, Paulus BC, Park D, et al. 2010. Assembly

history dictates ecosystem functioning: evidence from wood decomposer communities. *Ecol. Lett.* 13(6):675–84

38. Gilmore SP, Lankiewicz TS, Wilken SE, Brown JL, Sexton JA, et al. 2019. Top-Down Enrichment Guides in Formation of Synthetic Microbial Consortia for Biomass Degradation. *ACS Synth. Biol.* 8(9):2174–85
39. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, et al. 2018. Emergent simplicity in microbial community assembly. *Science*. 361(6401):469–74
40. Goodnight CJ. 1990. Experimental Studies of Community Evolution I: The Response to Selection at the Community Level. *Evolution*. 44(6):1614–24
41. Goodnight CJ. 1990. Experimental studies of community evolution II: The ecological basis of the response to community selection. *Evolution*. 44(6):1625–36
42. Goodnight CJ. 2000. Heritability at the ecosystem level
43. Goodnight CJ. 2011. Evolution in metacommunities. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366(1569):1401–9
44. Gore J, Youk H, van Oudenaarden A. 2009. Snowdrift game dynamics and facultative cheating in yeast. *Nature*. 459(7244):253–56
45. Gould AL, Zhang V, Lamberti L, Jones EW, Obadia B, et al. 2018. Microbiome interactions shape host fitness. *Proc. Natl. Acad. Sci. U. S. A.* 115(51):E11951–60
46. Guo X, Boedicker JQ. 2016. The Contribution of High-Order Metabolic Interactions to the Global Activity of a Four-Species Microbial Community. *PLoS Comput. Biol.* 12(9):e1005079
47. Guo X, Boedicker J. 2016. High-Order Interactions between Species Strongly

Influence the Activity of Microbial Communities. *Biophys. J.* 110(3):143a

48. Hall BG. 1978. Experimental evolution of a new enzymatic function. II. Evolution of multiple functions for ebg enzyme in E. coli. *Genetics*. 89(3):453–65
49. Hansen JJ, Sartor RB. 2015. Therapeutic Manipulation of the Microbiome in IBD: Current Results and Future Approaches. *Curr. Treat. Options Gastroenterol.* 13(1):105–20
50. Hauert C, Holmes M, Doebeli M. 2006. Evolutionary games and population dynamics: maintenance of cooperation in public goods games. *Proc. Biol. Sci.* 273(1600):2565–70
51. Hauert C, Wakano JY, Doebeli M. 2008. Ecological public goods games: cooperation and bifurcation. *Theor. Popul. Biol.* 73(2):257–63
52. Herrera Paredes S, Gao T, Law TF, Finkel OM, Mucyn T, et al. 2018. Design of synthetic bacterial communities for predictable plant phenotypes. *PLoS Biol.* 16(2):e2003962
53. Ho K-L, Lee D-J, Su A, Chang J-S. 2012. Biohydrogen from cellulosic feedstock: Dilution-to-stimulation approach. *Int. J. Hydrogen Energy*. 37(20):15582–87
54. Hu J, Xue Y, Li J, Wang L, Zhang S, et al. 2016. Characterization of a designed synthetic autotrophic–heterotrophic consortia for fixing CO2 without light. *RSC Adv.* 6(81):78161–69
55. Jochum MD, McWilliams KL, Pierson EA, Jo Y-K. 2019. Host-mediated microbiome engineering (HMME) of drought tolerance in the wheat rhizosphere. *PLoS One*. 14(12):e0225933

56. Kang D, Jacquiod S, Herschend J, Wei S, Nesme J, Sørensen SJ. 2019. Construction of Simplified Microbial Consortia to Degrade Recalcitrant Materials Based on Enrichment and Dilution-to-Extinction Cultures. *Front. Microbiol.* 10:3010
57. Karkaria BD, Treloar NJ, Barnes CP, Fedorec AJH. 2020. From Microbial Communities to Distributed Computing Systems. *Frontiers in Bioengineering and Biotechnology*. 8:834
58. Kato S, Haruta S, Cui ZJ, Ishii M, Igarashi Y. 2005. Stable coexistence of five bacterial strains as a cellulose-degrading community. *Appl. Environ. Microbiol.* 71(11):7099–7106
59. Klumpp S, Zhang Z, Hwa T. 2009. Growth rate-dependent global effects on gene expression in bacteria. *Cell*. 139(7):1366–75
60. Kucharzyk KH, Crawford RL, Paszczynski AJ, Soule T, Hess TF. 2012. Maximizing microbial degradation of perchlorate using a genetic algorithm: Media optimization. *J. Biotechnol.* 157(1):189–97
61. Kurkjian H, Akbari MJ, Momeni B. 2020. *The impact of interactions on invasion and colonization resistance in microbial communities*. Work. Pap.
62. Lee D-J, Show K-Y, Wang A. 2013. Unconventional approaches to isolation and enrichment of functional microbial consortium--a review. *Bioresour. Technol.* 136:697–706
63. Lee HL, Shen H, Hwang IY, Ling H, Yew WS, et al. 2018. Targeted Approaches for In Situ Gut Microbiome Manipulation. *Genes* . 9(7):
64. Leemhuis H, Stein V, Griffiths AD, Hollfelder F. 2005. New genotype-phenotype

linkages for directed evolution of functional proteins. *Curr. Opin. Struct. Biol.* 15(4):472–78

65. Levin BR, Kilmer WL. 1974. Interdemic selection and the evolution of altruism: a computer simulation study. *Evolution*. 28(4):527–45
66. Lewontin RC. 1970. The Units of Selection. *Annu. Rev. Ecol. Syst.* 1(1):1–18
67. Loreau M, Naeem S, Inchausti P, Bengtsson J, Grime JP, et al. 2001. Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science*. 294(5543):804–8
68. Lu N, Sanchez-Gorostiaga A, Tikhonov M, Sanchez A. 2018. *Cohesiveness in microbial community coalescence*. Work. Pap.
69. Macia J, Manzoni R, Conde N, Urrios A, de Nadal E, et al. 2016. Implementation of Complex Biological Logic Circuits Using Spatially Distributed Multicellular Consortia. *PLoS Comput. Biol.* 12(2):e1004685
70. Maxson T, Mitchell DA. 2016. Targeted Treatment for Bacterial Infections: Prospects for Pathogen-Specific Antibiotics Coupled with Rapid Diagnostics. *Tetrahedron*. 72(25):3609–24
71. Mueller UG, Juenger TE, Kardish MR, Carlson AL, Burns K, et al. 2016. *Artificial Microbiome-Selection to Engineer Microbiomes That Confer Salt-Tolerance to Plants*. Work. Pap.
72. Mueller UG, Sachs JL. 2015. Engineering Microbiomes to Improve Plant and Animal Health. *Trends Microbiol.* 23(10):606–17
73. Nale JY, Redgwell TA, Millard A, Clokie MRJ. 2018. Efficacy of an Optimised

Bacteriophage Cocktail to Clear Clostridium difficile in a Batch Fermentation Model. *Antibiotics (Basel)*. 7(1):

74. Packer MS, Liu DR. 2015. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* 16(7):379–94
75. Panke-Buisse K, Lee S, Kao-Kniffin J. 2016. Cultivated Sub-Populations of Soil Microbiomes Retain Early Flowering Plant Trait. *Microb. Ecol.*
76. Panke-Buisse K, Poole AC, Goodrich JK, Ley RE, Kao-Kniffin J. 2015. Selection on soil microbiomes reveals reproducible impacts on plant function. *ISME J.* 9(4):980–89
77. Puentes-Téllez PE, Falcao Salles J. 2018. Construction of Effective Minimal Active Microbial Consortia for Lignocellulose Degradation. *Microb. Ecol.* 76(2):419–29
78. Quistad SD, Doulcier G, Rainey PB. 2020. *Experimental manipulation of selfish genetic elements links genes to microbial community function*. Work. Pap.
79. Raman S, Rogers JK, Taylor ND, Church GM. 2014. Evolution-guided optimization of biosynthetic pathways. *Proc. Natl. Acad. Sci. U. S. A.* 111(50):17803–8
80. Ratcliff WC, Denison RF, Borrello M, Travisano M. 2012. Experimental evolution of multicellularity. *Proc. Natl. Acad. Sci. U. S. A.* 109(5):1595–1600
81. Ratcliff WC, Herron MD, Howell K, Pentz JT, Rosenzweig F, Travisano M. 2013. Experimental evolution of an alternating uni- and multicellular life cycle in Chlamydomonas reinhardtii. *Nat. Commun.* 4:2742
82. Rauch J, Kondev J, Sanchez A. 2017. Cooperators trade off ecological resilience and evolutionary stability in public goods games. *J. R. Soc. Interface*. 14(127):

83. Raynaud T, Devers M, Spor A, Blouin M. 2019. Effect of the Reproduction Method in an Artificial Selection Experiment at the Community Level. *Frontiers in Ecology and Evolution*. 7:416

84. Regot S, Macia J, Conde N, Furukawa K, Kjellén J, et al. 2011. Distributed biological computation with multicellular engineered networks. *Nature*. 469(7329):207–11

85. Reich PB, Tilman D, Naeem S, Ellsworth DS, Knops J, et al. 2004. Species and functional group diversity independently influence biomass accumulation and its response to CO2 and N. *Proc. Natl. Acad. Sci. U. S. A.* 101(27):10101–6

86. Rillig MC, Antonovics J, Caruso T, Lehmann A, Powell JR, et al. 2015. Interchange of entire communities: microbial community coalescence. *Trends Ecol. Evol.* 30(8):470–76

87. Rillig MC, Tsang A, Roy J. 2016. Microbial Community Coalescence for Microbiome Engineering. *Front. Microbiol.* 7:1967

88. Robert L, Paul G, Chen Y, Taddei F, Baigl D, Lindner AB. 2010. Pre-dispositions and epigenetic inheritance in the Escherichia coli lactose operon bistable switch. *Mol. Syst. Biol.* 6:357

89. Romero PA, Arnold FH. 2009. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* 10(12):866–76

90. Romero PA, Krause A, Arnold FH. 2013. Navigating the protein fitness landscape with Gaussian processes. *Proc. Natl. Acad. Sci. U. S. A.* 110(3):E193–201

91. Ronda C, Chen SP, Cabral V, Yaung SJ, Wang HH. 2019. Metagenomic engineering of the mammalian gut microbiome in situ. *Nat. Methods*. 16(2):167–70

92. Ross-Ibarra J, Morrell PL, Gaut BS. 2007. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. Natl. Acad. Sci. U. S. A.* 104 Suppl 1:8641–48

93. Sanchez A, Gore J. 2013. feedback between population and evolutionary dynamics determines the fate of social microbial populations. *PLoS Biol.* 11(4):e1001547

94. Sanchez-Gorostiaga A, Bajić D, Osborne ML, Poyatos JF, Sanchez A. 2019. High-order interactions distort the functional landscape of microbial consortia. *PLoS Biol.* 17(12):e3000550

95. Senay Y, John G, Knutie SA, Brandon Ogbunugafor C. 2019. *Deconstructing higher-order interactions in the microbiota: A theoretical examination*. Work. Pap.

96. Shepherd ES, DeLoache WC, Pruss KM, Whitaker WR, Sonnenburg JL. 2018. An exclusive metabolic niche enables strain engraftment in the gut microbiota. *Nature*. 557(7705):434–38

97. Sheth RU, Cabral V, Chen SP, Wang HH. 2016. Manipulating Bacterial Communities by in situ Microbiome Engineering. *Trends Genet.* 32(4):189–200

98. Shibasaki S, Mitri S. 2019. *Controlling evolutionary dynamics to optimize microbial bioremediation*. Work. Pap.

99. Sierocinski P, Milferstedt K, Bayer F, Großkopf T, Alston M, et al. 2017. A Single Community Dominates Structure and Function of a Mixture of Multiple Methanogenic Communities. *Curr. Biol.* 27(21):3390–95.e4

100. Smith GP, Petrenko VA. 1997. Phage Display. *Chem. Rev.* 97(2):391–410

101. Stein RR, Tanoue T, Szabady RL, Bhattarai SK, Olle B, et al. 2018.

Computer-guided design of optimal microbial consortia for immune system modulation. *Elife*. 7:

102. Stemmer WP. 1994. Rapid evolution of a protein in vitro by DNA shuffling. *Nature*. 370(6488):389–91

103. Swenson W, Arendt J, Wilson DS. 2000. Artificial selection of microbial ecosystems for 3-chloroaniline biodegradation. *Environ. Microbiol.* 2(5):564–71

104. Swenson W, Wilson DS, Elias R. 2000. Artificial ecosystem selection. *Proc. Natl. Acad. Sci. U. S. A.* 97(16):9110–14

105. Tikhonov M. 2016. Community-level cohesion without cooperation. *Elife*. 5:

106. Tilman D. 1999. The ecological consequences of changes in biodiversity: A search for general principles. *Ecology*. 80(5):1455–74

107. Tilman D, Reich PB, Knops J, Wedin D, Mielke T, Lehman C. 2001. Diversity and productivity in a long-term grassland experiment. *Science*. 294(5543):843–45

108. Tracewell CA, Arnold FH. 2009. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr. Opin. Chem. Biol.* 13(1):3–9

109. Vila JCC, Liu Y-Y, Sanchez A. 2020. Dissimilarity–Overlap analysis of replicate enrichment communities. *ISME J.*

110. Vorholt JA, Vogel C, Carlström CI, Müller DB. 2017. Establishing Causality: Opportunities of Synthetic Communities for Plant Microbiome Research. *Cell Host Microbe*. 22(2):142–55

111. Wade MJ. 1976. Group selections among laboratory populations of Tribolium. *Proc. Natl. Acad. Sci. U. S. A.* 73(12):4604–7

112. Wade MJ. 1977. An experimental study of group selection. *Evolution*. 31(1):134–53

113. Wade MJ. 1978. A Critical Review of the Models of Group Selection. *Q. Rev. Biol.* 53(2):101–14

114. Wade MJ. 2016. *Adaptation in Metapopulations: How Interaction Changes Evolution*. University of Chicago Press. 260 pp.

115. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, et al. 2009. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*. 460(7257):894–98

116. Westra ER, van Houte S, Oyesiku-Blakemore S, Makin B, Broniewski JM, et al. 2015. Parasite Exposure Drives Selective Evolution of Constitutive versus Inducible Defense. *Curr. Biol.* 25(8):1043–49

117. Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, et al. 2016. Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J.*

118. Williams HTP, Lenton TM. 2007. Artificial selection of simulated microbial ecosystems. *Proc. Natl. Acad. Sci. U. S. A.* 104(21):8918–23

119. Wilson DS. 1992. Complex Interactions in Metacommunities, with Implications for Biodiversity and Higher Levels of Selection. *Ecology*. 73(6):1984–2000

120. Wolfe BE, Button JE, Santarelli M, Dutton RJ. 2014. Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. *Cell*. 158(2):422–33

121. Wright RJ, Gibson MI, Christie-Oleza JA. 2019. Understanding microbial community dynamics to improve optimal microbiome selection. *Microbiome*. 7(1):85

122. Wright S. 1982. The shifting balance theory and macroevolution. *Annu. Rev. Genet.* 16:1–19

123. Xie L, Yuan AE, Shou W. 2019. Simulations reveal challenges to artificial community selection and possible strategies for success. *PLoS Biol.* 17(6):e3000295

124. Yao J, Carter RA, Vuagniaux G, Barbier M, Rosch JW, Rock CO. 2016. A Pathogen-Selective Antibiotic Minimizes Disturbance to the Microbiome. *Antimicrob. Agents Chemother.* 60(7):4264–73

125. Yokobayashi Y, Arnold FH. 2005. A Dual Selection Module for Directed Evolution of Genetic Circuits. *Nat. Comput.* 4(3):245–54

126. Yokobayashi Y, Weiss R, Arnold FH. 2002. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. U. S. A.* 99(26):16587–91

127. Zanaroli G, Di Toro S, Todaro D, Varese GC, Bertolotto A, 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

128. Zhang Y-X, Perry K, Vinci VA, Powell K, Stemmer WPC, del Cardayré SB. 2002. Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature*. 415(6872):644–46

TERMS AND DEFINITIONS LIST

Artificial selection: The process of intervening in the natural reproduction cycle of organisms to favor those individuals which exhibit desirable traits.

Directed evolution: An iterative process of randomization and selection that is used to engineer biological systems from the top-down.

Fitness landscape: The map between the genotype of a gene or organism and its phenotype or fitness.

Ecological Structure-function landscape: The map between community composition and the function or functions associated to it.

Generational-stability: A state reached by serially passaged batch cultures where the ecological successions in successive batches converge to be identical.

Enrichment approach: A method to engineer communities from the top-down by growing environmental microbiomes in selective media.

FIGURES

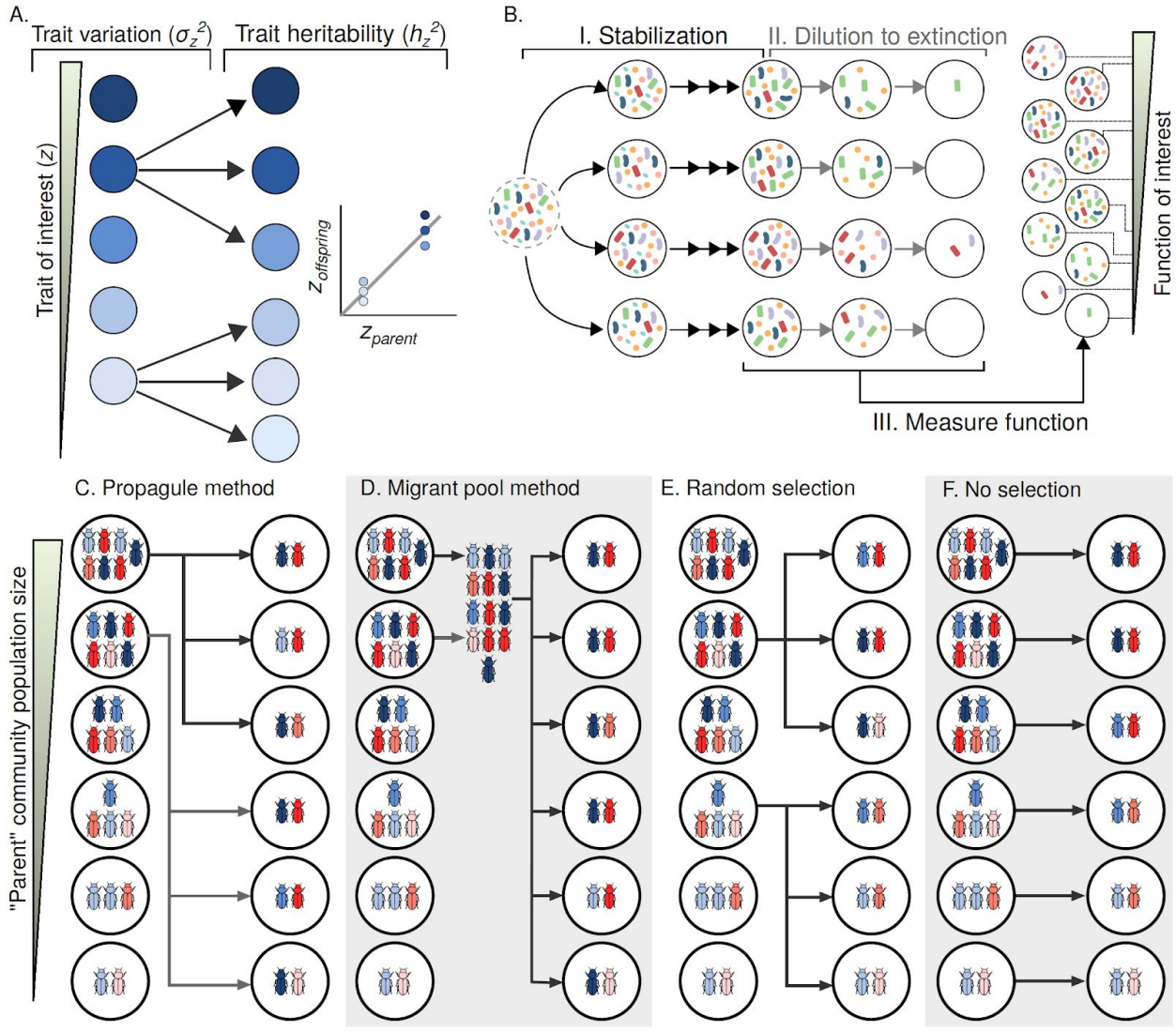

Fig. 1. Methods of top-down engineering above the individual organism**(A)** Any biological system can be subject to artificial selection as long as it exhibits variation along a trait of interest (z), and that trait is heritable, i.e., can be reliably passed into variants derived from it in a subsequent generation. **(B)** Typical workflow for an enrichment approach to engineering communities from the top-down. Multiple enrichment communities are set up by inoculating habitats from a species pool. Typically, the environment is selective for the desired function. The enrichment communities can be stabilized by serial passaging. Then, a severe bottleneck (dilution-to-extinction) is applied to subsample from the stable communities to find simpler communities that maintain the function, and the best amongst those is selected. **(C-F)** A depiction of the two main methods of artificial population-level selection, representing their original application in animal populations (40, 43, 111, 112): the propagule method and the migrant pool method, together with the random control and the no-selection control.

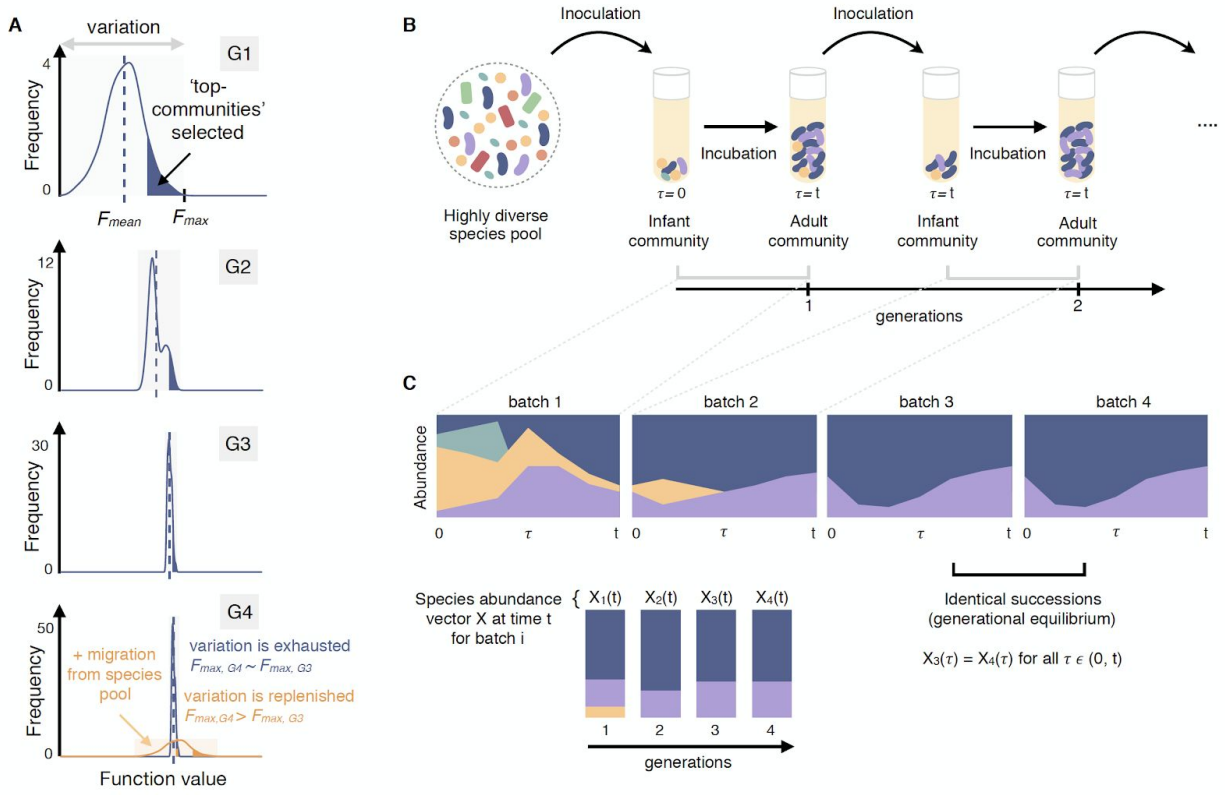

Fig. 2. Limitations of artificial selection at the level of communities. **(A)** Schematic illustrating the conflict between heritable variation and selection. As the top-performing communities get selected, the worst-performing communities get purged from the experiment, and as a result, F_{mean} increases and the amount of heritable variation decreases over generations (G). After multiple rounds of selection, and without any novel variants introduced, the heritable variation is fully exhausted and selection has nothing to act upon. Variation can be replenished by, for instance, introducing migrants from a species pool, which may allow communities to reach new function peaks (F_{max}). **(B)** Microbial community growth in serial batch culture (without selection). Communities are initially seeded from a highly diverse species pool into a new habitat ("infant" community), and then allowed to grow for an incubation time t (at which point they are an "adult" community). Without selection, a small and random fraction of the cells from the "adult" community are inoculated into a new habitat, forming a new 'infant'. This growth/dilution process is repeated multiple times. **(C).** Within each batch incubation, the species undergo an ecological succession as they grow and interact with each other. After multiple rounds of serial passage, communities reach a generationally stable equilibrium, which is seen when the species abundance vectors X during (and at the end of) the i-th and all subsequent incubation periods are identical, i.e. when $X_i(\tau)=X_{i+j}(\tau)$ for all $\tau \ \varepsilon \ (0,t)$. Without such "generational stability", community heritability is very low and the success of ecosystem-level selection at the level of communities is strongly reduced.

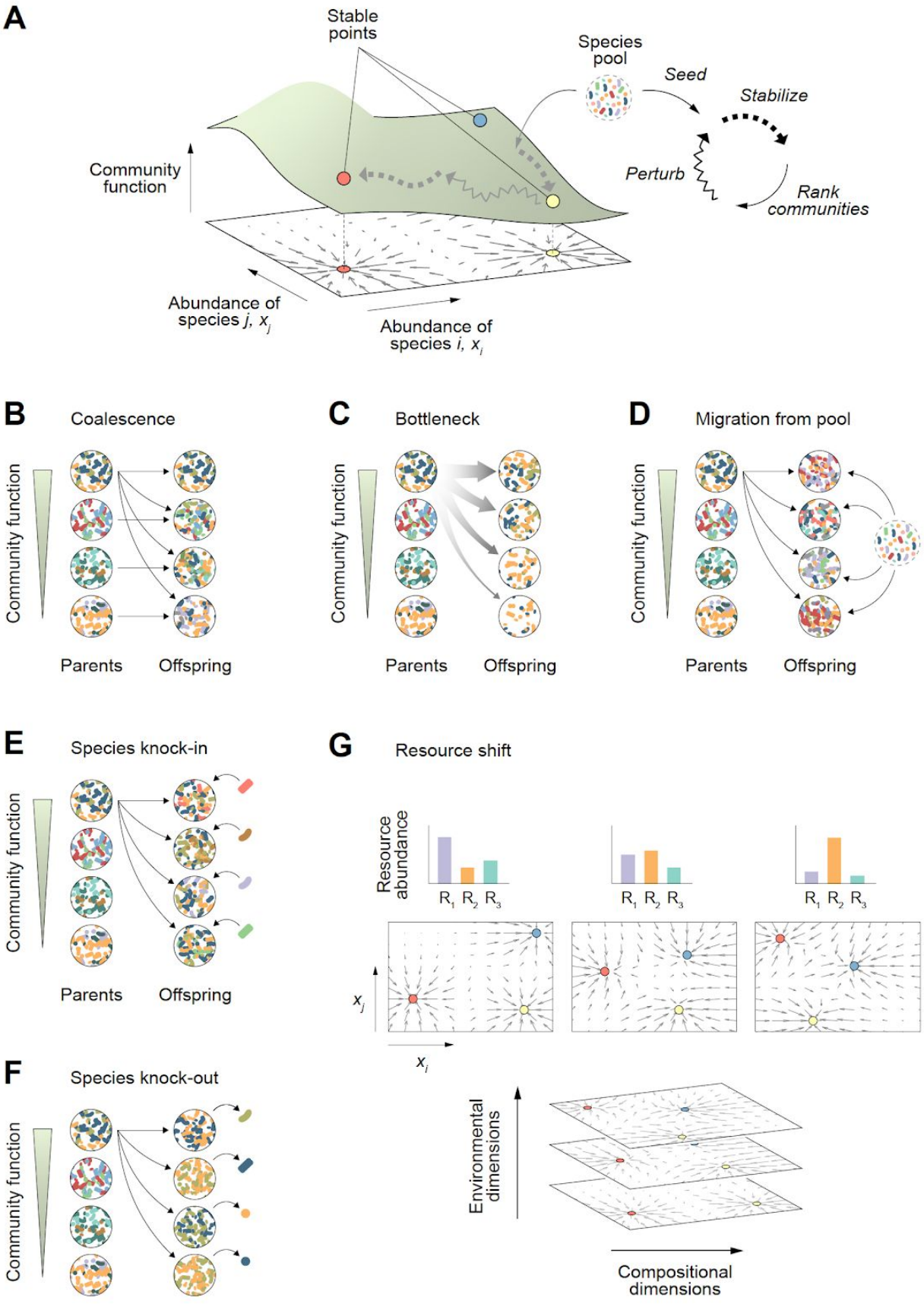
A
Stable points
Species pool
Seed
Stabilize
Perturb
Rank communities
Community function
Abundance of species j, x_j
Abundance of species i, x_i
B
Coalescence
Community function
Parents
Offspring
C
Bottleneck
Community function
Parents
Offspring
D
Migration from pool
Community function
Parents
Offspring
E
Species knock-in
Community function
Parents
Offspring
F
Species knock-out
Community function
Parents
Offspring
G
Resource shift
Resource abundance
R_1 R_2 R_3
R_1 R_2 R_3
R_1 R_2 R_3
x_j
x_i
Environmental dimensions
Compositional dimensions

Fig. 3. Directed Evolution as navigation of an ecological structure-function landscape. **(A)** In this cartoon, and for simplicity, we project the community function over an ecological space defined by the abundance of just two species (i and j). The depicted ecological dynamics are multistable, and communities converge to one of three different attractors (stable points), colored by red, yellow and blue circles. This ecological landscape can be navigated through an iterative process of perturbation, stabilization, ranking and selection. **(B-F)** Six different methods to create a library of "compositional variants" of the selected community. **(G)** Altering resource concentration can be a way to change the fitness of different species within the community and therefore of changing the composition of generationally stable communities (60).