

# Multi-cellularity without cooperation

Mikhail Tikhonov

*Center of Mathematical Sciences and Applications*

*School of Engineering and Applied Sciences and*

*Kavli Institute for Bionano Science and Technology, Harvard University, Cambridge, MA 02138, USA*

The hallmark of a major evolutionary transition, whereby independently replicating units became grouped together into larger wholes, is the “decoupling” of fitness of the collective from the individual fitnesses of the units. It is widely believed that the key element making such decoupling possible is extensive cooperation between the lower-level entities. Here, it is demonstrated that cohesive multicellular behavior can arise in a purely competitive setting as a generic consequence of division of labor. A minimal theoretical model of competitive coexistence on multiple resources provides an explicit manifestation of fitness decoupling in a rigorous mathematical sense. This form of multicellular behavior is not vulnerable to “cheaters” and can be expected to be widespread in microbial ecosystems.

In 1995, Maynard-Smith and Szathmary in their highly influential work [1] introduced the concept of a major evolutionary transition. Several of the transitions they identified constitute evolutionary transitions in individuality, e.g. the evolution of multicellularity, whereby independently replicating units became grouped together into larger wholes. A hallmark of this process is the “decoupling” of fitness of the collective from the individual fitnesses of the lower-level entities [2, 3].

This process is puzzling: how does a collective acquire interests that are distinct from the interests of its members? It is widely believed that the crucial role is played by cooperative interactions among lower-level entities [2–4]. This is highly intuitive: put simply, until cooperation between individual units binds them into a coherent whole, there is no whole, and therefore there can be no interests of the whole. Without cooperation, a fit collective must necessarily be merely a collective made of fit individuals: fitness decoupling cannot occur.

Accordingly, Michod’s model for the evolution of multicellularity [5] proceeds in two steps. First, cooperation between organisms evolves as an individual trait, typically (though not always [4]) conditioned on their high relatedness. Second, once cooperation becomes sufficiently strong, selection can act on collectives as true evolutionary “wholes”. The role cooperation is expected to play in major evolutionary transitions is one of the reasons this topic has enjoyed so much attention in the recent literature.

At the same time, it is known that, at least in laboratory conditions, artificial selection can successfully act on entire ecosystems such as soil microbiota [6, 7], where pervasive cooperation between organisms appears unlikely. These findings were confirmed and nuanced by simulation-based investigation [8] and could have important implications for evolutionary theory. However, until the mechanisms underlying apparent success of higher-level selection in genetically inhomogeneous communities are clearly understood theoretically, the relevance of such observations for natural evolutionary processes will re-

main unclear.

The purpose of this work is to demonstrate that cooperation is indeed not a necessary ingredient for cohesive multi-cellular behavior. It will be shown that the decoupling of collective fitness can occur in a purely competitive setting. Evolution of cooperation is famously prone to the emergence of “cheaters” and so requires very special circumstances, such as conflict suppression or “policing” mechanisms [9]. In contrast, competition for multiple resources as considered in this work is a generic occurrence in the highly diverse microbial communities that dominate our planet [10–12]. These results suggest that “acting as a whole”, in a rigorous mathematical, rather than purely metaphorical sense, may be expected to be ubiquitous in microbial ecology.

## THE MODEL: COMPETITIVE COEXISTENCE ON MULTIPLE RESOURCES

This work uses the “metagenome partitioning” model introduced in [13], which builds on Mac Arthur’s treatment of competitive coexistence on multiple resources [14] to model division of labor in large communities. For convenience, all the relevant definitions are reproduced below; for a detailed introduction and a discussion of the intuition behind the model, see [13].

Consider a world with  $N$  resources  $i \in \{1 \dots N\}$  denoted  $A$ ,  $B$ , etc. These resources can be harvested with “pathways”  $P_i$ . An organism is defined by its “functional type”, namely the pathways that it carries. There are  $2^N - 1$  possible functional types; they will be denoted using a binary vector of pathway presence/absence:  $\vec{\sigma} \in \{0, 1\}^N$ . Let  $n_{\vec{\sigma}}$  be the total number of organisms  $\vec{\sigma}$  in the population. The total benefit  $R_i$  from a resource  $i$  is equally distributed among all organisms carrying the pathway  $P_i$ ; their number will be denoted  $T_i$ :

$$T_i \equiv \sum_{\text{all } \vec{\sigma} \text{ carrying } i} n_{\vec{\sigma}} = \sum_{\vec{\sigma}} n_{\vec{\sigma}} \sigma_i.$$

The growth rate of every organism type is determined by its *resource surplus*  $\Delta\varphi$ :

$$\Delta\varphi_{\vec{\sigma}} = \sum_i \sigma_i \frac{R_i}{T_i} - \chi_{\vec{\sigma}}. \quad (1)$$

Here the first term is the total benefit harvested by all carried pathways, and the second term represents the maintenance costs of organism  $\vec{\sigma}$ . In [13], costs were modeled as possessing some structure stemming from cross-talk between pathways. For the purposes of this work, a simpler model of unstructured random costs will be sufficient:

$$\chi_{\vec{\sigma}} = |\vec{\sigma}|(1 + \epsilon \delta\chi_{\vec{\sigma}}). \quad (2)$$

Here the frozen disorder  $\delta\chi_{\vec{\sigma}}$  is a random variable drawn out of the standard normal distribution (chosen once for each organism type), and  $|\vec{\sigma}| \equiv \sum_i \sigma_i$  is the number of pathways carried by the organism. The reason for including this factor will become clear shortly. Note that  $N$ ,  $\epsilon$  and  $R_i$  are the only parameters in the model.

The resource surplus  $\Delta\varphi$  is used to generate biomass. The simplest approach is to equate the biomass of the organism with its cost, so that excess resources are “spent” on making new organisms.<sup>1</sup> The total biomass change due to resource-dependent growth or death is then given by:

$$\chi_{\vec{\sigma}} \frac{dn_{\vec{\sigma}}}{dt} = n_{\vec{\sigma}} \Delta\varphi_{\vec{\sigma}}. \quad (3)$$

In this model, a given random realization of the costs  $\chi_{\vec{\sigma}}$  summarizes all the biochemistry that makes different organisms types more or less efficient at processing their resources; our focus will therefore be on studying the dynamics of the model while the costs of organisms are held fixed. The types that are present at non-zero abundance in a given community  $C$  will be denoted  $\Omega(C)$ . As established in [13], any community  $C$  will eventually converge to an equilibrium  $C^*$  uniquely determined by the set of organism types  $\Omega(C) = P$  (irrespective of their initial abundance). At this equilibrium, certain types  $S^* = \Omega(C^*) \subset P$  establish at a non-zero abundance, while others “go extinct”, their abundance exponentially decreasing towards zero. Here and below, the starred quantities refer to equilibrium of ecological dynamics.

The equilibrium  $C^*$  will be stable with respect to any perturbations restricted to organisms in  $P$ , but may be invaded by some other organism  $\vec{\sigma} \notin P$ . Assuming a weak mutation rate for loss/acquisition of pathways, this model can be seen as “mesoscopic” population genetics

<sup>1</sup> This differs from the convention used in [13], where the biomass was equated to  $|\vec{\sigma}|$ . This difference is insignificant; the approach taken in this work is perhaps more appealing aesthetically.

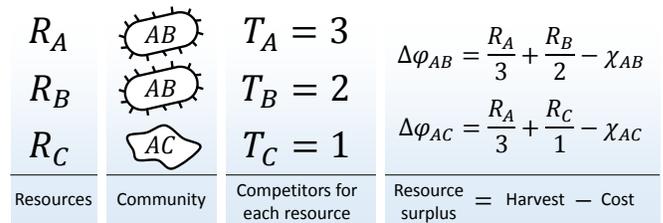


FIG. 1. **The metagenome partitioning model.** Organisms are defined by the pathways they carry, the benefit from each resource is equally partitioned among all organisms who can harvest it, and growth/death of each individual is determined by the resource surplus it experiences.

for bacteria evolving through horizontal gene transfer. For the purposes of this work, it will be assumed that mutation events introduce a random new type, but are sufficiently rare that ecological dynamics have time to equilibrate before a novel invader can arise, so that population proceeds from equilibrium to equilibrium. After infinite time, the community in this model will stabilize at the unique equilibrium that cannot be invaded by any organism. This state will be called the *final* equilibrium.

For simplicity, consider an environment  $\mathcal{E}_0$  with all resources supplied in equal abundance:  $R_i \equiv R$ . The final equilibrium community for one random realization of organism costs with  $R = 100$ ,  $N = 10$  and  $\epsilon = 10^{-3}$  is depicted in Fig. 2A. In this case it consists of 9 organism types present at varying abundances. It is natural to ask: for a given initial set of competing organisms  $P$ , what determines which organisms  $S^*$  survive?

### INDIVIDUAL FITNESS PREDICTS THE OUTCOME OF COMPETITION BETWEEN INDIVIDUALS

Intuitively, the fitness of an organism must be determined by its cost:  $\chi_{\vec{\sigma}}$  is the lowest resource harvest at which an organism can persist, and by analogy with Tilman’s  $R^*$  rule, low-cost types should have a competitive advantage.

To define fitness, consider an assay whereby an organism  $\vec{\sigma}$  is placed (at abundance  $n_{\vec{\sigma}} = 1$ ) in the environment  $\mathcal{E}_0$  with no other organisms present. Its initial growth rate is given by:

$$\chi_{\vec{\sigma}} \left. \frac{dn_{\vec{\sigma}}}{dt} \right|_{t=0} = \sum_i R_i \sigma_i - \chi_{\vec{\sigma}} = |\vec{\sigma}|R - \chi_{\vec{\sigma}},$$

so that

$$\left. \frac{dn_{\vec{\sigma}}}{dt} \right|_{t=0} = R \frac{|\vec{\sigma}|}{\chi_{\vec{\sigma}}} - 1.$$

As expected, the growth rate is determined by the organism’s cost, specifically its cost per pathway (this is

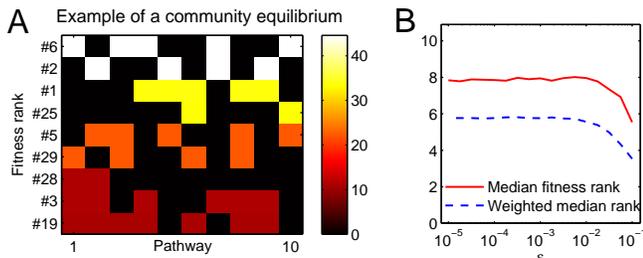
why the factor  $|\sigma|$  was included in the cost model (2)). Our goal is to define fitness as a quantity predicting survival; for this purpose, additive constants or rescaling is irrelevant. It is therefore natural to subtract the baseline  $|\sigma|/\chi_\sigma \approx 1$  and define the *individual fitness* of type  $\vec{\sigma}$  as

$$f_{\vec{\sigma}} \equiv \frac{|\sigma|}{\chi_\sigma} - 1. \quad (4)$$

This is a quantity of order  $\epsilon$ , and so with this definition,  $\epsilon$  serves as the natural scale. The fitness ranking between organism types is random and is set by the disorder  $\delta\chi$ .

To see that the individual fitness defined in this way is indeed predictive of the outcome of competition between types, consider the example equilibrium state depicted in Fig. 2A. Note that the set of 9 survivors predominantly consists of highly fit individuals; in particular, the top 3 types by fitness are all present, and the remaining types are all within the top 30 (out of 1023 total competing types).

To confirm the generality of this observation, Fig. 2B shows the median fitness rank of the types  $S^*$  that survive at equilibria of randomly initialized populations. Specifically, for a given cost structure, 10 random subsets  $P_i$  of 100 types each were equilibrated to determine survivors  $S_i^*$ . The procedure was repeated for  $\epsilon$  ranging from  $10^{-5}$  to 0.1, and for 10 random realizations of the cost structure at each  $\epsilon$ . Thus for each value of  $\epsilon$ , a total of 100 randomly constructed communities were evaluated. Fig. 2B shows the median fitness rank of survivors  $S^*$  within the respective subset  $P$ , averaged over these 100



**FIG. 2. Individual fitness is a good predictor of competition outcome.** **A:** Final (non-invadable) community equilibrium for one particular random realization of the model ( $N = 10$ ,  $\epsilon = 10^{-3}$ ). Rows correspond to surviving organism types and the pathways they carry; rows are labeled by fitness rank and colored to reflect organism abundance. For example, one of the survivors carries pathways 1 and 2 only; its fitness rank is #28 and abundance  $\approx 10$ . The final equilibrium consists of highly fit types (within top 30 out of 1023). **B:** The median fitness rank of survivors, weighted (dashed) or not weighted (solid) by abundance. Curves show mean over 100 random communities for each  $\epsilon$ ; the standard deviation across 100 instances is stable at approximately 40% of the mean for both curves, independently of  $\epsilon$  (not shown to reduce clutter). The panel indicates that fitness rank is predictive of both survival and abundance at final equilibrium.

instances, where the median was either weighted (blue dashed line) or not weighted (red solid line) by abundance of the type at equilibrium. The weighted median rank is considerably lower, indicating that fitter types are present at higher abundance. This figure confirms that the fitness rank of a type is strongly correlated both with its probability of survival and with its abundance at equilibrium.

Also apparent from Fig. 2B is that for small enough  $\epsilon$ , the structure of the final equilibria does not significantly depend on this parameter. This can be intuitively understood as follows. Consider two resources  $A, B$  and organisms  $\sigma_A = \{1, 0\}$ ,  $\sigma_B = \{0, 1\}$ , and  $\sigma_{AB} = \{1, 1\}$ . If

$$\chi_{AB} > \chi_A + \chi_B, \quad (5)$$

it easily follows that the “generalist” organism  $\sigma_{AB}$  will eventually be outcompeted by the two specialists  $\sigma_A$  and  $\sigma_B$ . Conversely, if the opposite inequality holds, then  $\sigma_A$  and  $\sigma_B$  cannot stably coexist in the final equilibrium, since  $\sigma_{AB}$  will always be able to invade, displacing one of them. In this way, in the metagenome partitioning model, community composition is shaped primarily by inequalities like (5), which are invariant under changes in  $\epsilon$  and depend only on the realization of disorder. The dominant effect of  $\epsilon$  is to set the time scale for the ecological dynamics. In what follows,  $\epsilon$  will be fixed at  $10^{-3} \approx 2^{-N}$ , sufficiently small to be in the regime where equilibrium structure is no longer sensitive to  $\epsilon$ . The large- $\epsilon$  regime will be discussed later.

While Fig. 2 demonstrates that the fitness rank is indeed correlated with organism abundance at equilibrium, it also shows that this notion of individual fitness is not wholly sufficient. For example, at the equilibrium shown in Fig. 2A, the type ranked 4th in individual fitness went extinct, but 6 other types ranked as low as #29 remained present, despite being individually “less fit”.

The origin of this is clear: the instantaneous growth rate  $r$ , and therefore fitness, of any organism type is a function of the traits it possesses, encoded by  $\vec{\sigma}$ , but also of the state of the environment  $\mathcal{E}$ , encoded by the resource harvest vector  $h_i \equiv R_i/T_i$ :

$$r(\vec{\sigma}, \mathcal{E}) \equiv r(\vec{\sigma}, \{h_i\}) = \frac{1}{\chi_{\vec{\sigma}}} \left( \sum_i \sigma_i h_i - \chi_{\vec{\sigma}} \right).$$

As organisms reproduce or die, they modify their environment and thus their own fitness and the fitness of all other types. This, of course, is the standard argument of niche construction theory [15]. For example, consider the three-resource world depicted in Fig. 1, and assume that  $\sigma_{AB} = \{1, 1, 0\}$  is the most fit type with a very low cost. As  $\sigma_{AB}$  multiplies, it depletes resources  $A$  and  $B$  (in the sense that the benefit  $R_i/T_i$  any organism can harvest from them is reduced). As a result, the final

equilibrium is highly likely to include the specialist organism  $\sigma_C = \{0, 0, 1\}$ , even if its cost is relatively high, and under other circumstances (if  $\sigma_{AB}$  were less fit) it would lose to  $\sigma_{AC} = \{1, 0, 1\}$  or  $\sigma_{BC} = \{0, 1, 1\}$ .

In general, for any type  $\bar{\sigma}_0$ , it is possible to adjust the cost of some *other* organism types to ensure that  $\bar{\sigma}_0$  will survive at final equilibrium. Therefore, any fitness function of  $\bar{\sigma}_0$  that could predict its fate with certainty would have to depend on the traits of *all* other organisms in the population. Constructing such a fitness function would be equivalent to solving the equilibrium population structure and declaring survivors to be the most fit; such concept of fitness would be tautological and have no predictive value. In contrast, a function that depends only on the traits  $\bar{\sigma}$  can be useful (see Fig. 2), but any such function is necessarily an imperfect predictor.

Motivated by these considerations, we will accept, for now, that the definition (4) constitutes the most useful notion of individual fitness that is available in this context, and continue with the argument. After discussing the results that stem from this decision, we will return to re-examine this issue.

To summarize, individual fitness as defined above is strongly predictive of the outcome of competition between individuals, but is not the whole story. How else can we characterize the ensemble of factors that determine survival or extinction of particular types?

## COMMUNITY FITNESS

A convenient property of the metagenome partitioning model is that the dynamics (3) possess a Lyapunov function [13, 14]:

$$F = \frac{1}{C} \left( \sum_i R_i \ln \frac{T_i}{R_i} - \sum_{\bar{\sigma}} \chi_{\bar{\sigma}} n_{\bar{\sigma}} \right) + F_0. \quad (6)$$

Here  $C$  and  $F_0$  are constants introduced for later convenience. Specifically, set  $C = \sum_i R_i$  and  $F_0 = 1$ ; this choice ensures that close to community equilibrium,  $F$  is of order  $\epsilon$  (see Supplementary Material; SM). This function, defined for  $n_{\bar{\sigma}} \geq 0$  and  $T_i > 0$ , has the property that  $C \frac{\partial F}{\partial n_{\bar{\sigma}}} = \Delta \varphi_{\bar{\sigma}}$ , and therefore  $F$  is monotonously increasing as the system is converging to equilibrium. To illustrate this, Fig. 3 shows 10 trajectories of ecological dynamics for the same system as in Fig. 2A, starting from random initial conditions (see SM). Far from equilibrium, mean individual fitness and  $F$  increase together (Fig. 3, inset), confirming that individual fitness is a useful predictor. However, as the equilibrium is approached, Fig. 3 shows that virtually every trajectory has portions whereby dynamics increase  $F$  at the expense of mean individual fitness. Note that each of the trajectories in Fig. 3 converges to the same equilibrium depicted in Fig. 2A. This is because  $F$  is convex and bounded from

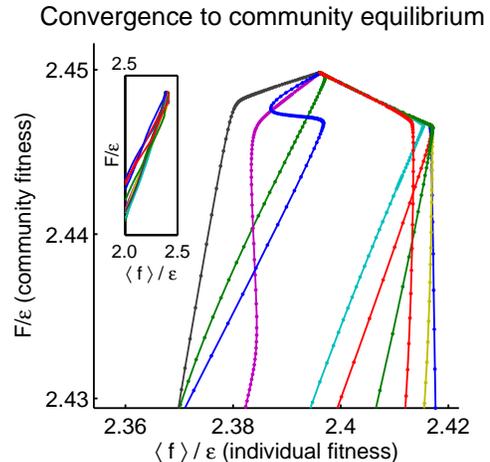


FIG. 3. **Community dynamics maximize a global objective function  $F$  at the expense of mean individual fitness.** 10 trajectories of ecological dynamics for an example system, starting from random initial conditions and converging to the equilibrium depicted in Fig. 2A. **Inset:** a zoomed-out version of the same plot; data aspect ratio as in the main panel.

above (see SM), and its maximum is therefore unique. It follows that community equilibrium always exists, is stable and is uniquely determined by the set of competing types. The goal of this section is to demonstrate that  $F$  possesses all the qualities one might expect from a *bona fide* fitness function of the community as a whole.

To speak of a whole-community fitness function, one first needs to define the relevant Darwinian process. Imagine a collection of islands  $\alpha$ , each experiencing identical influxes  $R_i^{(\alpha)} \equiv R_i$  of our  $N$  resources. Each island  $\alpha$  harbors a community  $C_\alpha^*$  of organisms at equilibrium of their ecological dynamics. On each island, a weak mutation occasionally introduces a random new organism type; if this type can invade, the community transitions to a new equilibrium and awaits a new mutation. This process defines evolution of each community independently. In addition, two islands  $\alpha$  and  $\beta$  may occasionally come into contact. The communities they harbor then merge, equilibrate, and by the time the islands separate again, both harbor the same community, which, by definition, will be called the result of competition between communities  $C_\alpha^*$  and  $C_\beta^*$ .

The first observation to make is that for each island,  $F$  is monotonously increasing throughout its evolutionary history. To see this, notice that  $F$  is continuous and non-singular in all  $n_\sigma$  (as long as  $T_i > 0$  for all  $i$ : it is assumed that each island has discovered all pathways; once discovered, it will never go extinct). Therefore, introducing a new type at a vanishingly small abundance will leave  $F$  unchanged, and if this type can invade the community, the convergence to the new equilibrium is a valid

trajectory of ecological dynamics on which  $F$  increases. Similarly, since the equilibrium state depends only on the organism types but not their initial abundance, the competition between communities  $S_\alpha^*$  and  $S_\beta^*$  can be recast as introducing types  $\Omega(C_\beta^*)$  at a vanishingly small abundance into community  $C_\alpha^*$  or vice versa; in both cases, therefore, the equilibration process can only increase  $F$  (or leave it intact if no types can invade).

Merely demonstrating that  $F$  is increased by evolution is however, not enough. For communities in this model to qualify as true fitness-bearing entities, there must be a sense in which a more fit community can be expected to “outcompete” one that is less fit. Therefore, the next section examines the competition between communities.

### COMMUNITY FITNESS IS MORE PREDICTIVE OF COMPETITION OUTCOME THAN INDIVIDUAL FITNESS

Consider a community  $C_\alpha^*$  coming into contact with community  $C_\beta^*$  as described above, and denote the equilibrated community as  $C^*$  with no subscripts. If none of the organism types from island  $\beta$  could invade the community  $C_\alpha^*$ , then  $C^* = C_\alpha^*$ . In this case community  $C_\alpha^*$  is the clear winner. In general, however, the space of competition outcomes is richer than merely one community taking over the other: both competitors  $C_\alpha^*$ ,  $C_\beta^*$  can contribute to  $C^*$ . If one contributes more types than the other, it can be pronounced more successful at competition, since the resulting community  $C^*$  is closer to it in terms of composition (since community equilibrium is uniquely determined by the set of competing types, similarity between communities is evaluated by counting shared organism types irrespectively of their abundance). What makes a community more likely to be successful?

Let  $C_\alpha^*$  and  $C_\beta^*$  each consist of  $k$  organism types. Then on the one hand, the community  $C^*$  is simply the result of competition between  $2k$  individual types (fewer if  $\Omega(C_\alpha^*)$  and  $\Omega(C_\beta^*)$  overlap), and we have established (Fig. 2) that the individual fitness rank of a type in the pool of competitors is generally a good predictor of its survival. It is therefore reasonable to expect that out of  $2k$  competing types, the survivors will predominantly be the ones with higher individual fitness. This predicts that the community whose members have higher average fitness should be more successful. On the other hand, we have also established that the ultimate equilibrium community that cannot be invaded by *any* type, and so will out-compete any other community, corresponds to the maximum of  $F$ , and generally not the maximum average individual fitness. This suggests that  $F$ , rather than individual fitness, should be the better predictor of the competition outcome.

To settle the competition between these two hypotheses, the following procedure was implemented. For a

given random realization of the cost structure  $\delta\chi$ ,  $M = 50$  random organism types were selected. These 50 types were used to construct all 230300 possible combinations of  $k = 4$  types. All such combinations were independently equilibrated; communities where the equilibrium state had fewer than  $k = 4$  types or where some pathways were not represented were excluded. The mean individual fitness and the putative “collective fitness”  $F$  of the remaining 70160 communities are shown in Fig. 4A. This procedure puts at our disposal multiple examples of communities where individual and collective fitness are both high, both low, or one is high while the other is low (the quadrants highlighted in Fig. 4A). Competing pairs of communities drawn from these pools will make it possible to determine which of the two factors, individual or collective fitness, is the better predictor of community competition outcome.

To begin, consider the cyan and magenta quadrants (I and III, respectively). Any community from the magenta quadrant is more fit than any cyan community, both in the collective sense and as measured by the average member fitness. Therefore, there is little doubt that the magenta communities should be more successful in pairwise competitions. To confirm this, Fig. 4B presents the results of an “elimination assay” competing communities from these quadrants. 500 random pairs were drawn, and correspond to columns in Fig. 4B. For each pair, the organism types from both communities (up to 8 types each time, fewer if the two sets of 4 had overlap) were equilibrated together; the rows in Fig. 4B correspond to these types, ordered by individual fitness: high (top) to low (bottom). For each organism type that went extinct during equilibration, its provenance was identified (did it come from the magenta or the cyan community?), and the corresponding rectangle in the image in Fig. 4B is colored accordingly; in cases when the eliminated type was present in both communities, it was colored yellow. The dominant color in the elimination assay in Fig. 4B is cyan, confirming that the cyan communities are less successful at contributing their members to the final equilibrium. Note also that the colored entries are predominantly located in the bottom half of the table: the eliminated types tend to also have low individual fitness. This is the expected result.

Now, consider the competition between blue and red quadrants (II and IV). An elimination assay conducted in an identical manner is presented in Fig. 4C. Now the colored entries are predominantly red and occupy the *top* half of the table. In other words, members of the red communities are being outcompeted despite the fact that their individual fitness is higher. Surprisingly, the individual fitness of an organism is less predictive of its ability to survive competition than the collective fitness of the community of which it was part.

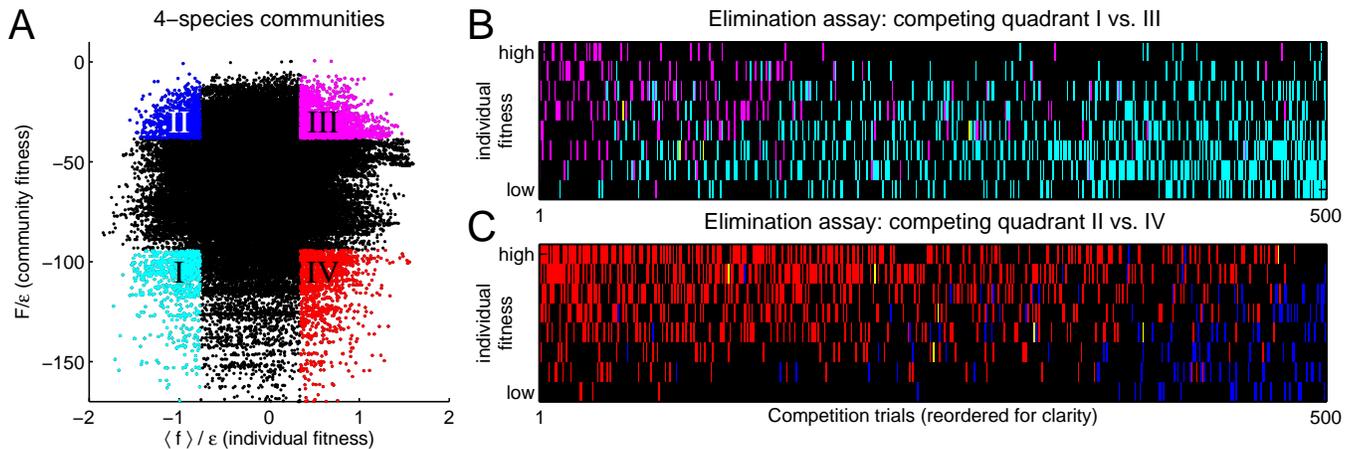


FIG. 4. **Community fitness is more predictive of competition outcome than individual fitness.** **A:** Community fitness *vs.* mean individual fitness, measured in units of  $\epsilon$ , for 70160 communities composed of 4 organism types (see text). Communities in which both characteristics are in the top 10% or bottom 10% are highlighted in color and labeled I through IV. **B:** Elimination assay competing quadrants I (cyan) vs. III (magenta). 500 randomly drawn community pairs (columns) were jointly equilibrated, with up to 8 organism types each time (rows; ordered by individual fitness). For each organism that went extinct during equilibration, the corresponding cell in the table is colored by the organism provenance. As expected, that most eliminated organisms were from the less fit cyan communities (there are more cyan cells than magenta). These organisms were also less fit individually (most colored cells are in the lower half of the table). **C:** Same, competing quadrants II (blue) vs. IV (red). The dominant color is now red: most eliminated types were from red communities, and went extinct despite being individually more fit than average (most colored cells are in the upper half of the table). Columns ordered by dominant color from magenta to cyan (panel B) and blue to red (panel C).

### COMMUNITY AS A FITNESS-BEARING UNIT

Consider an observer who is unaware of the internal structure of the community on each island, and who is able to measure only the overall response of each island (total expression of each pathway  $\vec{T} = \{T_i\}$ ) to the resource influx  $\vec{R} = \{R_i\}$ . First, consider an island  $\alpha_G$  harboring a single organism type: the complete generalist  $\sigma_G = \{1, 1 \dots 1\}$ . Its abundance at equilibrium will be  $n_G = T_i = R_{\text{tot}}/\chi_G$ , where  $R_{\text{tot}} = \sum_i R_i$ . Although resources may be supplied in varying abundance, the island  $\alpha_G$  can only express all pathways at the same level.

Another island  $\alpha_S$  harbors a community of perfect specialists:  $\sigma_A = \{1, 0, 0 \dots\}$ ,  $\sigma_B = \{0, 1, 0 \dots\}$ , etc. Faced with an uneven supply of resources, this island will exactly adjust expression levels  $T_i$  to track the supply vector  $R_i$ , so that  $T_i = R_i/\chi_i$ , where  $\chi_i$  is the cost of each respective specialist. Our external observer will conclude that island  $\alpha_C$  is able to sense its environment and up-regulate or downregulate individual pathways.

Such perfect regulation is, however, costly: typically,  $\sigma_A$ ,  $\sigma_B$ , etc. will not be the most cost-efficient combinations. As a result, allowing the island  $\alpha_S$  to evolve while holding  $\vec{R}$  fixed, one will obtain an island  $\alpha_C$  harboring a different multi-organism community  $C$ . Unlike  $\alpha_S$ , it will generally be unable to sense all resources independently and adjust expression patterns accordingly; however, evolution will trade some of the sensing capacity to fit the particular resource influx  $\vec{R}$  with more efficient

pathway combinations.

To an external observer, therefore, each island is a coherent whole with an evolving complex behavior striving to better adjust its response  $\vec{T}$  to the environment  $\vec{R}$  it experiences, and in this particular model, the existence of a function  $F$  satisfying all the properties of a valid fitness function makes this metaphor exact. Importantly, however, the inner regulatory logic of this “coherent whole” has the form of purely “selfish” competition between different motifs of partial expression of pathways, inviting a parallel with other instances of “competition serving the good of the multicellular whole” [16]. To conclude this section, let us compute the community fitness  $F$  of the single-organism community  $\alpha_G$  for the case  $R_i \equiv R$  that was considered throughout this work. Applying the definition (6), and using  $T_i = n_G = NR/\chi_G$  one finds:

$$\begin{aligned}
 F &= \frac{1}{\sum_i R_i} \left( \sum_i R_i \ln \frac{T_i}{R_i} - n_G \chi_G \right) + 1 \\
 &= \ln \frac{N}{\chi_G} = \ln(1 + f_G) \approx f_G
 \end{aligned}$$

where  $f_G$  is the individual fitness (4) of organism  $\sigma_G$ , and the approximate equality holds because  $f_G$  is of order  $\epsilon$ . In other words, for a community consisting of a single organism type, the community fitness coincides with the individual fitness of this type, reinforcing the interpretation of a community as a “smarter” individual that had evolved internal division of labor. This interpretation is

specific to the particular model explored here, but within this model, the metaphor is exact.

## DISCUSSION

It is now time to revisit the definition of individual fitness used in this work. By definition of the dynamics (3), each organism merely pursues its own “selfish interest” in the environment where it happens to be. In each instance of “community competition” assayed in Fig. 4, whenever some organism type invaded a community, it was because its fitness in *that particular environment* was higher than the fitness of types already present. One could therefore argue that the surprising finding of Fig. 4, namely that the competition outcome is better predicted by collective, rather than individual, fitness is entirely a consequence of the unsatisfactory definition of individual fitness that sought to attribute a number to a genotype irrespectively of its environment.

This observation is, of course, correct. However, the view proposed in this work is that the distinction between the individual growth rate and the useful notion of fitness is precisely what underlies the phenomenon of “fitness decoupling” during an evolutionary transition, at least in the model described here. In many circumstances, for an entity placed in a sufficiently stable environment, and sufficiently clearly demarcated from it, its likelihood to survive can be reasonably approximated by a quantity that depends only on its intrinsic traits; this constitutes a “useful” notion of fitness. However, when the environment experienced by individual organisms is variable enough that their relative fitness rank is extensively reorganized with each perturbation, this approximation fails. From this perspective, the two quantities that become “decoupled” are not the individual fitness and collective fitness. Rather, they are the true environment-dependent growth rate and the approximation of individual fitness that would have been useful in other circumstances. Importantly, this effect is a consequence of diversity of interactions and would be obscured by the “historic focus on pairwise interactions decoupled from their more complex community context”, as stressed by the authors of Ref. [17].

If this interpretation of the origin of “fitness decoupling” in this model is correct, then reducing the degree to which environmental perturbations affect relative fitness of individuals should make fitness decoupling less pronounced. To test this prediction, consider the effect of increasing  $\epsilon$ , the parameter that determines the width of the distribution of individual fitnesses. Intuitively, returning to the example used previously, if organism  $\sigma_{AC}$  is more fit than  $\sigma_C$  while resources are equally abundant, reducing the availability of resource  $A$  can inverse this relation, but the larger the difference in organism costs  $\chi_{AC}$  and  $\chi_C$ , the more extreme such resource depletion

would have to be. Therefore, increasing  $\epsilon$  will reduce the relative effect that changing environment has on fitness rank ordering. Fig. 5 repeats the analysis of Fig. 4A for  $\epsilon = 0.1$  (rather than  $\epsilon = 10^{-3}$  used until now). As predicted, the collective fitness is now strongly associated with the fitness of individuals. In fact, a first indication of this phenomenon can already be observed in Fig. 2B: as  $\epsilon$  is increased, the median fitness rank of survivors at the final equilibrium begins to reduce. At high  $\epsilon$ , it is increasingly true that high collective fitness is merely a reflection of high individual fitness of community members.

To summarize, changing  $\epsilon$  tunes the magnitude of fitness “decoupling”, and Fig. 2B can be seen as documenting the transition between a largely individualistic regime (large  $\epsilon$ ) and a regime where the multi-individual, genetically inhomogenous assembly of organisms increasingly acts “as a whole”, in a rigorous mathematical, rather than purely metaphorical sense (small  $\epsilon$ ). Importantly, this “whole” is not bound together by any explicitly cooperative interactions. The community-level description was appropriate not because of a high degree of cooperation between members, but because the community was the smallest unit whose environment was well-defined and fixed (the overall resource influx  $\vec{R}$ ).

The discussion of cooperation requires a comment. Consider yet again the example of a world with 3 resources, where a highly fit organism  $\sigma_{AB}$  makes it possible for  $\sigma_C$  to survive at the final equilibrium despite it relatively high cost. One could say that  $\sigma_{AB}$  provides a benefit to  $\sigma_C$  by suppressing its direct competitors, so that the “alliance”  $\sigma_{AB} + \sigma_C$  is selected over other possible combinations. This scenario, summarized in the formula “enemy of my enemy is my friend” can be interpreted as

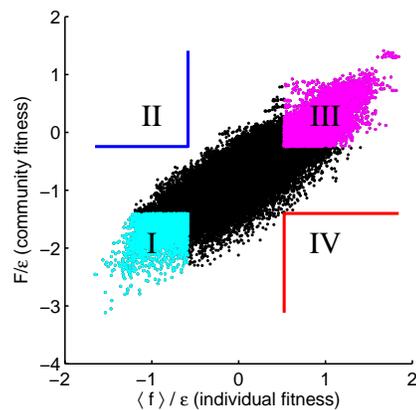


FIG. 5. **Parameter  $\epsilon$  tunes the magnitude of fitness decoupling.** Same as Fig. 4A, for  $\epsilon = 0.1$ . Increasing  $\epsilon$  reduces the relative importance of environment in determining the organisms’ fitness ranking. As a result, collective and individual fitness remain strongly coupled. Defining quadrants as in Fig. 4A leaves the blue and red quadrants empty.

a form of cooperation [17]. Importantly, however, such “cooperation” does not admit cheating. As a result, the form of cohesive multicellular behavior described here is not subject to the same vulnerabilities that make evolution of cooperation in the traditional sense a difficult process. It is a generic consequence of division of labor and can be expected to be a common occurrence in the highly diverse microbial communities that dominate our planet.

In the particular model considered here, the community-wide dynamics happened to take the form of optimizing a global “community fitness”. This lucky feature made it easier to demonstrate how environment-dependent individual fitness translates into selection effectively acting on collective traits, e.g. favoring more efficient alliances of co-evolved organisms rather than organisms themselves. In general, of course, collective dynamics are almost never reducible to optimizing one global function, and one does not expect community to possess a well-defined “fitness”. This, however, is hardly a limitation of the model, as the same can be said about the fitness of individual organisms. Fitness is not a fundamental property of any entity: organisms have traits correlated with their survival, which only translate into fitness under some very specific circumstances when the environment is sufficiently stable and is sufficiently clearly demarcated from the system (“individuality” [2, 9, 18]). Such circumstances can arise via artificial selection in the hands of an experimenter imposing a standardized performance test in a particular environment [6, 7], or as a consequence of evolution, e.g. a pitcher plant or a host animal with its symbiont organisms. It seems likely that the conditions experienced by the symbionts themselves are not in this class, in which case the “behavior as a whole” regime described here may be highly relevant for characterizing consortia of microorganisms within real communities. Some experimental data even provides circumstantial evidence for truly whole-community scale behavior, e.g. the efficacy of fecal matter transplant in treating *C. difficile* infections [20] or “lean microbiota” outcompeting “obese microbiota” in mice [21].

The discussion above relates the results presented here to several themes and ideas widely discussed in the literature, particularly those concerning the role of the environment, co-evolution, and non-transitive competitive structures, reviewed in Refs. [17, 19]. The central aim of this paper was to present a rigorous mathematical structure exhibiting the phenomenon of “fitness decoupling”, to which these ideas can be applied and which can be interpreted both in terms of multi-level selection theory and niche construction theory. It should be noted that the same mathematics can admit different, and sometimes conflicting, interpretations, as illustrated by the ongoing debates in multi-level selection theory [9]. However, the interpretation proposed here suggests that at least some forms of simple multicellularity can be ex-

pected to be a widespread phenomenon that does not require high relatedness or cooperation.

## ACKNOWLEDGMENTS

I thank Ariel Amir, Michael P. Brenner, Simon A. Levin, Anne Pringle, Ned S. Wingreen and David Zwicker for helpful discussions. This work was supported by the Harvard Center of Mathematical Sciences and Applications and the Simons Foundation.

- 
- [1] Maynard Smith J, Szathmary E (1995) *The Major Transitions in Evolution* (Freeman, Oxford).
  - [2] Michod RE, Nedelcu AM (2003) On the reorganization of fitness during evolutionary transitions in individuality. *Integr Comp Biol* 43 (1): 64–73.
  - [3] Hammerschmidt K, Rose CJ, Kerr B, Rainey PB (2014) Life cycles, fitness decoupling and the evolution of multicellularity. *Nature* 515 (7525): 75–.
  - [4] Levin SR, Brock DA, Queller DC, Strassmann, JE (2015) Concurrent coevolution of intra-organismal cheaters and resisters *J Evol Biol* 28(4): 756–765.
  - [5] Michod RE (1999) *Darwinian dynamics: evolutionary transitions in fitness and individuality* (Princeton University Press, Princeton, NJ).
  - [6] Swenson W, Wilson DS, Elias R (2000) Artificial ecosystem selection. *Proc Natl Acad Sci U S A* 97(16): 9110–9114.
  - [7] Swenson W, Arendt J, Wilson DS (2000) Artificial selection of microbial ecosystems for 3-chloroaniline biodegradation. *Environ Microbiol* 2:564571.
  - [8] Williams HTP, Lenton TM (2007) Artificial selection of simulated microbial ecosystems. *Proc Natl Acad Sci U S A* 104(21): 8918–8923.
  - [9] Okasha S (2006) *Evolution and the levels of selection* (Clarendon Press, Oxford).
  - [10] Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R (2012) Diversity, stability and resilience of the human gut microbiota. *Nature* 489: 220–230.
  - [11] Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486(7402): 207–214.
  - [12] Gilbert JA, Jansson JK, Knight R (2014) The Earth Microbiome project: successes and aspirations. *BMC Biology* 12: 69.
  - [13] Tikhonov M (2015) Theoretical ecology without species. arXiv:1504.02550
  - [14] Mac Arthur R (1969) Species packing, and what interspecies competition minimizes. *Proc Natl Acad Sci U S A* 64(4): 1369–71.
  - [15] McIntire EJB, Fajardo A (2014) Facilitation as a ubiquitous driver of biodiversity. *New Phytol* 201(2): 403–416.
  - [16] Moreno E, Rhiner C (2014) Darwin’s multicellularity: from neurotrophic theories and cell competition to fitness fingerprints *Curr Opin Cell Biol* 31: 16–22.
  - [17] Hay ME, Parker JD, Burkepile DE, Caudill CC, Wilson AE, Hallinan ZP, Chequer AD (2004) Mutualisms and

- aquatic community structure: The enemy of my enemy is my friend. *Annu Rev Ecol Evol Syst* 35: 175–197.
- [18] Buss LW (1987) *The evolution of individuality* (Princeton University Press, Princeton, NJ).
- [19] Calcott B, Sterelny K, eds. (2011) *The major transitions in evolution revisited* (MIT Press, Cambridge, MA).
- [20] Bakken JS, Borody T, Brandt LJ, Brill JV, Demarco DC, Franzos MA, et al. (2011). Treating *Clostridium difficile* infection with fecal microbiota transplantation. *Clin Gastroenterol Hepatol* 9(12): 1044–1049.
- [21] Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, et al. (2013) Gut microbiota from twins discordant for obesity modulate metabolism in mice *Science* 341(6150): 1241214.

SUPPLEMENTARY MATERIAL

**Lyapunov function for community dynamics**

This section reproduces portions of the supplementary material to [13]. Consider the following objective function:

$$\tilde{F} = \sum_i R_i \ln T_i - \sum_{\vec{\sigma}} \chi_{\vec{\sigma}} n_{\vec{\sigma}}, \quad (\text{S1})$$

defined for  $\{n_{\vec{\sigma}} \geq 0\}$ , and differing from the definition of Eq. (6) only by normalization.

**$\tilde{F}$  is bounded from above**

To see this, note the inequalities:

$$\sum_i T_i = \sum_{\vec{\sigma}} |\vec{\sigma}| n_{\vec{\sigma}} \leq N \sum_{\vec{\sigma}} n_{\vec{\sigma}}$$

and for  $\alpha, \beta > 0$ :

$$\alpha \ln x - \beta x \leq \alpha \ln \frac{\alpha}{e\beta}$$

Using these, and setting  $\min_{\vec{\sigma}} \chi_{\vec{\sigma}} = \chi^* > 0$ , one can write:

$$\begin{aligned} \tilde{F} &\leq \sum_i R_i \ln T_i - \chi^* \sum_{\vec{\sigma}} n_{\vec{\sigma}} \leq \sum_i \left( R_i \ln T_i - \frac{\chi^*}{N} T_i \right) \\ &\leq \sum_i R_i \ln \frac{NR_i}{e\chi^*} \end{aligned}$$

**$\tilde{F}$  is convex**

To see this, note that for any function  $f(\vec{n})$ , the following two operations leave its convexity invariant ( $M$  is an arbitrary matrix):

1. adding a linear function of its arguments:

$$f(\vec{n}) \mapsto g(\vec{n}) = f(\vec{n}) + M\vec{n};$$

2. performing a linear transformation of its arguments:

$$f(\vec{n}) \mapsto h(\vec{n}) = f(M\vec{n}).$$

Given these observations, convexity of  $\tilde{F}$ , and therefore also the convexity of  $F$  as defined in (6), directly follows from the convexity of the logarithm.

**$\tilde{F}$  is increasing on trajectories**

Since  $T_i = \sum_{\vec{\sigma}} \sigma_i n_{\vec{\sigma}}$ , the gradient of  $\tilde{F}$  is precisely the ‘‘resource surplus’’ experienced by different organism types:

$$\frac{\partial \tilde{F}}{\partial n_{\vec{\sigma}}} = \sum_i \frac{R_i}{T_i} \sigma_i - \chi_{\vec{\sigma}} = \Delta \varphi_{\vec{\sigma}}$$

Therefore,  $\tilde{F}$  is always increasing along the trajectories of the model:

$$\frac{d\tilde{F}}{dt} = \sum_{\vec{\sigma}} \frac{\partial \tilde{F}}{\partial n_{\vec{\sigma}}} \dot{n}_{\vec{\sigma}} = \sum_{\vec{\sigma}} \frac{n_{\vec{\sigma}}}{|\vec{\sigma}|} (\Delta \varphi_{\vec{\sigma}})^2 \geq 0.$$

For any initial community state  $C$ , ecological dynamics converge to the equilibrium corresponding to the unique maximum of  $\tilde{F}$  on the domain  $\{n_{\vec{\sigma}} \geq 0 \text{ for } \vec{\sigma} \in \Omega(C)\}$ . Since  $\tilde{F}$  is bounded and convex, the final equilibrium always exists and is unique and stable.

**Normalization of community fitness**

The typical value of  $\tilde{F}$  as defined in equation S1 for a community close to equilibrium can be estimated as follows.

To estimate the first term, note that the cost per pathway of all organisms is close to  $\chi^* = 1$ , and therefore the overall expression  $T_i$  is approximately  $T_i \approx R_i / \chi^* = R_i$ .

The second term is the total cost of all organisms in the population  $\sum_{\vec{\sigma}} n_{\vec{\sigma}} \chi_{\vec{\sigma}}$ . At any equilibrium, it is equal to the total resource abundance  $R_{\text{tot}} \equiv \sum_i R_i$ . This can be seen in two ways. One approach is to use the equilibria conditions to express the cost of all present organisms in terms of resources:

$$\forall \vec{\sigma} \in \Omega(C): \chi_{\vec{\sigma}} = \sum_i \sigma_i \frac{R_i}{T_i}$$

Therefore,

$$\sum_{\vec{\sigma}} n_{\vec{\sigma}} \chi_{\vec{\sigma}} = \sum_i \left( \sum_{\vec{\sigma}} n_{\vec{\sigma}} \sigma_i \right) \frac{R_i}{T_i} = \sum_i R_i.$$

Alternatively, this same equation can be derived from the condition of maximization of  $\tilde{F}$ , by setting  $n_{\vec{\sigma}} \equiv Mp_{\vec{\sigma}}$ , and requiring  $\frac{\partial \tilde{F}}{\partial M} = 0$ .

Putting these observations together, the expectation for the value of  $\tilde{F}$  at any equilibrium is therefore

$$\begin{aligned} \tilde{F} &= \sum_i R_i \ln T_i - \sum_{\vec{\sigma}} \chi_{\vec{\sigma}} n_{\vec{\sigma}} = \sum_i R_i \ln T_i - \sum_i R_i \\ &\approx \sum_i R_i \ln R_i - \sum_i R_i \equiv \tilde{F}_0 \quad (\text{S2}) \end{aligned}$$

When defining community fitness, it is natural to subtract this baseline value from  $\tilde{F}$  as defined in (S1), and normalize by  $R_{\text{tot}}$ :

$$F = \frac{\tilde{F} - \tilde{F}_0}{\sum_i R_i}.$$

This is the normalization chosen in equation (6) in the main text.

### **Simulation of dynamics and numerical determination of community equilibria**

Calculations were performed in Matlab (Mathworks, Inc.). Scripts performing calculations and reproducing Figs. 2-5 are available as Supplementary File 1. Briefly, community dynamics were computed using MatLab ODE solver `ode15s`. Availability of all resources was set to  $R_i = 100$ . Equilibria corresponding to a given set of types  $\Omega_0$  were determined as follows. Community was initialized by setting the abundance of all types in  $\Omega_0$  to 1. Dynamics was simulated until the absolute magnitude of all time derivatives  $\dot{n}_{\vec{s}}$  fell below threshold  $10^{-4}\epsilon$ . At this point, most types present in the community are present at vanishingly small abundances that are in the process of exponential extinction. To ensure that all low-abundance types are indeed going extinct, the following protocol was

followed. All types with abundance below  $10^{-4}$  were removed from the population, the pruned community was re-equilibrated (to account for any tiny adjustments this removal might have caused), and the resulting state  $C^*$  was tested for being a non-invadeable equilibrium.

If any of the initially present types in  $\Omega_0$  could invade this community  $C^*$ , these types were re-initialized at abundance 1 and the equilibration process was repeated. This extremely rare occurrence captured the cases where the number of types going extinct was so large that their cumulative effect was sufficient to mask a potential survivor, driven to low abundance during the early stages of equilibration but now increasing. If no invaders were detected, the state  $C^*$  was identified as the equilibrium corresponding to the set of types  $\Omega_0$ . This protocol ensures that in the community  $C^*$ , the list of survivors is exact, and their abundance is within acceptable numerical error. The protocol always converged due to convexity of “community fitness”  $F$ .

### **Initial conditions for Fig. 3**

The trajectories displayed in Fig. 3 were simulated for time  $T = 10^6$  starting from 10 random initial conditions whereby each of the 1023 types was set to an abundance value drawn out of a log-uniform distribution between  $10^{-5}$  and 100.