Controlling microbial communities: a theoretical framework

Marco Tulio Angulo,[1,2,*] Claude H. Moog,[3] and Yang-Yu Liu[4,5,†]

[1]*National Council for Science and Technology (CONACyT), Ciudad de México 03940, México.*

[2]*Institute of Mathematics, Universidad Nacional Autónoma de México, Juriquilla 76230, México.*

[3]*Laboratoire des Sciences du Numrique de Nantes, UMR CNRS 6004, Nantes 44321, France.*

[4]*Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA.*

[5]*Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA.*

Microbial communities perform key functions for the host they associate with or the environment they reside in. Our ability to control those microbial communities is crucial for maintaining or even enhancing the well-being of their host or environment. But this potential has not been fully harvested due to the lack of a systematic method to control those complex microbial communities. Here we introduce a theoretical framework to rigorously address this challenge, based on the new notion of structural accessibility. This framework allows the identification of minimal sets of "driver species" through which we can achieve feasible control of the entire community. We apply our framework to control the core microbiota of a sea sponge and the gut microbiota of gnotobiotic mice infected with *C. difficile*. This control-theoretical framework fundamentally enhances our ability to effectively manage and control complex microbial communities, such as the human gut microbiota. In particular, the concept of driver species of a microbial community holds translational promise in the design of probiotic cocktails for various diseases associated with disrupted microbiota.

*Electronic address: `mangulo@im.unam.mx`

†Electronic address: `yyl@channing.harvard.edu`

INTRODUCTION

Microbial communities (MCs) play vital roles in the well-being of their hosts or environment [1–4]. For example, the human gut microbiota —the aggregate of microorganisms that resides in our intestines— plays a very important role in human physiology and diseases. Many gastrointestinal diseases such as inflammatory bowel disease, irritable bowel syndrome, and *C. difficile* Infection, as well as a variety of non-gastrointestinal disorders as divergent as autism and obesity, have been associated with disrupted gut microbiota [5–7]. For soil microbiota, its disruption may reduce the resistance of crops to diseases [8, 9]. For ocean microbiota, the disruption may impact global climate by altering carbon sequestration rates in the oceans [3, 4] Controlling these disrupted MCs to restore their healthy or normal states will be important for addressing a variety of challenges, including complex human diseases, global warming and sustainable agriculture [10–13].

In general, MCs can be affected by four types of control actions. Bacteriostatic agents and bactericides are antibiotics that decrease the abundance of targeted species* by inhibiting their reproduction or directly killing them, respectively [14]. Prebiotics are chemical compounds that selectively stimulate the growth of targeted species [15]. The fourth type of control action is transplantation, in which certain combination of species (e.g., all species from a "healthier" MC) is introduced to an MC at discrete intervention time instants. For the human gut microbiota, probiotics administration [16] and fecal microbiota transplantations (FMTs) [17] are examples of transplantations. To date, the empirical use of those control actions have already shown their efficacy for restoring disrupted MCs. For example, transplanting adequate soil species (i.e., soil inoculation) has been shown to be the key to successful restoration of terrestrial ecosystems [18]. And FMT is so far the most successful therapy in treating patients with recurrent *C. difficile* Infection [17]. But in order to harvest their full potential for restoring healthy MCs from disrupted ones, a systematic and rational control design method is required [10, 11, 19].

There are two fundamental challenges. First and foremost, we lack an efficient algorithm to identify minimal sets of species, which we call driver species, whose control can help us steer the whole community

* In this paper the term "species" is used in the general context of ecology, i.e., a set of organisms adapted to a particular set of resources in the environment. It doesn't necessarily represent the lowest major taxonomic rank. In fact, one could think of organizing microbes by strains, genus or operational taxonomical units as well.

to desired states. Consequently, we don't quite understand why microbiota transplantation works for some conditions but fails for others. Second, even if we have those driver species right in our hands, microbial dynamics are typically complicated and highly uncertain, rendering the systematic design of control strategies extremely difficult. Here, we developed a new control-theoretic framework that systematically address those two challenges. We first define the new notion of structural accessibility and derive a graph-theoretical characterization of it. This enables us to efficiently identify minimal sets of driver species of any MC purely from the topology of its underlying ecological network. Once the driver species are identified, the notion of structural accessibility also allows us to systematically design control strategies that steer an MC towards desired states. We demonstrate our framework in restoring the core microbiota of a sea sponge and the gut microbiota of gnobiotic mice infected by *C. difficile*.

PROBLEM STATEMENT

Define the state of an MC as the abundance profile $x \in \mathbb{R}^n$ of its n species. Consider that host or environmental factors influencing the MC remain constant during the time-interval that the control is to be performed (see Remark 1 in SI-1.2). Then, the evolution of the state along time t can be described by a general population dynamics model in the form of the ordinary differential equation

$$\dot{x}(t) = f(x(t)), \tag{1}$$

where the function $f : \mathbb{R}^n \to \mathbb{R}^n$ models the intrinsic growth and inter/intra-species interactions of the n species in the MC.

The exact functional form of f is often unknown because species can interact via a multitude of mechanisms [20], forming complex ecological networks [21] that give rise to various population dynamics models even at the scale of two species [22]. This leads us to only assume that (i) the topology of the underlying ecological network $\mathcal{G}$ of the MC is known; and (ii) f is some meromorphic function (i.e., its n entries $f_i(x)$ are the quotient of analytic functions of x). The ecological network of an MC has one state node $\mathsf{X} = \{\mathsf{x}_1, \cdots, \mathsf{x}_n\}$ per species and edges $(\mathsf{x}_j \to \mathsf{x}_i) \in \mathcal{G}$ iff the j-th species directly promotes or inhibits the growth of the i-th one (Fig.1A). Those interactions can be directly inferred from co-culture experiments [21, 23], or indirectly inferred from time-resolved metagenomics data using system identification tech-

niques [24, 25]. Assumption (ii) is very mild and is satisfied for most population dynamics models (Remark 2 in SI-1), including the classical Generalized Lotka-Volterra (GLV) model: $f(x) = \mathsf{diag}(x)[Ax + r]$ with $A \in \mathbb{R}^{n\times n}$ and $r \in \mathbb{R}^n$ the inter-species interaction matrix and the intrinsic growth rate vector, respectively [17, 19, 21, 26–32].

Consider now m different control inputs $u \in \mathbb{R}^m$ applied to an MC with population dynamics as Eq. (1), aiming to steer it from an initial towards a desired state (Fig. 1B). To describe which species are actuated by (i.e., directly susceptible to) the control inputs, we introduce the controlled ecological network $\mathcal{G}^{\mathsf{c}}$. This network contains the same set of state nodes X and inter-species interaction edges as $\mathcal{G}$, as well as an additional control node set $\mathsf{U} = \{\mathsf{u}_1, \cdots, \mathsf{u}_\mathsf{m}\}$ —one per control input— and edges of the form $(\mathsf{u}_j \to \mathsf{x}_i) \in \mathcal{G}^{\mathsf{c}}$ denote that the i-th species is actuated by the j-th control input (Fig.1A).

We consider two control schemes to analyze the effect of the control inputs on the state of the MC. The first control scheme models a combination of prebiotics ($u_i(t) > 0$) and bacteriostatic agents ($u_i(t) < 0$) as continuous control signals modifying the growth of certain species (Fig.1C):

$$\dot{x}(t) = f\left(x(t)\right) + g\left(x(t)\right)u(t), \quad t \in \mathbb{R}. \tag{2}$$

The second control scheme models a combination of probiotic administrations or microbiota transplantations ($u_i(t) > 0$) and bactericides ($u_i(t) < 0$) as impulsive control signals (Fig. 1D), which instantaneously modify the abundance of the actuated species at discrete time instants $\mathbb{T} = \{t_1, t_2, \cdots\}$:

$$\dot{x}(t) = f\left(x(t)\right) \text{ if } t \notin \mathbb{T}, \quad x(t^+) = x(t) + g\left(x(t)\right)u(t) \text{ if } t \in \mathbb{T}. \tag{3}$$

Here $x(t^+) = \lim_{\sigma \searrow t} x(\sigma)$ denotes the value of x "right after time t". Notice how the state $x(t)$ "jumps" if $u(t) \neq 0$ because of the instantaneous increase or decrease of the abundance of the actuated species (Fig. 1C). For simplicity's sake, we assume that the control signals are periodically applied every $\tau > 0$ time units (i.e., $t_{k+1} = t_k + \tau$ for all $k \in \mathbb{N}$).

In both control schemes, the entry $g_{ij}(x)$ of the susceptibility matrix $g(x) = [g_{ij}(x)] \in \mathbb{R}^{n\times m}$ describes the susceptibility of the i-th species to the j-th control input. If $g_{ij} \not\equiv 0$, then the i-th species is actuated by the j-th control input. The detailed mechanism of the susceptibility of a species to certain control actions is typically complicated and not exactly known. But we can just assume that g is some unknown meromorphic function.

In the trivial case when each species is directly actuated by an independent control input (e.g., $g(x)$ is the identity matrix $I_{n\times n}$), the state of the MC can of course be fully controlled, indicating that the set of all species is a trivial set of driver species. This control strategy is certainly overkill, since it requires as many controls as species in the MC. Next we show that knowledge of the topology of $\mathcal{G}^{\mathrm{c}}$ can actually be exploited to significantly reduce the number of driver species.

RESULTS

Given a particular pair of meromorphic functions $\{f, g\}$ in Eqs. (2) or (3), the topology of its controlled ecological network $\mathcal{G}_{f,g}$ is given by $(\mathsf{x}_j \rightarrow \mathsf{x}_i) \in \mathcal{G}_{f,g}$ iff x_j appears in the right-hand side of $\dot{x}_i$ for continuous control, or in the right-hand side of $\dot{x}_i$ or $x_i(t^+)$ for impulsive control. For both control schemes, $(\mathsf{u}_j \rightarrow \mathsf{x}_i) \in \mathcal{G}_{f,g}$ iff $g_{ij} \not\equiv 0$. Knowing only the topology of the ecological network of the MC, we are led to consider the class $\mathfrak{G}$ of all controlled dynamics of the form of Eqs. (2) or (3) with the same network topology $\mathcal{G}^{\mathrm{c}}$ (i.e., all meromorphic f's and g's such that $\mathcal{G}_{f,g} = \mathcal{G}^{\mathrm{c}}$, see SI-3.1).

Structural accessibility

To identify minimal sets of driver species, we introduce the notion of structural accessibility of $\mathfrak{G}$. The class $\mathfrak{G}$ is <u>structurally accessible</u> if at least one of its pairs $\{f, g\}$ lacks <u>autonomous elements</u> —that is, internal variables of the system, involving certain combination of species, that are completely unaffected by the control inputs (Definitions 1 and 3 in SI-2). We further proved that this condition implies that almost all other pairs in $\mathfrak{G}$ lack autonomous elements as well (Theorem 3 in SI-3). The absence of autonomous elements guarantees that the dimension of their set of reachable states equals the dimension of the state space itself, so the control actions are effective enough to locally steer and independently manipulate the whole state of the MC (Fig.2A). Therefore, if $\mathfrak{G}$ is structurally accessible, its corresponding set of actuated species is defined as a set of driver species, and the condition of structural accessibility of $\mathcal{G}^{\mathrm{c}}$ can be used to determine minimal sets of driver species.

By contrast, $\mathfrak{G}$ is not structurally accessible if all its pairs $\{f, g\} \in \mathfrak{G}$ have autonomous elements. This implies that, for all population dynamics f and all susceptibility matrices g, the set of states that can be reached by using some control input u is constrained to a low-dimensional manifold, representing

an underlying constraint between some species abundances that the control inputs cannot brake (Fig.2B). Consequently, those actuated species (encoded in the susceptibility matrix g) cannot be a set of driver species.

We also say that a pair $\{f, g\} \in \mathfrak{G}$ is structurally accessible if $\mathfrak{G}$ is structurally accessible. We proved that if a pair $\{f, g\}$ is structurally accessible but does have autonomous elements, then there always exists an infinitesimal deformation than can be made to $\{f, g\}$ such that the new pair $\{\tilde{f}, \tilde{g}\}$ is free of autonomous elements (Proposition 1 in SI-3). If one had originally chosen $\{f, g\}$ as an adequate dynamics model for a controlled MC, the new pair $\{\tilde{f}, \tilde{g}\}$ free of autonomous elements is as adequate as the original one, since it predicts essentially the same temporal behavior of the MC. Therefore, in practice, for a structurally accessible class of controlled microbial dynamics models $\mathfrak{G}$, those pairs with autonomous elements can be always circumvented.

Characterizing the structural accessibility of $\mathfrak{G}$ requires us to define the notion of autonomous element for impulsive control systems for the first time (Definitions 3 in SI-2), since this notion has been studied only for systems with continuous control inputs [33]. We then characterized the conditions for the absence of autonomous elements for a given pair $\{f, g\}$, finding that, surprisingly, such conditions are identical for systems with continuous and impulsive control (Theorem 2 and Remark 3 in SI-2). This indicates that, for controlling MCs, impulsive control actions can be as effective as continuous ones, making unnecessary the use of continuous control —which is hard to implement anyway for many MCs, especially the human gut microbiota. Note that other existing approaches that could steer an MC towards desired states, such as "clamping" the abundance of the species in the so-called Feedback Vertex Set of $\mathcal{G}$ to their values at the desired state [34], also require continuous actuation. This encouraging result assures us to further develop microbiome-based therapies such as probiotic cocktails and FMTs in an impulsive control manner.

Finally, we used the above obtained conditions to derive a graph-theoretical characterization of the structural accessibility of $\mathfrak{G}$ (Theorem 5 in SI-3). We proved that $\mathfrak{G}$ is structurally accessible if and only if (i) each state (species) node in $\mathcal{G}^{\mathsf{c}}$ is the end-node of a path that starts in the control input node set U; and (ii) $\mathcal{G}^{\mathsf{c}}$ does not have <u>pure dilations of the control inputs</u>. Let $\mathsf{S} \subseteq \mathsf{X}$ be a set of state nodes and denote its neighborhood set by $T(\mathsf{S}) \subseteq \mathsf{X} \cup \mathsf{U}$ (i.e., the set of all nodes that point to S). The controlled ecological network $\mathcal{G}^{\mathsf{c}}$ has a pure dilation of the control inputs (Definition 8 in SI-3) if there exists a set S of state nodes

such that $T(\mathsf{S})$ only contains control input nodes (i.e., $T(\mathsf{S}) \subseteq \mathsf{U}$) and the size of $T(\mathsf{S})$ is smaller than the size of S, i.e., $|T(\mathsf{S})| < |\mathsf{S}|$ (Fig.2B).

Finding minimal sets of driver species

The above graph-theoretical characterization of structural accessibility provides a systematic method to identify minimal sets of driver species for any MC that has known ecological network $\mathcal{G}$ and meromorphic functions for its controlled population dynamics model. Algorithmically, a minimal set of driver species can be obtained by choosing one node in each of the root Strongly Connected Components (SCCs) of $\mathcal{G}$ (Fig. 2A). Here, an SCC of $\mathcal{G}$ is a maximal subgraph such that there is a directed path between any two of its nodes. A root SCC is an SCC without incoming edges. If one node per root SCC is actuated by at least one independent control input, the resulting controlled ecological network $\mathcal{G}^{\mathsf{c}}$ cannot have pure dilations in the control input, rendering the set of all possible controlled dynamical systems $\mathfrak{G}$ structurally accessible. Note that the SCCs of general directed graphs can be found in linear time [35, pp. 552-557].

Notice that once $\mathcal{G}^{\mathsf{c}}$ satisfies the conditions for structural accessibility, this cannot be undone by adding new edges to the network. Thus, for example, if we find the driver species of an ecological network $\mathcal{G}$ that includes only the high-confidence interactions in an MC, the identified driver species remain driver species even if additional interactions exist. Note also that adding or removing self-loops do not change the driver species of $\mathcal{G}$, since its SCCs remain unchanged.

Designing control signals that steer MCs to desired states

Next we discuss the design of the control signals $u(t)$ that should be applied to those driver species to steer the entire MC towards some desired state. We will focus on feedback control signals $u(x(t))$, as they provide robustness against uncertainty in the system dynamics [36]. We consider two scenarios: when the controlled dynamics of the MC is approximately known, and when it is highly uncertain. In the first scenario, we show that the knowledge of the population dynamics can be used to "optimally" steer the MC towards a desired state. In the second scenario, the controller needs to be designed to cope with large uncertainties in the population dynamics of the MC.

If a structurally accessible pair $\{f, g\} \in \mathfrak{G}$ is known to adequately model the controlled population dynamics of the MC, then the absence of autonomous elements can be used to systematically build feedback

controllers $u(x(t))$ that steer the system towards desired states using the so-called "feedback linearization" methodology [33, pp. 119]. Such methodology yields controllers that steer the MC towards arbitrary desired states, excepting a set of "singularities" of zero measure. In addition of steering the MC to desired states, the feedback controller can be designed to satisfy certain optimality conditions using a linear quadratic regulator [36]. For example, the controller can provide an optimal tradeoff between the convergence rate towards the desired state x_f and the "control energy" that is spent (e.g., a proxy for total abundance of species that is transplanted or eliminated by bactericides), by minimizing the quadratic index

$$\int_0^\infty [z^\intercal(x(t))Qz(x(t)) + v^\intercal(t)Rv(t)]dt,$$

for continuous control. For impulsive control, the above integral should be replaced by a sum over $\mathbb{T}$. Here $z = \Phi(x) \in \mathbb{R}^n$ and $v \in \mathbb{R}^m$ are the linearized state and control inputs, respectively, produced by the feedback linearization methodology. The symmetric positive semi-definite matrices $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{m\times m}$ are design parameters: "large" R highly penalizes the control effort, while "large" Q enhances the convergence rate.

As a concrete example, consider the toy MC shown in Fig.1. This small MC contains 3 species and the underlying ecological network $\mathcal{G}$ is depicted in Fig.1A. $\mathcal{G}$ has only one root SCC and this root SCC contains only one species x_3, corresponding to the only driver species that needs to be actuated. The controlled network $\mathcal{G}^{\mathsf{c}}$ does satisfy the graph-theoretical criteria of structural accessibility. Hence the microbial system is structurally accessible. Indeed, a controller can be designed using the feedback linearization methodology (see Examples 1 and 2 in SI-2) to obtain a suitable control input $u(t)$ (either continuous or impulsive) that can be applied to the driver species x_3 to steer steer the whole MC towards the desired steady-state (Fig.1B). The obtained continuous (or impulsive) control input $u(t)$ is shown in Fig.1C (or Fig.1D), respectively. The time evolution of the species abundances are also shown in Fig.1C and Fig.1D.

ℓ-structural accessibility

Without detailed knowledge of an adequate pair of functions $\{f, g\}$ that models the controlled population dynamics of the MC, a systematic controller design for arbitrary MCs would be very challenging if not impossible [36, 37]. Indeed, the only general class of uncertain systems for which systematic controller design methods exist is linear systems. However, the selection of driver species by making $\mathcal{G}^{\mathsf{c}}$ structurally

accessible does not always guarantee that those robust linear control techniques can be applied. This is because, given an structurally accessible class $\mathfrak{G}$, it is possible that only its nonlinear function pairs lack autonomous elements (Example 3 in SI-3). This implies that there exist states that cannot be reached using any linear controller, despite being infinitesimally close to the initial state of the MC.

To circumvent the above limitation and allow for the design of feedback controllers that can (at least locally) steer an arbitrary MC with uncertain dynamics, we introduce the notion of ℓ-structural accessibility. The class $\mathfrak{G}$ is said ℓ-structural accessible if at least one of its linear pairs $\{Ax, B\}$ is free of autonomous elements. We can also prove that this condition implies that almost all linear pairs in $\mathfrak{G}$ are free of autonomous elements (Proposition 7 in SI-3). Note that the zero-nonzero patterns of the matrices $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{m\times n}$ are fully determined by $\mathcal{G}^{\mathsf{c}}$. Note also that, for linear systems $\dot{x} = Ax + Bu$, the absence of autonomous elements is equivalent to their (linear) controllability [33] —the intrinsic ability to steer these linear systems between two arbitrary states, which can be verified by the famous Kalman's rank condition: $\mathsf{rank}[B, AB, A^2B, \ldots, A^{n-1}B] = n$ [36].

If $\mathfrak{G}$ is ℓ-structurally accessible and $\{f, g\} \in \mathfrak{G}$ is any pair that adequately models the population dynamics of the controlled MC, it is always possible to rewrite this pair as

$$\{f, g\} = \{Ax + w_x, B + w_u\}, \tag{4}$$

where $(w_x, w_u) = (f - Ax, g - B)$ are some nonlinear functions, and the pair (A, B) is linearly controllable. The selection of (A, B) now becomes a part of the controller design, with better selections (i.e., those minimizing w_x and w_u) leading to controllers that can steer the system to more remote states. Of course, the best selection would be the linearization of $\{f, g\}$ at the initial state of the MC with zero control. This, however, requires exact knowledge of the uncertain $\{f, g\}$. To circumvent this requirement, we propose selecting some proxy (A, B) of this linearization; for example, the A matrix can simply be the weighted adjacency matrix of $\mathcal{G}$. Note that, regardless of the selection, the controllability of (A, B) guarantees that optimal robust linear feedback controllers $u = Kx$ can always be designed to locally steer the MC. Indeed, the feedback gain matrix $K \in \mathbb{R}^{m\times n}$ can be systematically designed using a variety of methodologies, such as linear quadratic regulators or $\mathcal{H}_\infty$ control (see details in SI-4).

We proved that, with either continuous or impulsive control, $\mathfrak{G}$ is ℓ-structurally accessible iff it is structurally accessible and, additionally, its controlled ecological network $\mathcal{G}^{\mathsf{c}}$ has no dilations in the state nodes

(Theorem 6 is SI-3). A dilation in the state nodes exists if there is a subset of state nodes S such that it neighborhood set $T(\mathsf{S}) \subseteq \mathsf{U} \cup \mathsf{X}$ has smaller size than S. These conditions turn out to be equivalent to the conditions for linear structural controllability of $\mathcal{G}^{\mathsf{c}}$ [38], which has recently received renewed attention in the context of network science [39, 40].

Finding minimal sets of ℓ-driver species

Minimal sets of ℓ-driver species can be efficiently found by solving the maximum matching problem on the directed graph $\mathcal{G}$ [39]. Indeed, in addition of having a path that starts in U to each state node, it is necessary that X is covered by a disjoint union of cycles and paths (a.k.a. the cactus structure in linear structural control theory [38]). Notice that for general population dynamics model, each species node in the ecological network $\mathcal{G}$ will usually have a self-loop due to the intrinsic growth and the intra-species interactions. Consequently, all state nodes are matched by themselves and the sets of ℓ-driver species and driver species coincide. This result suggest that, for most MCs, it might be enough to find their sets of driver species to design robust linear controllers.

Using control to restore the gut microbiota of mice

We identified a minimal set of driver species and ℓ-driver species for the gut microbiota of germ-free mice that are pre-colonized with a mixture of human commensal bacterial type stains and then infected with *C. difficile* spores [24]. The underlying ecological network $\mathcal{G}$ was inferred from time-resolved metagenomics data. This MC has fourteen strains (Fig. 3A), with *R. obeum* and *R. mirabilis* (x_1 and x_{12}) being the root SCCs of $\mathcal{G}$, thus forming the minimum set of driver species of the community. As discussed earlier, if all species have intra-species interactions, this set of driver species would also be a set of ℓ-driver species.

To further validate our conclusion that a set of driver species or ℓ-driver species remains valid after adding new edges to the network, we consider finding a set of ℓ-driver species for the ecological network in which all self-loops are removed. This makes necessary adding three more actuated species —for example, *B. ovalus, C. ramnosum* and *A. muciniphila* corresponding to $\{\mathsf{x}_2, \mathsf{x}_6, \mathsf{x}_{10}\}$— to obtain a disjoint union of cycles and paths covering all state nodes (Fig. 3A), providing a minimal set of ℓ-driver species. We validated by simulation the effectivity of the above five identified ℓ-driver species for steering the system towards a desired state despite uncertainty in the controlled system dynamics and the presence of self-loops. With this

aim, the simulation results use the GLV model, which implicitly adds a self-loop to each state node (see SI-5 for details on the dynamic model used for the simulation). To design the controller, we selected the A matrix in Eq. (4) as the weighted adjacency matrix of $\mathcal{G}$ without self-loops. The zero-nonzero pattern of B is determined by the selected driver species, while the value of its non-zero entries were randomly chosen. Linear optimal robust controllers were designed (see SI-4.4) and the resulting control signal was applied to the driver species to steer the system from an initial "diseased" state $x_0 \in \mathbb{R}^{14}$, in which *C. difficile* is overabundant compared to the rest of the species, towards a desired state with a more balanced abundance of species (Fig. 3B-D). The efficacy of the control signal also shows that the designed controller is robust enough to steer the system despite not knowing the presence of self-loops.

Controlling the core microbiota of a sea sponge

We also identified and validated a minimal set of driver species and ℓ-driver species for controlling the core microbiota of the sea sponge *Ircina oros* [25], see SI-5 for details of the dynamic model used for the simulation. This MC has twenty species; 14 of them are driver species, and two more are needed to obtain a minimal set of ℓ-driver species (Fig. 3E-H).

DISCUSSION

The notion of structural accessibility we introduced here is a generalization of the notion of structural controllability of linear systems (see e.g.[38–40] and Remark 7 in SI-3). It is also a generalization of the control theoretic notion of accessibility —defined as the absence of autonomous elements of a controlled system with known dynamics, a keystone notion in nonlinear control [33]— to systems with uncertain dynamics.

Although it has been noted before the that notion of controllability could be used to predict the success of ecosystem management strategies [41], such notion is not completely adequate for MCs and many other biological systems. First, for those systems, some states are unreachable simply due to their nature (e.g., states with negative abundances for MCs) and not due to ineffective control actions. Thus, it would be overkill to demand that the control actions provide controllability of the system. Second, precise dynamic models for some particular MCs might be unknown and very difficult to infer, making impossible to an-

alyze their controllability. The notion of structural accessibility that we introduced here overcomes those two limitations, opening the door for controlling other complex biological systems beyond MCs. Gene regulatory systems, for example, have highly uncertain and very nonlinear dynamics, but their underlying gene regulatory networks have continued being mapped in exquisite detail.

To achieve the ultimate goal of controlling MCs, there are still several steps to be taken to further enhance our control theoretical framework. First of all, despite structural accessibility is insensitive to missing edges in the ecological network of the MC, the design of feedback controllers is not. Hence, the systematic design of controllers that can be very robust to missing interactions will be necessary in order to control real MCs. Second, we could pose the design of controllers with certain constraints, such as only taking positive values. For impulsive control inputs, for example, this corresponds to constraining the control actions to be only microbiota transplantation or probiotic administration. Along this direction, our framework would provide the basis for a systematic design of probiotic cocktails that restore disrupted MCs.

In conclusion, by identifying driver species, our framework shows that correctly inferring the ecological networks underlying MCs could open the door for the systematic and rational design of control strategies to steer MCs towards desired states. It will be necessary to design robust control strategies, letting us calculate adequate control signals despite having large uncertainties on the dynamics of MCs. Therefore, in order to fully harvest the potential benefits of controlling microbial communities for ourselves and our environment, a stronger synergy between microbiology, ecology and control theory will be necessary.

---

[1] John W Pepper and Simon Rosenfeld. The emerging medical ecology of the human gut microbiome. Trends in ecology & evolution, 27(7):381–384, 2012.

[2] Natasha DeLeon-Rodriguez, Terry L Lathem, Luis M Rodriguez-R, James M Barazesh, Bruce E Anderson, Andreas J Beyersdorf, Luke D Ziemba, Michael Bergin, Athanasios Nenes, and Konstantinos T Konstantinidis. Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. Proceedings of the National Academy of Sciences, 110(7):2575–2580, 2013.

[3] Alison Buchan, Gary R LeCleir, Christopher A Gulvik, and José M González. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nature Reviews Microbiology, 12(10):686–698, 2014.

[4] Jenni Hultman, Mark P Waldrop, Rachel Mackelprang, Maude M David, Jack McFarland, Steven J Blazewicz, Jennifer Harden, Merritt R Turetsky, A David McGuire, Manesh B Shah, et al. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. Nature, 521(7551):208–212, 2015.

[5] Rodney R Dietert and Janice M Dietert. The microbiome and sustainable healthcare. In Healthcare, volume 3, pages 100–129. Multidisciplinary Digital Publishing Institute, 2015.

[6] Jacek Karczewski, Barbara Poniedziałek, Zygmunt Adamski, and Piotr Rzymski. The effects of the microbiota on the host immune system. Autoimmunity, 47(8):494–504, 2014.

[7] Laura M Cox and Martin J Blaser. Antibiotics in early life and obesity. Nature Reviews Endocrinology, 11(3):182–190, 2015.

[8] Roger East. Microbiome: Soil science comes to life. Nature, 501(7468):S18–S19, 2013.

[9] Ulrich G Mueller and Joel L Sachs. Engineering microbiomes to improve plant and animal health. Trends in microbiology, 23(10):606–617, 2015.

[10] A. P. Alivisatos, M. J. Blaser, E. L. Brodie, M. Chun, J. L. Dangl, T. J. Donohue, P. C. Dorrestein, J. A. Gilbert, J. L. Green, J. K. Jansson, R. Knight, M. E. Maxon, M. J. McFall-Ngai, J. F. Miller, K. S. Pollard, E. G. Ruby, and S. A. and Taha. A unified initiative to harness earth's microbiomes. Science, 350(6260):507–508, 2015.

[11] M. McFall-Ngai N. Dubilier and L. Zhao. Create a global microbiome effort. Nature, 2015.

[12] Martin J Blaser, Zoe G Cardon, Mildred K Cho, Jeffrey L Dangl, Timothy J Donohue, Jessica L Green, Rob Knight, Mary E Maxon, Trent R Northen, Katherine S Pollard, et al. Toward a predictive understanding of earths microbiomes to address 21st century challenges. mBio, 7(3):e00714–16, 2016.

[13] Stefanie Widder, Rosalind J Allen, Thomas Pfeiffer, Thomas P Curtis, Carsten Wiuf, William T Sloan, Otto X Cordero, Sam P Brown, Babak Momeni, Wenying Shou, et al. Challenges in microbial ecology: building predictive understanding of community function and dynamics. The ISME journal, 2016.

[14] Selman A Waksman. What is an antibiotic or an antibiotic substance? Mycologia, 39(5):565–569, 1947.

[15] Ronald S Oremland and Douglas G Capone. Use of specific inhibitors in biogeochemistry and microbial ecology. In Advances in microbial ecology, pages 285–383. Springer, 1988.

[16] Jürgen Schrezenmeir and Michael de Vrese. Probiotics, prebiotics, and synbioticsapproaching a definition. The American journal of clinical nutrition, 73(2):361s–364s, 2001.

[17] Charlie G Buffie, Vanni Bucci, Richard R Stein, Peter T McKenney, Lilan Ling, Asia Gobourne, Daniel No, Hui Liu, Melissa Kinnebrew, Agnes Viale, et al. Precision microbiome reconstitution restores bile acid mediated resistance to clostridium difficile. Nature, 517(7533):205–208, 2015.

[18] ER Jasper Wubs, Wim H van der Putten, Machiel Bosch, and T Martijn Bezemer. Soil inoculation steers restoration of terrestrial ecosystems. Nature plants, 2:16107, 2016.

[19] Travis E Gibson, Amir Bashan, Hong-Tai Cao, Scott T Weiss, and Yang-Yu Liu. On the origins and control of community types in the human microbiome. PLoS Comput Biol, 12(2):e1004688, 2016.

[20] Vanessa V Phelan, Wei-Ting Liu, Kit Pogliano, and Pieter C Dorrestein. Microbial metabolic exchange [mdash] the chemotype-to-phenotype link. Nature chemical biology, 8(1):26–35, 2012.

[21] Karoline Faust and Jeroen Raes. Microbial interactions: from networks to models. Nature Reviews Microbiology, 10(8):538–550, 2012.

[22] Christian Jost and Stephen P Ellner. Testing for predator dependence in predator-prey dynamics: a non-parametric approach. Proceedings of the Royal Society of London B: Biological Sciences, 267(1453):1611–1620, 2000.

[23] Jonathan Friedman, Logan M Higgins, and Jeff Gore. Community structure follows simple assembly rules in microbial microcosms. Nature Ecology & Evolution, (1):0109, 2017.

[24] Vanni Bucci, Belinda Tzen, Ning Li, Matt Simmons, Takeshi Tanoue, Elijah Bogart, Luxue Deng, Vladimir Yeliseyev, Mary L Delaney, Qing Liu, et al. Mdsine: Microbial dynamical systems inference engine for microbiome time-series analyses. Genome biology, 17(1):121, 2016.

[25] Torsten Thomas, Lucas Moitinho-Silva, Miguel Lurgi, Johannes R Björk, Cole Easson, Carmen Astudillo-García, Julie B Olson, Patrick M Erwin, Susanna López-Legentil, Heidi Luter, et al. Diversity, structure and convergent evolution of the global sponge microbiome. Nature Communications, 7, 2016.

[26] John C Moore, Peter C de Ruiter, H William Hunt, David C Coleman, and Diana W Freckman. Microcosms and soil ecology: critical linkages between fields studies and modelling food webs. Ecology, 77(3):694–705, 1996.

[27] Jérôme Mounier, Christophe Monnet, Tatiana Vallaeys, Roger Arditi, Anne-Sophie Sarthou, Arnaud Hélias, and Françoise Irlinger. Microbial interactions within a cheese microbial community. Applied and environmental microbiology, 74(1):172–181, 2008.

[28] Richard R Stein, Vanni Bucci, Nora C Toussaint, Charlie G Buffie, Gunnar Rätsch, Eric G Pamer, Chris Sander, and João B Xavier. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. PLoS Comput Biol, 9(12):e1003388, 2013.

[29] Georg K Gerber. The dynamic microbiome. FEBS letters, 588(22):4131–4139, 2014.

[30] Katharine Z Coyte, Jonas Schluter, and Kevin R Foster. The ecology of the microbiome: Networks, competition, and stability. Science, 350(6261):663–666, 2015.

[31] Amir Bashan, Travis E Gibson, Jonathan Friedman, Vincent J Carey, Scott T Weiss, Elizabeth L Hohmann, and Yang-Yu Liu. Universality of human microbial dynamics. Nature, 534(7606):259–262, 2016.

[32] Phuongan Dam, Luis L Fonseca, Konstantinos T Konstantinidis, and Eberhard O Voit. Dynamic models of the complex microbial metapopulation of lake mendota. NPJ Systems Biology and Applications, 2:16007, 2016.

[33] Giuseppe Conte, Claude H Moog, and Anna Maria Perdon. Algebraic methods for nonlinear control systems. Springer Science & Business Media, 2007.

[34] Bernold Fiedler, Atsushi Mochizuki, Gen Kurosawa, and Daisuke Saito. Dynamics and control at feedback vertex sets. i: Informative and determining nodes in regulatory networks. Journal of Dynamics and Differential Equations, 25(3):563–604, 2013.

[35] Thomas H.. Cormen, Charles Eric Leiserson, Ronald L Rivest, and Clifford Stein. Introduction to algorithms, volume 6. MIT press Cambridge, 2001.

[36] Karl Johan Aström and Richard M Murray. Feedback systems: an introduction for scientists and engineers. Princeton university press, 2010.

[37] Rodolphe Sepulchre, Mrdjan Jankovic, and Petar V Kokotovic. Constructive nonlinear control. Springer Science & Business Media, 2012.

[38] Ching Tai Lin. Structural controllability. Automatic Control, IEEE Transactions on, 19(3):201–208, 1974.

[39] Yang-Yu Liu, Jean-Jacques Slotine, and Albert-László Barabási. Controllability of complex networks. Nature, 473(7346):167–173, 2011.

[40] Yang-Yu Liu and Albert-László Barabási. Control principles of complex systems. Reviews of Modern Physics, 88(3):035006, 2016.

[41] Craig Loehle. Control theory and the management of ecosystems. Journal of applied ecology, 43(5):957–966, 2006.

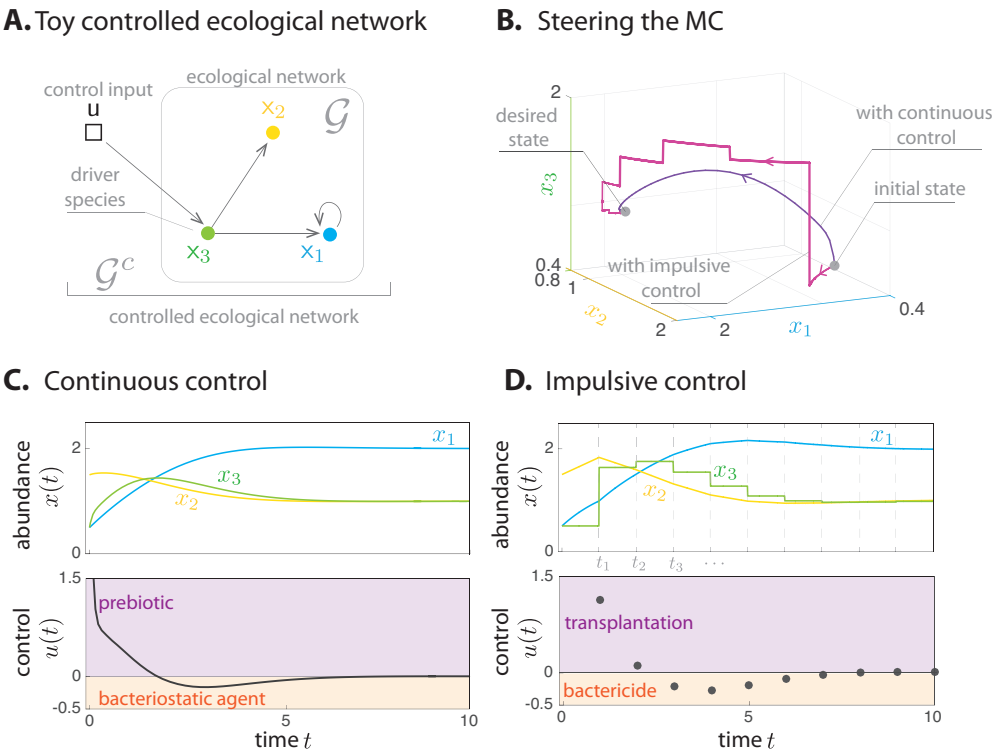

FIG. 1: **Controlling a small microbial community towards a desired state.** **A.** The controlled ecological network $\mathcal{G}^{\mathbf{c}}$ of a toy MC with $n = 3$ species (green, yellow, blue) and $m = 1$ control actions (square) actuating the species x_3. Our theoretical framework allows us to identify this actuated species as the minimal set of driver and ℓ-driver species, since $\mathcal{G}^{\mathbf{c}}$ is structurally accessible and ℓ-structurally accessible. **B.** Based on the identified driver species, we can design a continuous $u(x(t))$ or an impulsive feedback controller $u(x(t_k))$ to steer the MC from a initial to a desired equilibrium state.These two control signals are displayed in panels C and D, respectively. **C.** In the case of continuous control, the effect of the control actions $u(t) \in \mathbb{R}$ on the state $x(t) \in \mathbb{R}^3$ are modeled as a continuous signals modifying the growth of the actuated species. Positive and negative control inputs correspond to prebiotics and bacteriostatic agents, respectively. In simulating this toy controlled MC, for illustration purpose, we used the dynamics $\dot{x}_1 = (1 - x_1 + x_3)/(1 + x_3)$, $\dot{x}_2 = (1 - x_3)/(1 + x_3)$ and $\dot{x}_3 = u$, which has strong nonlinearities and might not necessarily follow any classic population dynamics. Note that real MCs could have less nonlinear dynamics (the Generalized Lotka-Volterra equations, for example) and more self-loops in their ecological networks, facilitating the design of controllers. **D.** In the case of impulsive control, the control actions $u(t) \in \mathbb{R}$ are modeled as inputs applied at discrete time instants $\mathbb{T} = \{t_1, t_2, \cdots\}$, instantaneously modifying the state of the actuated species. Positive and negative control inputs correspond to transplantation and bactericides, respectively. In simulating this toy community, we use the dynamics as in C, except that equations for x_3 are replaced by $\dot{x}_3 = 0$ and $x_3(t^+) = x_3(t) + u(t)$ if $t \in \mathbb{T}$.

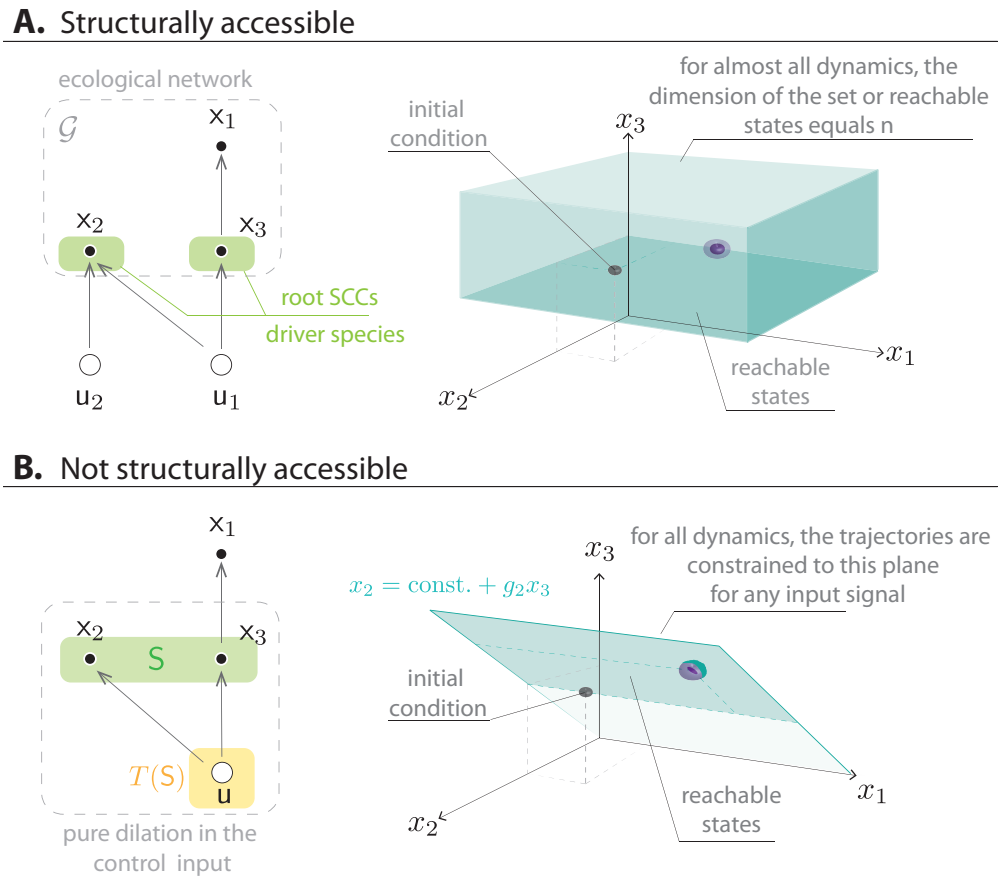

FIG. 2: **Structural accessibility and its graph-theoretical characterization.** **A.** For an structurally accessible class of controlled systems $\mathfrak{G}$, for almost all its pairs, their set of reachable states has dimension n, guaranteeing that the control actions can locally steer the whole system. Consequently, the condition of structural accessibility can be used to find minimal sets of driver species that provide control of the rest of the species in the MC. A minimal set of driver species is obtained by arbitrary selecting one node per root SCCs of the ecological network $\mathcal{G}$. In this example, $\mathcal{G}$ has two root SCCs characterizing its two driver species. If these two driver species are actuated by two inputs u_1 and u_2, the resulting controlled ecological network $\mathcal{G}^{\mathbf{c}}$ is structurally accessible. **B.** The control actions do not provide full control when the trajectories of the system are always constrained to a low-dimensional manifold for population dynamics and susceptibility matrix $\{f, g\} \in \mathfrak{G}$ and all inputs $u \in \mathbb{R}^m$. In such case, the class $\mathfrak{G}$ is not structurally accessible. In this example, the controlled ecological network $\mathcal{G}^{\mathbf{c}}$ indicates that the possible dynamics that the MC can have are of form $\dot{x}_1 = f_1(x_3), \dot{x}_2 = g_2 u, \dot{x}_3 = g_3 u$, for some meromorphic function f_1 and constants g_2, g_3. For all those controlled systems, the function $\xi = g_3 x_2 - g_2 x_3$ is an autonomous element because $\dot{\xi} = 0$ and $\xi(t^+) = \xi(t)$. Consequently, the trajectories of the system are always constrained to an hyperplane of the form $x_2 = \text{const.} + (g_2/g_3)x_3$. Our analysis shows that the presence of such autonomous element can be deduced purely from the topology of $\mathcal{G}^{\mathbf{c}}$, since it contains a pure dilation in the control input.

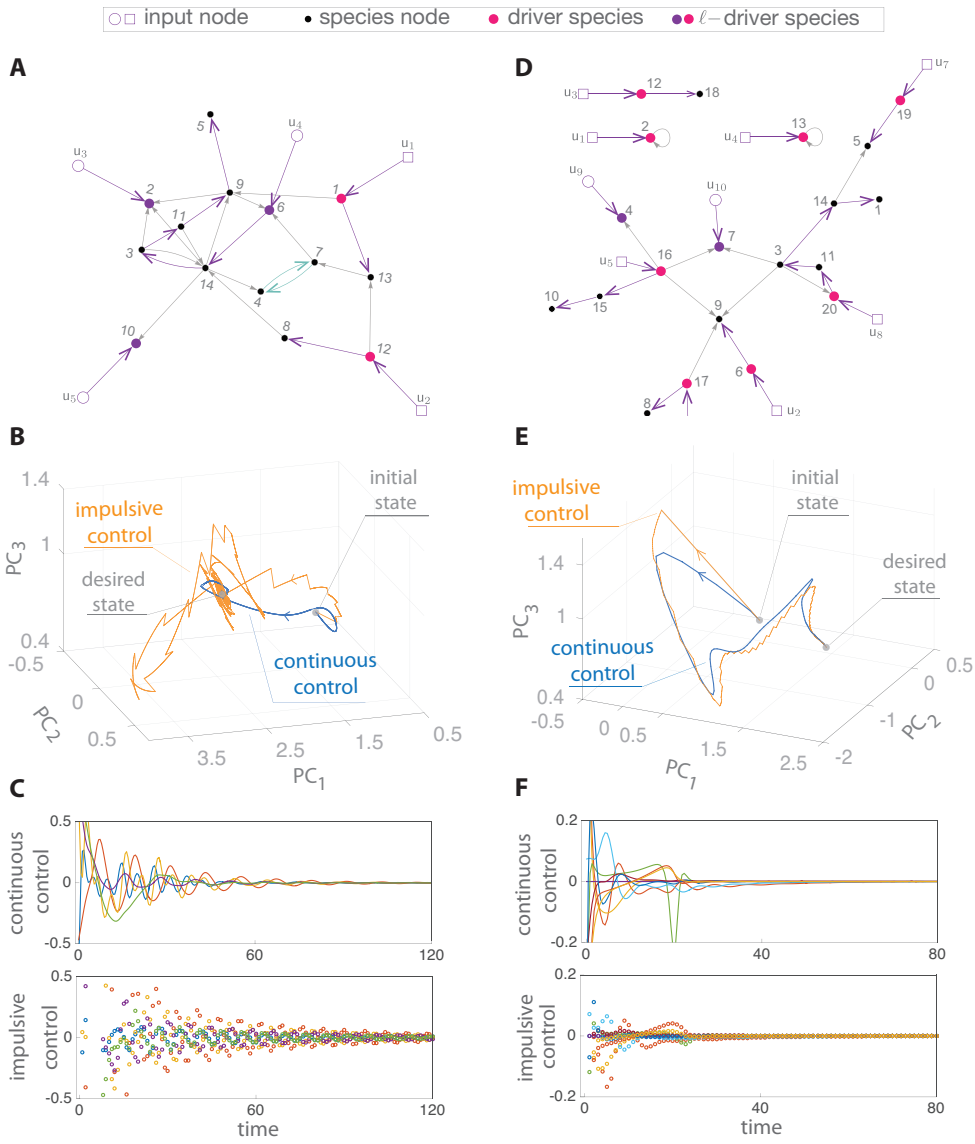

FIG. 3: **Controlling two real microbial communities.** **A.** Ecological network of the gut microbiota of gnotobiotic mice that are pre-colonized with a mixture of human commensal bacterial type stains and then infected with C. difficile. A minimal set of driver species and ℓ-driver species is shown. In particular, the minimal set of ℓ-driver species provides a disjoint union of paths that start in the control input set U (purple) and cycles (green) that cover all species nodes X. Refer to Table 1 SI-5 for species name (C. difficile corresponds to species 7). **B.** The temporal response of the MC to the control input was simulated using the GLV equations. Without control, the parameters of the system were adjusted to have a diseased equilibrium state x_0 in which C. difficile is overabundant compared to the rest of species. The desired healthy state x_f is chosen as an equilibrium with a more balanced species abundance profile. Assuming inexact knowledge of its dynamics, feedback controller were designed to obtain the control signals applied to the driver species in order to steer the MC towards the desired healthy state (see SI-5 for details). The figure shows the projection of the high-dimensional abundance profiles (states of the MC) into the first three principal components (PCs). The temporal response of the complete state of the MC is shown in Fig SI-2. **C.** Continuous and impulsive control signals produced by the feedback controllers (see SI-5 for details). **D.** Ecological network of the core microbiota of Ircina oros (refer to Table 1 in SI-5 for species names). A minimal set of driver species and ℓ-driver species is shown. **E.** The temporal response of the MC to the control input was also simulated using the GLV equations (see SI-5 for details). Without control, the parameters of the system were adjusted to have an unbalanced equilibrium. Feedback control applied to the driver species was used to steer the abundance profile towards the desired healthy state. Only the first three principal components of the state trajectory are shown. The temporal response of the complete state of the MC is shown in Fig SI-2. **F.** Continuous and impulsive control signals produced by the feedback controllers (see SI-5 for details).