Coarse-graining ecological dynamics in the face of unknown microscopic details

Jacob Moran
Department of Physics, Washington University in St Louis, St Louis, MO, USA

Mikhail Tikhonov*
Department of Physics and Center for Science and Engineering of Living Systems, Washington University in St Louis, St Louis, MO, USA

(Dated: July 17, 2021)

Any description of an ecosystem necessarily ignores some details of the underlying diversity. What predictions can be robust to such omissions? Here, building on the theoretical framework of resource competition, we introduce an eco-evolutionary model that allows organisms to be described at an arbitrary, potentially infinite, level of detail, enabling us to formally study the hierarchy of possible coarse-grained descriptions. Within this model, we demonstrate that a coarse-graining scheme may enable ecological predictions despite grouping together functionally diverse strains. However, this requires two conditions: the strains we study must remain in a diverse ecological context, and this diversity must be derived from a sufficiently similar environment. Our model suggests that studying individual strains of a species away from their natural eco-evolutionary context may eliminate the very reasons that make a species-level characterization an adequate coarse-graining of the natural diversity.

Any characterization of an ecological state or ecological dynamics, whether experimental or theoretical, is always incomplete: both experiments and models necessarily operate at a coarse-grained level of description that ignores some details. A bigger magnifying glass will always be able to resolve finer distinctions – between subspecies, clonal lineages, or even individuals. Which of these details, if any, are indeed safe to ignore?

The omissions come in two flavors: some are known, and others, more worryingly, unknown. In the former case, we knowingly choose to ignore some features deemed irrelevant for the specific question at hand. At the moment, theory offers little guidance for making such choices systematically, and they usually remain an educated guess of the experimenter or the modeler. On the positive side, the validity of such approximations can, at least in principle, be verified (by including the omitted factor into the model, or controlling for it in a new experiment).

The unknown omissions are a deeper issue, and become especially concerning in the context of microbial ecology. As an example, consider the human gut, a habitat harboring hundreds of species of bacteria, experimental access to which is often limited to proxies like fecal samples [1]. How important is the exact geometry of the gut epithelium? the effect of peristalsis and flow on small-scale bacterial aggregates? the exact role of the vast diversity of uncharacterized secondary metabolites [2, 3]? It seems plausible that the complete list of factors shaping these ecosystems includes many we will never even know about, let alone include in our models. Can any predictions be robust to such ignorance? In physics, the validity of simple effective theories despite the complete microscopic description being unknown, and perhaps even unknowable, is guaranteed by the renormalization group construction. There is no known analog in an ecological context.

It seems reasonable to hope that not all details matter, but the experimental evidence is conflicting. For example, the notion of a bacterial species is undoubtedly useful, despite collapsing together strains that can differ from each other by as much as half of their genome [4]; in fact, by some metrics, the species-level characterization of a community appears to be *too* detailed, and can be coarse-grained further [5]. At the same time, numerous studies have highlighted the role of strain-level variation in shaping the functional repertoire of a population, both in microbial context [6–9] and beyond [10]. Recently, Goyal et al. found that fine-scale variants differing by as few as 100 base pairs can exhibit vastly different dynamics, concluding that strains might indeed be "the relevant unit of interaction and dynamics in microbiomes, not merely a descriptive detail" [11]. Seemingly small details can matter not only when characterizing the organisms, but also when describing the environment. For example, Kinsler et al. [12] report that the relative growth rate of different strains of yeast can be measurably and consistently different in flasks differing only by their shape. Developing methods for systematically constructing and evaluating coarse-grained descriptions appropriate for a given question is an urgent challenge for ecological and eco-evolutionary theory [13].

Here, we introduce a theoretical framework to begin addressing these questions. Placing ourselves in a microbial context for concreteness, we develop an explicit eco-evolutionary model allowing organisms to be described at an arbitrary, potentially infinite, level of detail. Within this framework, we reproduce the notions of core and accessory traits, and reconcile the apparent paradox mentioned above, whereby different strains can behave very differently, yet lumping them together into

* tikhonov@wustl.edu

coarse-grained "operational taxonomic units" (OTUs) often appears to be an adequate description in practice. Remarkably, we will show how the validity of a coarse-graining scheme can even be robust to the presence of unknown factors of a certain kind. However, this holds only as long as the communities we consider remain in their natural environment and retain their natural ecological context. In particular, we will show that if we pluck a set of strains from their natural context, the outcome of their interaction may become formally impossible to predict, requiring a complete (and therefore unattainable) microscopic knowledge.

I. AN ECO-EVOLUTIONARY FRAMEWORK FOR A HIERARCHICAL DESCRIPTION OF THE INTERACTING PHENOTYPES

In order to study the hierarchy of possible coarse-graining schemes for ecosystems, we need an eco-evolutionary framework that would (a) describe players functionally, by a list of characteristics than can be made longer (more detailed) or shorter (more coarse-grained), and (b) recognize that the *complete* list of relevant characteristics of the environment, or the relevant traits of an organism, is virtually infinite.

A. The eco-evolutionary dynamics

A given environment presents various opportunities that organisms can exploit to gain a competitive advantage. Imagine a world where all such opportunities or "niches" are enumerated with index $i \in \{1 \dots L_\infty\}$. The notation L_∞ highlights that in general, one expects this to be a very large number, corresponding to a complete, infinitely detailed microscopic description. A strain μ is phenotypically described by enumerating which of these opportunities it exploits, i.e. by a string of numbers of length L_∞ which we will denote $\sigma_{\mu i}$. For simplicity, we will assume $\sigma_{\mu i}$ to be binary ($\sigma_{\mu i} \in \{0, 1\}$): strain μ either can or cannot benefit from opportunity i. This will allow us to think of evolution as acting via bit flips $0 \mapsto 1$ and $1 \mapsto 0$, corresponding to the acquisition or loss of the relevant machinery ("trait i") via horizontal gene transfer events or loss-of-function mutations.

We will assume that the fitness benefit from carrying trait i is largest when the opportunity is unexploited, and declines as the competition increases. For a given set of phenotypes present in the community, the ecological dynamics are determined by the feedback between strain abundance and niche exploitation: Briefly, the strain abundances n_μ determine the total exploitation level of niches $T_i \equiv \sum_\mu n_\mu \sigma_{\mu i}$. The exploitation levels determine the fitness benefit $h_i \equiv h_i(T_i)$ from carrying the respective trait; we will choose $h_i(T_i)$ of the form $h_i(T_i) = \frac{E_i}{1+T_i/N_i}$. These h_i, in turn, determine the growth or decline of the strains. Specifically, we postulate the following ecological dynamics:

$$\frac{\dot{n}_\mu}{n_\mu} = \sum_i \sigma_{\mu i}\, h_i - \chi_\mu \qquad \text{strain abundance} \tag{1a}$$

$$h_i = h(T_i) \equiv \frac{E_i}{1 + T_i/N_i} \qquad \text{benefit from carrying } i \tag{1b}$$

$$T_i \equiv \sum_\mu n_\mu \sigma_{\mu i} \qquad \text{exploitation of } i \tag{1c}$$

In these equations, the parameters E_i and N_i describe the environment, with E_i being the fitness benefit of being the first to discover the niche i (at zero exploitation $T_i = 0$), and the "carrying capacity" N_i describing how quickly the benefit declines as the exploitation level T_i increases. The quantities χ_μ are interpreted as the "maintenance cost" of being an organism carrying a given set of traits; more on this below.

The dynamics (1) is basically the MacArthur model of competition for L_∞ substitutable "resources" [14–16].[1] To these dynamics we add the stochastic arrival of new phenotypes arising through bit flips ("mutations"). The combined eco-evolutionary process is simulated using a hybrid discrete-continuous method as described in the SI Appendix A. In this way, our eco-evolutionary model is similar to, e.g., Ref. [17]. Note, however, that typically the interpretation of resources in such models is metabolic [18–24]; for example, i might label the different forms of carbon available to a carbon-limited microbial community. Here, we adopt a more general perspective, where environmental opportunities need not be specifically metabolic.

As an example, one way for a strain to survive in chemostat conditions is to develop an ability to adhere to the walls of the device [25]. The wall surface is finite, and provides an example of a non-metabolic limited resource. Similarly, being physically bigger, or carrying a rare toxin could be a useful survival strategy, but in both cases the benefit decreases as the trait becomes widespread in the community. Unlike the forms of carbon, which may be numerous but are certainly countable and finite, the list of exploitable opportunities of this kind could be arbitrarily long ($L_\infty \to \infty$). Modeling fitness benefits as additive, relegating all trait interactions ("epistasis") to the cost term χ_μ is certainly a simplification [see the SI in Ref. 17]. It is also worth noting that the model (1) is special in the fact that it possesses

[1] Note that, by construction, our model allows coexistence of a very large number of phenotypes. In many studies, explaining such coexistence is the aim; here, it is our starting point. Rather than asking how a given environment enables coexistence of a diverse community, we start from the observation that natural communities are extremely diverse; interpret this as evidence for the existence of a very large number of limiting factors, many of which remain unknown; and ask whether such diversity of types can be usefully coarse-grained.

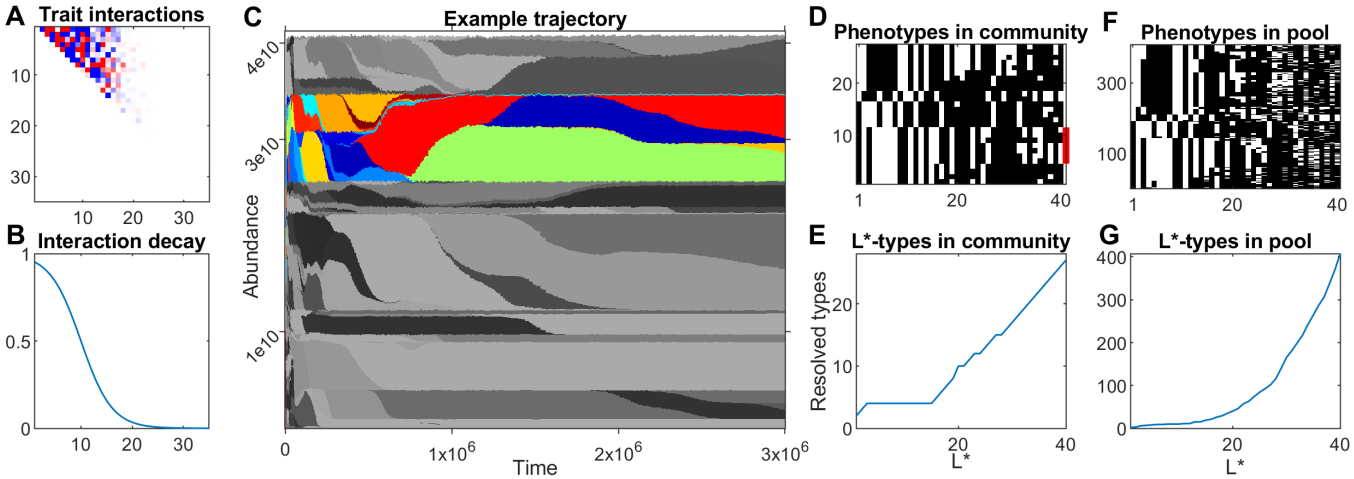

FIG. 1. In the eco-evolutionary model we consider, organisms are described phenotypically by $L_\infty = 40$ binary traits. Carrying a trait incurs a cost but enables the organism to benefit from the corresponding opportunity present in the environment (see text). Organisms engage in ecological competition for the limited opportunities and evolve by gaining or losing traits. **A, B:** The traits carried by a given phenotype interact with each other to determine its "maintenance cost" (see text). Combinations that interact unfavorably are costly to maintain; as a result, not all phenotypes are competitive. The matrix J_{ij} of pairwise interactions is drawn randomly and is the same for all phenotypes, encoding the "biochemical constraints"; panel A shows an example (J_{ij} is triangular with one element per trait pair $i \neq j$). We assume an interaction structure such that a few traits interact strongly while others interact weaker and weaker (panel B). **C:** An example of an eco-evolutionary trajectory in our model. Shading corresponds to different phenotypes; one of the coarse-grained types (see below) is highlighted in color. **D:** The phenotypes present at the endpoint of the trajectory shown in C. Each of 27 phenotypes is a row of length $L_\infty = 40$ (white pixels are carried traits); note the hierarchical structure. The seven highlighted strains are identical in traits 1–24. We will say that they belong to the same "L^*-type", for level of coarse-graining $L^* = 24$. **E:** The number of L^*-types in the community of panel D, shown as a function of L^*. At a coarse-grained level, the community appears to consist of only 4 types (one of these is highlighted in C); resolving finer substructure requires $L^* > 15$. **F, G:** Same as D, E for a broader set of strains, pooled over $N_{\rm env} = 50$ similar environments. The hierarchical structure is maintained. Here, we ask: in what sense, if any, could the phenotypic details beyond $L^* \approx 20$–25 be coarse-grained away in this model?

a Lyapunov function [26]; we will return to this point below. Nevertheless, this is a good starting step for our program, namely understanding the circumstances under which coarse-grained descriptions are adequate. Most crucially, a suitable choice of the cost model χ_μ will allow us to naturally obtain hierarchically structured communities, reproducing the notion of "core" and "accessory" traits [27].

B. A simple cost model leads to hierarchically structured communities

Several studies investigated dynamics like (1) with costs assigned randomly [e.g. 5, 19, 20, 24, 28–30]. We, however, seek to build a model where the phenotypes in the community are not random, but are hierarchically structured, reproducing phenomena such as fine-scale strain diversity found within a species. For this, consider the following cost structure:

$$\chi_\mu = c + \sum_i \chi_i \sigma_{\mu i} + \sum_{i<j} J_{ij} \sigma_{\mu i} \sigma_{\mu j}. \qquad (2)$$

The parameter c encodes a baseline cost of essential housekeeping functions (e.g. DNA replication). χ_i is the cost of carrying trait i (e.g. synthesizing the relevant machinery); for most of our discussion, we will set $c = 0.1$, and set all $\chi_i \equiv \chi_0 = 0.5$ for simplicity. The most important object for us is the matrix J_{ij}, which encodes interactions between traits and shapes the pool of viable (low-cost) phenotypes. As an example, the enzyme nitrogenase is inactivated by oxygen, so running nitrogen fixation and oxygen respiration in the same cell would require expensive infrastructure for compartmentalizing the two processes from each other; in our model, this would correspond to a strongly positive J_{ij} (carrying both traits is costly). For an opposite example, imagine two enzymes that both require an atypically low pH for optimal function. An organism that had already invested into creating such conditions for one can carry the other with relatively little extra investment. Such circumstances would correspond to a beneficial interaction (a negative J_{ij}). Crucially, in our model, the parameters c, χ_i and J_{ij} are the same for all organisms; we will refer to them as encoding the "biochemistry" of our eco-evolutionary world.

We now make our key choice. For our matrix J_{ij}, we generate a random matrix of progressively smaller elements, as illustrated in Fig. 1A. Specifically, we will be drawing the element J_{ij} out of a Gaussian distribution with zero mean and standard deviation $J_0\, f(\max(i,j))$,

with a sigmoid-shaped $f(n) = \frac{1}{1+\exp(\frac{n-n^*}{\delta})}$ (see Fig. 1B). Throughout this work, we set $J_0 = 0.2$, $n^* = 10$ and $\delta = 3$. We claim that this choice enables us to implement the notions of core and accessory traits. Intuitively, since high-cost phenotypes are poor competitors, we can think of the interactions J_{ij} as determining the "sensible" associations of traits. For strongly interacting traits only some combinations are competitive; in contrast, a weakly interacting trait can be gained, lost, or remain polymorphic, as dictated by the environment. An example might be a gene encoding a costly pump that enables the organism to live in otherwise inaccessible regions of (toxin-laden) space. Such a trait is "weakly interacting" if the cost of running the pump does not depend on the genetic background. As we will see, our model will naturally give rise to hierarchically structured sets of strains that share "core" functions but differ in such "accessory" traits, distinctions which we may or may not be able to coarse-grain away.

C. Environment defines a strain pool

To build some intuition about the model defined above, consider Figure 1C that shows an example of these eco-evolutionary dynamics for one random biochemistry, and an environment where we set $E_i \equiv E_0 = 1$ for simplicity, and $N_i = N_0 = 10^{10}$ to set the scale of population size as appropriate for bacteria. The community was initialized with a single (randomly drawn) phenotype. Shading corresponds to distinct phenotypes. Starting from about $t \simeq 10^5$, the dynamics resemble a stable coexistence of several coarse-grained "species" (one is highlighted in color), whose overall abundance remains roughly stable even as individual strains continue to emerge and die out.

As we continue the simulation, the dynamics converge to an eco-evolutionary equilibrium (a state where the coexisting types are in ecological equilibrium, and no single-bit-flip mutant can invade). In this example, it consists of 27 coexisting phenotypes and is shown in Fig. 1D. Note that, confirming our expectations, it appears to possess a hierarchical structure. The seven highlighted strains are identical over the first 24 components, and differ only in the "tail" (components 25-40). A coarse-grained description that characterized organisms only by the first $L^* = 24$ traits would be unable to distinguish these strains; we will say that these strains belong to the same L^*-type with $L^* = 24$. Fig. 1E plots the number of L^*-types resolved at different levels of coarse-graining L^*. For $L^* =$ 3-15, the number of types remains stable at just 4; the color in Fig. 1C highlights one of them. Beyond $L^* = 15$, adding more details begins to resolve additional types, up until $L^* = L_\infty$ when the number of L^*-types coincides with the total number of microscopic strains.

Of course, when discussing the diversity of strains one expects to find in a given environment, it is important to remember that no real environment is exactly static, and no real community is in evolutionary equilibrium. To take this into account while keeping the model simple, we will consider not a single equilibrium, but a collection of communities assembled in $N_{\text{env}} = 50$ similar environments where we randomly perturb the carrying capacity of all niches ($N_i = N(1 + \epsilon\eta_i)$, with $\epsilon = 0.1$ and η_i are i.i.d. from a standard Gaussian) and wait a fixed amount of simulation time (see SI Appendix A 1). Figure 1F shows the set of strains pooled over the 50 ecosystems assembled in this way. This *strain pool* is the central object we will seek to coarse-grain. We stress that its construction explicitly depends on the environment. (Or, more specifically, the particular random set of N_{env} similar environments, but $N_{\text{env}} = 50$ is large enough that the results we present are robust to their exact choice.)

As we see in Figure 1F, adding more strains to the pool makes its hierarchical structure even more apparent. Microscopically, perturbing the environment favors new strains, but at a coarse-grained level, these new strains are variations of the same few types. This is precisely the behavior that we were aiming to capture in our model. Although the number of L^*-types (Fig. 1G) no longer shows a clear plateau, the number of resolved types begins to grow very rapidly beyond $L^* \approx 20 - 25$. Can this diversity be coarse-grained away? Is there a precise sense in which these tail-end traits are "just details"? As we will see, the answer depends on which ecosystem properties we would like our coarse-graining to capture. But first, we need to make this question quantitative.

II. COARSE-GRAINING: IMPLEMENTATION AND EVALUATION

Naively, assessing the validity of a coarse-graining is about comparing two models: one more detailed, one simplified; the coarse-graining is acceptable if the predictions of the two models are within a tolerable margin of error. However, as described in the introduction, this formulation is insufficient. If the complete microscopic description is unattainably complex, any "detailed" model we could ever consider is itself only an approximation of the experimental reality. There can be no model where we account for everything, only an infinite sequence of ever-more-detailed models where we account for more and more factors. This is the situation we can implement in our model.

A. Coarse-graining L-dimensional models of the L_∞-dimensional world

The L_∞-dimensional description we defined will serve as our unattainable "ground truth" representing the complete list of niches and opportunities present in a natural habitat (the *world*). As modelers, we can only construct a hierarchy of increasingly detailed *models* aware of traits 1 through L, with $L < L_\infty$. This parameter L indexes

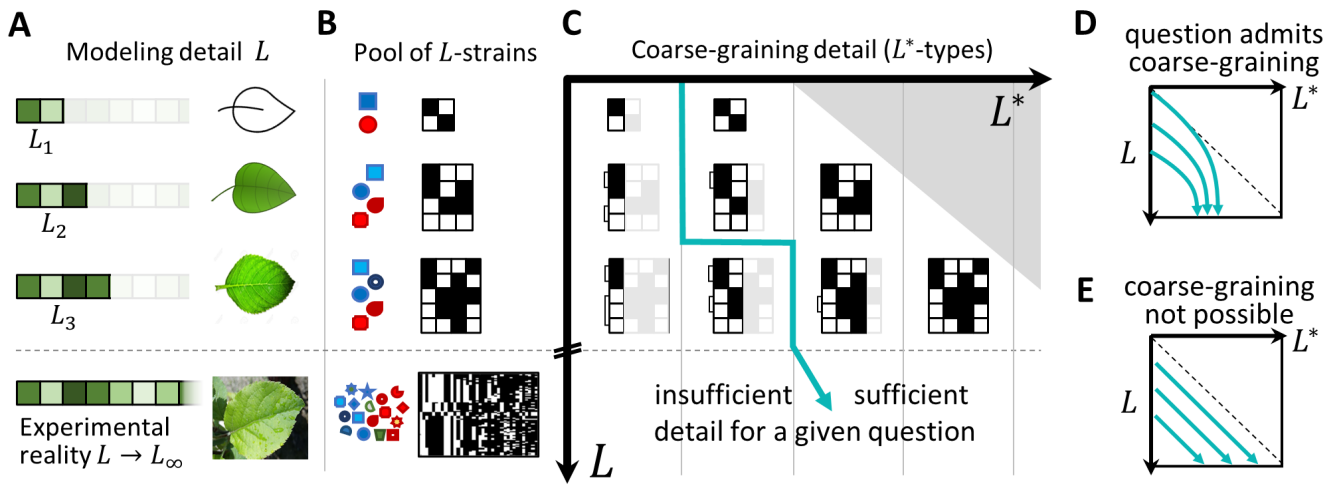

FIG. 2. **Evaluating a hierarchy of coarse-graining schemes.** **A:** Since the complete microscopic description of an ecosystem is unknowable, a coarse-graining scheme must be evaluated against a hierarchy of ever-more-detailed models indexed by level of modeling detail L. The experimental reality corresponds to $L = L_\infty$ assumed unattainably large. **B:** For each L, the model makes a prediction for the pool of strains we expect to encounter (the pool of "L-strains"). **C:** The set of L-strains can be coarse-grained to a varying level of detail $L^* \leq L$. Consider some criterion identifying when the coarse-graining becomes sufficiently detailed for a given ecological question (cyan line). Whether the question admits a coarse-grained description is determined by the asymptotic behavior of this separating line as $L \to \infty$. **D, E:** More generally, whether a question admits coarse-graining is encoded in the behavior of the isolines of a (question-dependent) coarse-graining quality metric $Q(L, L^*)$.

the level of *modeling detail* (Fig. 2A). For any L, the corresponding model will predict which L-dimensional phenotypes (which we will call L-strains) can stably coexist when competing for the L known niches. Note that increasing modeling detail can both refine such predictions or invalidate them: including a yet another niche in the model may reveal that an L-phenotype previously believed to be viable cannot actually survive in this environment (Fig. 2B).

Within the context of any one model, we can compare the performance of various coarse-grained descriptions of length $L^* \leq L$, whereby the L-strains are grouped into L^*-types; we will refer to L^* as the level of *coarse-graining detail*. Let $Q(L, L^*)$ be some measure of coarse-graining quality, which will manifestly depend on the question of interest; we will discuss specific examples below. Within any one L-model, increasing L^* will of course provide a better approximation. The key question, however, is how $Q(L, L^*)$ behaves for a fixed L^*, with increasing L.

Indeed, to say that a given question admits coarse-graining is to say that a desired level of accuracy can be achieved with a fixed L^* even as new, previously unknown factors are added to the model. Mathematically, this "robustness to ignorance" is now encoded in the behavior of the isolines of $Q(L, L^*)$ (Fig. 2D-E). Remarkably, as we show below, coarse-grainable questions do exist. However, as we will see, this property is context-specific: a coarse-graining that works in the organisms' natural eco-evolutionary context is easily broken if community is assembled in the non-native environment or if the natural ecological diversity is removed. To see this, we turn to specific metrics of coarse-graining quality $Q(L, L^*)$, corresponding to different questions of interest.

B. Defining coarse-graining quality $Q(L, L^*)$

Consider a community at ecological equilibrium for simplicity. A given L^* specifies a grouping of L-strains into a smaller number of coarse-grained classes (Fig. 3A). Instead of describing the community composition microscopically, we can describe it in a coarse-grained way, specifying the identity and abundance of the L^*-types that are represented. How do we assess whether such a description constitutes a "good" coarse-graining?

The reconstitution test: One possible criterion is the *reconstitution test*. Drawing a random representative for each of the L^* types in the strain pool, we seed an identical environment with the representatives we chose, allowing them to reach an ecological equilibrium (Fig. 3B). If the details ignored by the coarse-graining are indeed irrelevant, we would expect such "reconstituted" replicates to all be alike. If the reconstituted communities are found to be highly variable depending on exactly which representative we happened to pick, this will signal that the distinctions we attempted to ignore are, in fact, significant.

Quantitatively, for each L^*-type μ_*, let us denote $n_{\mu_*}^{(\alpha)}$ its final abundance in the reconstituted replicate α. The standard deviation of $n_{\mu_*}^{(\alpha)}$ over α (denoted std_α) pro-

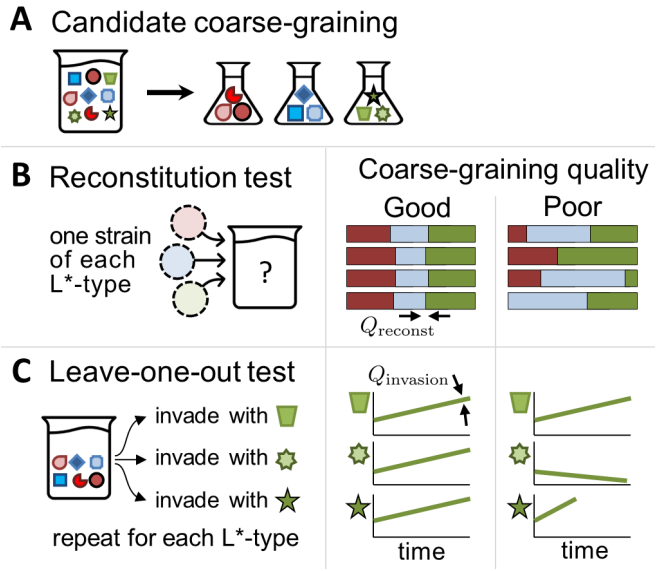

FIG. 3. **Specific criteria for assessing coarse-graining quality** $Q(L, L^*)$**. A:** In this cartoon, the community is coarse-grained into three OTUs (L^*-types in our model). **B:** The reconstitution test. Under this criterion, grouping strains into coarse-grained OTUs is justified if reconstituting a community from a single representative of each OTU yields similar communities regardless of which representatives we pick. As a quantitative measure, we compare the OTU abundances across replicates. **C:** The "leave-one-out test". Under this criterion, grouping strains into coarse-grained OTUs is justified if the strains constituting OTU X (green in this cartoon) all behave similarly when introduced into a community missing X. As a quantitative measure, we compare the invasion rates of the left-out strains.

vides a natural measure of variability across replicates. To combine these into a single number, we compute the average such variability over all L^*-types μ_*, weighted by their total abundance N_{μ_*} in the pool (i.e. the combined abundance of all strains belonging to μ_* observed across the set of N_{env} environments used to define the pool):

$$Q_{\text{reconst}} = \frac{\sum_{\mu_*} N_{\mu_*} \ \text{std}_\alpha \left[n_{\mu_*}^{(\alpha)} \right]}{\sum_{\mu_*} N_{\mu_*}}.$$

A perfect reconstitution would have $Q_{\text{reconst}} = 0$. Conveniently, this is automatically the case if $L^* = L$ (no coarse-graining).

The leave-one-out test: As we will see, the criterion defined above is extremely stringent and is rarely satisfied. In this section, we introduce a weaker version that we will find more useful. Instead of requiring the strains grouped together to be interchangeable in absolute terms, we will ask that they behave similarly *in the context of the assembled community.*

Specifically, for a given scheme grouping strains into coarse-grained types, consider assembling a community missing a particular coarse-grained type μ_* (the ecological equilibrium reached when combining all the strains in the pool, except those belonging to type μ_*; Fig. 3C). We will judge the coarse-graining as consistent if the different strains constituting the missing type μ_* all behave similarly when introduced into this community. As one example, we can compare their initial growth rates if introduced into the community at low abundance, called henceforth "invasion rate" (other possible choices include the abundance the strain will reach if established, or the level of niche exploitation in the resulting community; these are considered in the SI Fig. S1). If the invasion rates are similar, describing the community as missing the coarse-grained type μ_* would indeed be consistent. If, however, the invasion rates vary strongly, we will conclude that the features our coarse-graining is neglecting are, in fact, important.

Quantitatively, denote the invasion rate of strain μ into a community missing type μ_* as r_{μ,μ_*}. We define

$$Q_{\text{invasion}} = \frac{\sum_{\mu_*} N_{\mu_*} \ \text{std}_{\mu \in \mu_*} \, r_{\mu,\mu_*}}{\sum_{\mu_*} N_{\mu_*}},$$

where $\text{std}_{\mu \in \mu_*}$ denotes the standard deviation over all strains belonging to type μ_*. Once again, at $L^* = L$ we automatically have $Q_{\text{invasion}} = 0$.

To illustrate the difference between the two criteria, consider the statement that a community consisting of *Tetrahymena thermophila* and *Chlamydomonas reinhardtii* cannot be invaded by *Escherichia coli* [31]. What meaning should we ascribe to this statement when phrased in terms of coarse-grained units, rather than specific strains? Under the first criterion, we would require that if we combine any single strain of *T. thermophila*, any strain of *C. reinhardtii*, and any strain of *E. coli*, only the first two would survive. Under the second criterion, we would combine a vial labeled *T. thermophila*, containing the entire diverse ensemble of its strains, with a similarly diverse vial of *C. reinhardtii*, and verify that the resulting community cannot be invaded by any individual strain of *E. coli*.[2]

Note that in our model, the existence of a Lyapunov function [26] means the ecological equilibrium is uniquely determined by the environment and the identity of the competing strains; their initial abundance or the order of their introduction does not matter. While this is a simplification, this property is very useful for our purposes, since any lack of reproducibility between reconstituted communities is then clearly attributable to faulty coarse-graining. In a model where even identical phenotypes could assemble into multiple steady states, distinguishing this variability from the variability due to strain differences would add a layer of complexity to our analysis.

[2] Although we use this as an example here, we should note that in the original reference [31] this statement was not actually meant as a species-level claim, but indeed referred to the three specific strains used in the experiment.

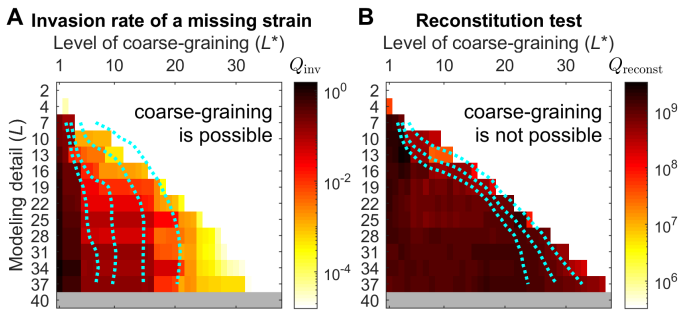

FIG. 4. **A coarse-graining scheme may enable predictions even when grouping functionally diverse strains. A:** If the quantity we seek to predict is the invasion rate of a phenotype missing from the community (the leave-one-out test), this question is readily coarse-grainable. The acceptable level of coarse-graining is determined by the error bar we are willing to tolerate (different isolines of Q), but is robust to including more modeling detail (compare to Fig. 2D). Nevertheless, our ability to ignore tail-end traits for this purpose does not make them any less relevant in other circumstances. **B:** In particular, no amount of coarse-graining is acceptable for the reconstitution test (compare to Fig. 2E). Both heatmaps represent a single random biochemistry; isolines are averaged over 10 biochemistries (see SI Appendix B). The row $L = 40$ is grayed out as a reminder that we treat the complete microscopic description $L = L_\infty$ as unattainable.

III. RESULTS

A. A coarse-graining may enable predictions despite grouping functionally diverse strains

Fig. 4A plots $Q(L, L^*)$ for the leave-one-out test comparing the invasion rates of different strains falling into the same coarse-grained types. We find that specifying a desired accuracy determines a sufficient L^*. Although increasing the level of modeling detail will continue to resolve more and more strains, these differences have almost no effect on their invasion rates. Since all the strains in the same L^*-type behave similarly by our metric, grouping them together appears justified.

And yet, it would be wrong to conclude that the traits beyond a given L^* are safe to ignore. This is clearly demonstrated by the reconstitution test (Fig. 4B). If we attempt to reconstruct the community from its members, *every* detail matters: no amount of coarse-graining is acceptable. We will now explain this apparent paradox within our model.

Consider a community at an ecological equilibrium, and let us focus on a particular phenotype σ_i carrying one of the weakly interacting (tail-end) traits i_0: $\sigma_{i_0} = 1$. What would be the fitness effect of losing this trait? Losing the benefit h_{i_0} from opportunity i_0 is offset by the reduction in maintenance cost; for a weakly interacting trait, the contribution from the term $\sum_j J_{ji_0}\sigma_j\sigma_{i_0}$ is negligible, and the change in cost is simply χ_{i_0}. We conclude that the fitness effect of losing the trait is $\delta f = \chi_{i_0} - h_{i_0}$. At an evolutionary equilibrium, we would therefore have $h_{i_0} = \chi_{i_0}$ (the "functional attractor" state [24]).

Even though our community is not at the evolutionary equilibrium, the simulations show that a sufficiently diverse strain pool will similarly ensure that $h_{i_0} \approx \chi_{i_0}$ for weakly interacting traits; we will say that such niches are "equilibrated". The key observation, then, is that whenever a weakly interacting niche is equilibrated, carrying the respective trait becomes effectively neutral. In particular, the ability of a strain to invade is entirely determined by its phenotypic profile over non-equilibrated niches, explaining the results of Fig. 4A.

Crucially, this approximate neutrality applies only in the environment created by the assembled community, and does not mean that the distinctions are functionally negligible. For instance, consider the (Lotka-Volterra-style) interaction term for a given pair of strains $\mu \neq \nu$:

$$A_{\mu\nu} \equiv \frac{1}{n_\mu}\frac{\partial \dot{n}_\mu}{n_\nu} = \sum_i \sigma_{\mu i}\sigma_{\nu i}\frac{h_i^2}{E_i N_i} = \frac{\sum_i \sigma_{\mu i}\sigma_{\nu i} h_i^2}{E_0 N_0},$$

where we substituted $E_i \equiv E_0$ and $N_i \equiv N_0$ for our environment. Even when tail-end niches are equilibrated with $h_i \approx \chi_i = \chi_0$, we find that each of them contributes equally to the interaction term: no detail is negligible.

This argument directly relates the observed effect to the distinction between a trait that is truly neutral, and one that is effectively neutral in the assembled community only. A truly neutral trait, one incurring almost no cost and bringing almost no benefit, would have $h_i \to 0$ and its contribution to the interaction term $A_{\mu\nu}$ would indeed be small. And indeed, if we repeat our analysis for a scenario where both E_i and χ_i decline with i, we find that neglecting the tail-end traits becomes an adequate coarse-graining also for the reconstitution test (see SI, Fig. S2).

The conclusion from contrasting Fig. 4A and 4B is worth emphasizing. In the example we constructed, the coarse-grained description is valid *sensu* panel 4A. This means that, for instance, we can meaningfully say that "a community assembled of OTU#1 and OTU#2 can be invaded by OTU#4". We can even measure e.g. the invasion rate, and be assured that it is quantitatively reproducible, with a bounded error bar, across the many strains that actually constitute OTU#4 at the microscopic level. Despite all this, the *interaction* between the OTUs as coarse-grained units is not actually definable: any specific pair of strains of OTU#1 and OTU#4 will interact differently with each other, as is indeed observed experimentally [11].

B. Validity of a coarse-graining scheme does not transfer between environments

Above, our focus was to show that the "appropriate" way to coarse-grain a pool of strains depends on the question of interest. In this final section, we will show that

it depends also on the environment where the strains are placed. As we modify the native environment from which the strain pool is derived, the details that were previously negligible will cease to be so.

To see this, let us briefly recall the logic leading to Fig 4A. We considered a set of similar environments; used them to define a strain pool; and showed that certain ecological properties of these strains, such as their invasion rate into an assembled community, did not require a full microscopic knowledge, but could be well captured with a coarse-grained description. As we explained, the reason many traits could be neglected was because the diverse set of strains was able to successfully equilibrate the remainder of weakly interacting niches. However, as we are about to see, this is only true if the diversity had evolved in (was sampled from) a sufficiently similar set of environments.

In Fig. 5, the same pool of strains is used to populate environments that are increasingly further away from the native environment $N_i = N_0$, $E_i = E_0$. Specifically, we use

$$N_i = \begin{cases} N_0 & \text{for } i < 21 \\ N_0(1 + \epsilon\, \eta_i) & \text{for } i \geq 21 \end{cases}$$

where η_i are drawn from the standard normal distribution. All E_i are left at $E_i = E_0$ for simplicity. The figure shows the performance of different L^*-coarse-grainings at predicting the invasion rate of a missing type (the leave-one-out test) under the model with $L = 37$ (chosen to match the last row of Fig. 4A).

At $\epsilon = 0$, this is the same as the last row of Fig. 4A; we see that if we are willing to tolerate an error bar of 10^{-2} on the invasion rate, we need only measure 19 traits. However, as the environment is modified, the same coarse-graining becomes insufficient. This is despite the fact that the changes in the environment were restricted to the supposedly negligible, and potentially unknown, niches 21–40. In summary, we have shown that the conditions under which functionally diverse strains can be meaningfully grouped into coarse-grained units include two requirements, both all of which are necessary: the strains we study must remain in a diverse ecological context, and this diversity must be derived from a sufficiently similar environment.

IV. DISCUSSION

In this work, we defined an eco-evolutionary model enabling us to describe interacting phenotypes in a hierarchical manner, increasing or reducing the level of detail as desired. We used this setup to define a hierarchy of coarse-grained descriptions, and demonstrated that a coarse-graining can be adequate for answering some ecological questions even if the microscopic strains it groups together can be shown to behave very differently from each other.

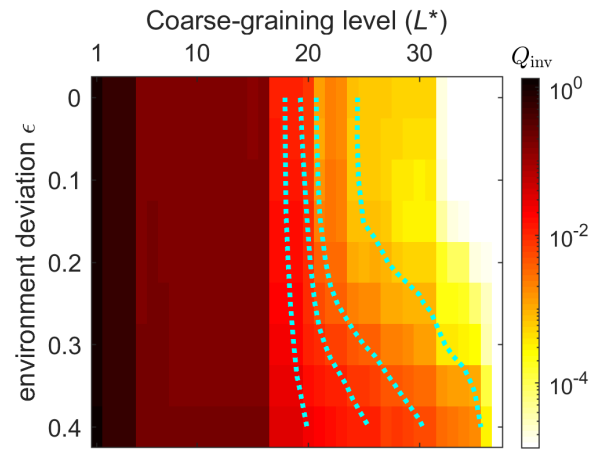

FIG. 5. **A coarse-graining scheme works best in the environment from which the strain pool is derived.** The same strain pool as in Fig. 4A is placed into an increasingly different environment, with differences restricted to the tail-end components 21-40 (see text). The coarse-graining quality is assessed by the leave-one-out test, and shown as a function of L^* and environment difference ϵ (L is fixed at $L = 37$ to match the last row of Fig. 4A). As the environment is modified, the traits that were previously negligible can no longer be coarse-grained. Heatmap and isolines shown for same random biochemistry as in Fig. 4A; for $\epsilon \neq 0$ each pixel is averaged over 5 environments.

One way to approach the coarse-graining problem is to only group together the individuals that are to a sufficient extent interchangeable. This is the criterion we introduced as a "reconstiution test", and is the criterion implicitly assumed by virtually all compositional models of ecosystem dynamics. However, the existing experimental evidence [6–9, 11] suggests that unless we are willing to resolve types differing by as few as 100 bases, this criterion is likely violated in most practical circumstances. It is certainly violated when grouping strains into taxonomic species or families [4, 32–35]. One expects, therefore, that explaining the practical successes of such descriptions would require a different definition of what makes a coarse-graining scheme adequate.

We proposed that this can be achieved with only a subtle change to the criterion: namely, by requiring that the grouped strains be approximately interchangeable not in all conditions, but in the conditions created by the assembled community itself. As long as the strains we study remain in a diverse ecological context, and as long as this diversity is derived from a sufficiently similar environment, we find that the coarse-grained description can be consistent in the sense that certain properties of interest (invasion rate, post-invasion abundance...) are approximately consistent across the strains grouped together. Crucially, the residual variability within a coarse-grained class is determined by the level of coarse-graining, and not the level of modeling detail. The validity of a coarse-grained description is retained even as we include new

microscopic details into the model, such as additional niches previously unaccounted for.

In this paper, we focused on a case where the traits were differentiated only by the strength of their interactions, which established a unique hierarchy among them (a clear order in which to include them in the hierarchy of models, or the hierarchy of coarse-grained descriptions). In the more general case, the trait cost χ_i, or the trait usefulness in a given environment (E_i, N_i) will set up alternative, potentially conflicting hierarchies. We expect the model to have a rich phenomenology in this regime, which we have not considered here. Another obvious limitation of our analysis is that our model includes only competitive interactions. A simple way to extend our framework would be to include cross-feeding interactions. Although such a framework would still not be entirely general, consumer-resource models with cross-feeding were shown to capture a surprisingly wide set of experimental observations [5, 20, 23, 29]. We leave this extension of the model for future work.

In conclusion, there are many reasons to believe that analyzing a species in artificial laboratory environments might be of limited utility for understanding its function or interactions in the natural environment [13]. Usually, however, the concern is that the laboratory conditions are too simple, and in reality, many more details may matter. Here, we use our model to propose that the opposite can be true: understanding the interaction of two strains in the foreign conditions of the Petri dish may require a much more detailed knowledge of microscopic idiosyncracies. Removing individual strains of a species from their natural eco-evolutionary context may eliminate the very reasons that make a species-level characterization an adequate coarse-graining of the natural diversity.

ACKNOWLEDGMENTS

We thank J. Grilli, C. Holmes, and C. Strandkvist for helpful discussions.

[1] T. Yatsunenko, F. E. Rey, M. J. Manary, I. Trehan, M. G. Dominguez-Bello, M. Contreras, M. Magris, G. Hidalgo, R. N. Baldassano, A. P. Anokhin, *et al.*, Human gut microbiome viewed across age and geography, nature **486**, 222 (2012).

[2] P. Vernocchi, F. Del Chierico, and L. Putignani, Gut microbiota profiling: metabolomics based approach to unravel compounds affecting human health, Frontiers in microbiology **7**, 1144 (2016).

[3] M. R. Wilson, L. Zha, and E. P. Balskus, Natural product discovery from the human microbiome, Journal of Biological Chemistry **292**, 8546 (2017).

[4] O. Lukjancenko, T. M. Wassenaar, and D. W. Ussery, Comparison of 61 sequenced escherichia coli genomes, Microbial ecology **60**, 708 (2010).

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

[6] A. Tett, K. D. Huang, F. Asnicar, H. Fehlner-Peach, E. Pasolli, N. Karcher, F. Armanini, P. Manghi, K. Bonham, M. Zolfo, *et al.*, The prevotella copri complex comprises four distinct clades underrepresented in westernized populations, Cell host & microbe **26**, 666 (2019).

[7] S. Zhao, T. D. Lieberman, M. Poyet, K. M. Kauffman, S. M. Gibbons, M. Groussin, R. J. Xavier, and E. J. Alm, Adaptive evolution within gut microbiomes of healthy people, Cell host & microbe **25**, 656 (2019).

[8] T. D. Lieberman, K. B. Flett, I. Yelin, T. R. Martin, A. J. McAdam, G. P. Priebe, and R. Kishony, Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures, Nature genetics **46**, 82 (2014).

[9] M. A. Lawson, I. J. O'Neill, M. Kujawska, S. G. Javvadi, A. Wijeyesekera, Z. Flegg, L. Chalklen, and L. J. Hall, Breast milk-derived human milk oligosaccharides promote bifidobacterium interactions within a single ecosystem, The ISME journal **14**, 635 (2020).

[10] F. Marroni, S. Pinosio, and M. Morgante, Structural variation and genome complexity: is dispensable really dispensable?, Current Opinion in Plant Biology **18**, 31 (2014).

[11] A. Goyal, L. S. Bittleston, G. E. Leventhal, L. Lu, and O. X. Cordero, Interactions between strains govern the eco-evolutionary dynamics of microbial communities, BioRxiv (2021).

[12] G. Kinsler, K. Geiler-Samerotte, and D. A. Petrov, Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation, Elife **9**, e61271 (2020).

[13] J. Bergelson, M. Kreitman, D. A. Petrov, A. Sanchez, and M. Tikhonov, Functional biology in its natural context: A search for emergent simplicity, Elife **10**, e67646 (2021).

[14] R. MacArthur and R. Levins, The limiting similarity, convergence, and divergence of coexisting species, The american naturalist **101**, 377 (1967).

[15] R. MacArthur, Species packing and competitive equilibrium for many species, Theoretical population biology **1**, 1 (1970).

[16] P. Chesson, Macarthur's consumer-resource model, Theoretical Population Biology **37**, 26 (1990).

[17] B. H. Good, S. Martis, and O. Hallatschek, Adaptation limits ecological diversification and promotes ecological tinkering during the competition for substitutable resources, Proceedings of the National Academy of Sciences **115**, E10407 (2018).

[18] A. Posfai, T. Taillefumier, and N. S. Wingreen, Metabolic trade-offs promote diversity in a model ecosystem, Physical review letters **118**, 028103 (2017).

[19] M. Tikhonov and R. Monasson, Collective phase in resource competition in a highly diverse ecosystem, Physical review letters **118**, 048103 (2017).

[20] R. Marsland III, W. Cui, J. Goldford, A. Sanchez, K. Ko-

rolev, and P. Mehta, Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities, PLoS computational biology **15**, e1006793 (2019).

[21] M. Advani, G. Bunin, and P. Mehta, Statistical physics of community ecology: a cavity solution to macarthur's consumer resource model, Journal of Statistical Mechanics: Theory and Experiment **2018**, 033406 (2018).

[22] A. Altieri and S. Franz, Constraint satisfaction mechanisms for marginal stability and criticality in large ecosystems, Physical Review E **99**, 010401 (2019).

[23] P.-Y. Ho, B. H. Good, and K. Huang, Competition for fluctuating resources reproduces statistics of species abundance over time across wide-ranging microbiotas, bioRxiv (2021).

[24] L. Fant, I. Macocco, and J. Grilli, Eco-evolutionary dynamics lead to functionally robust and redundant communities, bioRxiv (2021).

[25] D. E. Dykhuizen and D. L. Hartl, Selection in chemostats, Microbiological reviews **47**, 150 (1983).

[26] R. Mac Arthur, Species packing, and what competition minimizes, Proceedings of the National Academy of Sciences **64**, 1369 (1969).

[27] H. Tettelin, V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, N. L. Ward, S. V. Angiuoli, J. Crabtree, A. L. Jones, A. S. Durkin, *et al.*, Genome analysis of multiple pathogenic isolates of streptococcus agalactiae: implications for the microbial "pan-genome", Proceedings of the National Academy of Sciences **102**, 13950 (2005).

[28] W. Cui, R. Marsland III, and P. Mehta, Effect of resource dynamics on species packing in diverse ecosystems, Physical Review Letters **125**, 048101 (2020).

[29] R. Marsland, W. Cui, and P. Mehta, A minimal model for microbial biodiversity can reproduce experimentally observed ecological patterns, Scientific reports **10**, 1 (2020).

[30] M. Tikhonov, Community-level cohesion without cooperation, Elife **5**, e15747 (2016).

[31] H. Mickalide and S. Kuehn, Higher-order interaction between species inhibits bacterial invasion of a phototroph-predator microbial community, Cell Systems **9**, 521 (2019).

[32] S. Louca, L. W. Parfrey, and M. Doebeli, Decoupling function and taxonomy in the global ocean microbiome, Science **353**, 1272 (2016).

[33] C. Burke, P. Steinberg, D. Rusch, S. Kjelleberg, and T. Thomas, Bacterial community assembly based on functional genes rather than species, Proceedings of the National Academy of Sciences **108**, 14288 (2011).

[34] B. Segerman, The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories, Frontiers in cellular and infection microbiology **2**, 116 (2012).

[35] R. G. Beiko, Microbial malaise: how can we classify the microbiome?, Trends in microbiology **23**, 671 (2015).

[36] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions, The journal of physical chemistry **81**, 2340 (1977).

[37] E. L. Haseltine and J. B. Rawlings, Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics, The Journal of chemical physics **117**, 6959 (2002).

[38] J. G. Lopez and N. S. Wingreen, Noisy metabolism can drive the evolution of microbial cross-feeding, bioRxiv (2021).

[39] M. M. Desai and D. S. Fisher, Beneficial mutation–selection balance and the effect of linkage on positive selection, Genetics **176**, 1759 (2007).

SUPPLEMENTARY INFORMATION

Appendix A: Simulating eco-evolutionary dynamics

The eco-evolutionary world in which the dynamics take place is described by the environment (constant in time), the biochemistry (also constant in time), and the state of the ecosystem (dynamically evolving). At any given moment of time, the state of the ecosystem is described by the following information: (1) The identity of each of the phenotypes, described microscopically as vectors of length L_∞; (2) The current abundance (population size) of each of these phenotypes. All our simulations are performed using the "ground truth" microscopic model (description of length L_∞). The approximate L-models are implemented by zeroing out the environmental niches from $L+1$ onwards.

At the level of individual bacteria, for any moment in time, the next "event" to occur will be one of the following: (1) an individual dies; (2) an individual divides, giving rise to an identical sibling; or (3) an individual divides, giving rise to a mutant sibling. Of course, an individual-based simulation is both impractical and unnecessary; instead, we think of these dynamics as a combination of purely ecological updates of phenotype abundances (which can be modeled with continuous ODEs), and discrete dynamics whereby some strains go extinct, and others are introduced into the population by mutation.

To implement such discrete events, the standard way is to employ a Gillespie scheme [36]. A slight complication here is that when overlaid with ecological dynamics, the rates of such Gillespie events become time-dependent (a mutation favorable right now may cease to be so as ecological dynamics continue); however, this complication is easily resolved, as standard methods for implementing such a hybrid stochastic-deterministic Gillespie scheme have been developed in the literature [37]; briefly, instead of drawing the "time to next event", one must draw a probability threshold, and propagate the continuous dynamics while integrating the rate of an event to occur, up until that accumulated probability crosses the threshold (see [38] for an introduction that is both short and intuitive). As a result, to describe our simulation we just need to define how the state-dependent rates of such events are computed.

To do so, we adapt for our purposes the results of Ref. [39]. From the evolutionary standpoint, the candidate new strain is a mutant that has a chance of escaping drift and become *established* in the population. The probability of becoming established is proportional to the mutation rate; the population size of the parent strain; and the fitness effect δf of the mutation (i.e. the growth rate[3] of the candidate new strain). Once a strain is established, the stochastic effects become negligible and its subsequent dynamics can be modeled deterministically.

The simulation can be summarized with the following pseudocode:

1. For all single mutants of existing strains, determine the rate at which they would establish in the population, as explicit functions of the abundances of extant strains.

2. Propagate ecological dynamics "for an appropriate length of time" as per standard technique [37].

3. Pick the lucky new strain among the beneficial first mutants; add it to the community at a population size $1/\delta f$; the population of the parent strain is, for consistency, reduced by the same amount.

4. Remove any strain whose abundance is below some predetermined threshold and is declining. (In our simulations, this threshold is set at relative abundance 10^{-6}.)

5. Repeat until the required simulation time has elapsed.

The mutation rate we use is $\mu = 10^{-10}$ per individual per unit time. For context, recall that population size is set by the carrying capacity of environmental niches $N = 10^{10}$ (see Fig. 1C), so our choice corresponds to $\mu N = 1$. However, as explained below, the mutation rate parameter plays only a minor role in our analysis.

1. Evolutionary equilibrium is not required

In our analysis of coarse-graining schemes, the community we study is never technically at an evolutionary equilibrium. We explicitly constructed our procedure to avoid making such an unrealistic postulate. Specifically, note that

[3] Note that in many evolutionary models, the fitness effect is computed relative to the (ever-increasing) mean fitness in the population. For us here, fitness is not an abstract property that could increase arbitrarily, but is directly defined as a growth rate; and limited resources automatically ensure that the population-mean growth rate is zero.

we assemble our strain pool from evolutionary equilibria[4] obtained in *similar* environments, but we then study the interaction of these strains in the original, unperturbed environment, where the condition of evolutionary equilibrium was never imposed.

Of course, we then observe that a sufficiently diverse set of strains derived from sufficiently similar environments assembles into a community that is very close to the evolutionary equilibrium also in the original environment, and this proximity is largely responsible for the behaviors reported in this work. This, however, is not a caveat, but a feature, as we expect the same to be largely true for real communities as well: the large diversity of strains is quite plausibly sufficient to populate the available niches without requiring *de novo* mutations, relying exclusively on the standing variation.

2. Mutation rate is not a key parameter

A corollary of the previous point is that for specifically our purposes here, mutation rate is not a key parameter of our model. Indeed, we only invoke evolution when constructing the strain pool, but each of the combined states is an evolutionary equilibrium, which in this model is guaranteed to be unique. One small caveat is that our evolutionary process is simulated at a finite resolution, and considers first-mutants only. As a result, the trajectory could get stuck in a *locally* non-invadeable equilibrium rather than the unique true one, something that would be enhanced by setting the mutation rate too low. In this way, the evolutionary stochasticity (and thus the mutatiton rate) does technically play a weak role, but we found it to be essentially irrelevant for the parameters used here (specifically, running replicate eco-evolutionary trajectories from random initial phenotypes generated virtually indistinguishable final states).

Appendix B: Averaging in figures

All heatmaps shown in figures correspond to a single random biochemistry. The small amount of "graininess" seen in the heatmaps is caused by the quirks of exactly when the L^* coarse-graining is able to resolve new subtypes; it is shown as is, with no additional averaging or smoothing. In contrast, the isolines overlaid on plots are meant to capture the qualitative trends transcending the quirks of a given biochemistry. They are computed by repeating the same analysis for 10 random biochemistry realizations: the 10 heatmaps are averaged, smoothed with a 1-pixel-wide Gaussian kernel, and the isolines are picked as contour lines of the result.

The only exception to this procedure is Fig. 5 in the main text which investigates the effect of perturbing the environment. In this figure, each pixel is an average over 5 random perturbations (of magnitude specified by epsilon), and the isolines are computed directly from the heatmap shown (smoothed with a 1-pixel-wide Gaussian kernel).

Appendix C: Examples of other ecological properties and their compatibility with L^*-coarse-graining

As described in the main text, the leave-one-out scheme judges a grouping of strains into coarse-grained OTUs as justified if the strains constituting OTU X all behave similarly when introduced into a community missing X. However, the "similarity of behavior" could itself be assessed by a variety of criteria. In the main text, we focused on comparing the invasion rates of the left-out strains. Some equally interesting alternatives include, for instance, the abundance reached by the invading strain (which might be zero if the strain cannot invade), or the niche occupancy in the resulting community.

Invading strain abundance: In this example, we denote n_{μ,μ_*} the abundance reached by strain μ after being introduced in a community missing the L^*-type μ_* (by construction, μ is a representative of μ_*). We look at the variability of n_{μ,μ_*} over all $\mu \in \mu_*$, and define the coarse-graining quality Q_{abd} by the weighted average of this quantity over all L^*-types:

$$Q_{\text{abd}} = \frac{\sum_{\mu_*} N_{\mu_*} \, \text{std}_{\mu\in\mu_*} \, n_{\mu,\mu_*}}{\sum_{\mu_*} N_{\mu_*}}.$$

[4] In the main text, we were careful to state that in practice, we merely wait a fixed simulation time. In the presence of slow dynamical modes due to weakly interacting traits, the state we obtain is indeed not always the true equilibrium, but is very close.

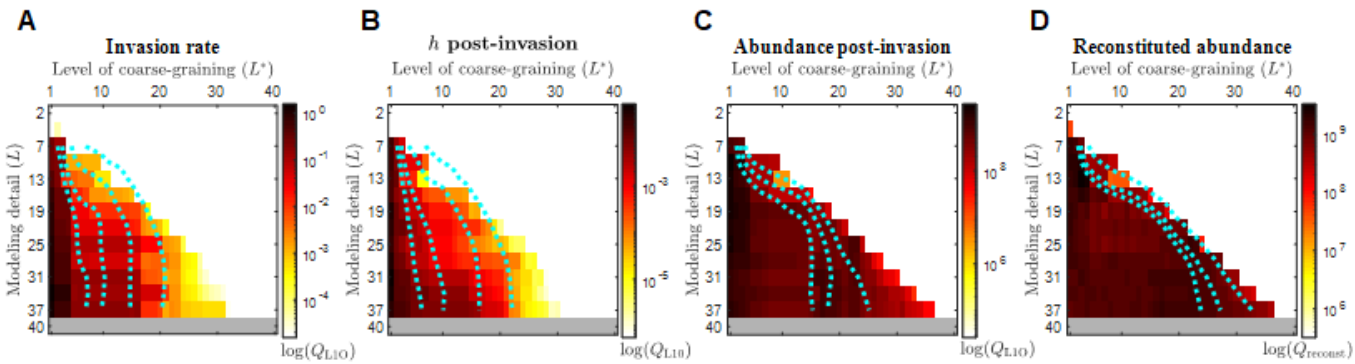

FIG. S1. Four questions whose compatibility with the L^* coarse-graining scheme ranges from excellent to non-existent. **A:** Invasion rate of a missing strain; coarse-grained description sufficient. **B:** Niche occupancy post-invasion; coarse-grained description sufficient. **C:** Abundance reached by a strain post-invasion; coarse-graining compatibility is poor. **D:** Reconstitution test (requires functional equivalence of strains); no coarse-graining is possible. The notation Q_{L1O} highlights that the coarse-graining quality metrics shown in the panels A-C are all assessed in the framework of the leave-one-out test.

Niche occupancy post-invasion: Here, instead of looking at a compositional property (abundance of a particular coarse-grained type), we ask how similar the effect of the invading strains is on the post-invasion *function* of the community as a whole. The functional properties are encoded in community-wide niche occupancy T_i — or, equivalently, in the niche benefit h_i at equilibrium (the benefit from carrying trait i). Since the environment is kept fixed, T_i and h_i are directly related. The reason we choose h over T is because in a metabolic interpretation of the resources (community in a chemostat), h_i is directly measurable as resource concentration in the eluent.

Quantitatively, let $h_i^{(\mu,\mu_*)}$ be the h vector of the equilibrium community after strain μ invades the community missing the L^*-type μ_* (again, μ is a representative of μ_*). For our measure of variability, we compute the average component-wise standard deviation of this vector (across all $\mu \in \mu_*$). We then compute the weighted average of this quantity over all L^*-types as before:

$$Q_h(L, L^*) = \frac{\sum_{\mu_*} N_{\mu_*} \left[\frac{1}{L} \sum_{i=1}^{L} \sqrt{\left\langle \left(h_i^{(\mu,\mu_*)} - \langle h_i^{(\mu,\mu_*)} \rangle_{\mu\in\mu_*} \right)^2 \right\rangle_{\mu\in\mu_*}} \right]}{\sum_{\mu_*} N_{\mu_*}}.$$

In this expression, we use $\langle \dots \rangle_{\mu\in\mu_*}$ to denote averaging over all microscopic representatives of a given L^*-class.

These two new measures of coarse-graining quality, combined with the two considered in the main text, give four metrics $Q(L, L^*)$ encoding four different questions of interest. We chose them to span both compositional and functional properties of a community. How amenable are these four questions to coarse-graining?

The answer is presented in Fig. S1. The panels are ordered by "degree of coarse-grainability" as inferred from the shape of the isolines. Our findings are consistent with the recurring observation in the experimental literature that functional properties tend to be more predictable or reproducible than compositional ones [5, 32, 33]: the hierarchy of the four questions presented in figure S1 can be summarized as saying that functional properties can be adequately predicted with fewer microscopic details.

Indeed, the invasion rate of a missing strain (panel A) is basically a functional property of the pre-invasion community (the growth rate of a strain is determined by the environmental conditions constructed by the rest of the community). Next comes the niche occupancy (panel B) — a functional property assessed post-invasion. Next is the abundance reached by the invading strain (panel C), a compositional property that we expect to be distinctly less coarse-grainable (indeed, recall how strain-strain interactions are strongly affected by tail-end niches). However, the reconstitution test is still last in the list (panel D), epitomizing the goal of a bottom-up compositional description and requiring the knowledge of *all* microscopic details.

Appendix D: Coarse-graining truly neutral traits

One simple way to sanity-check our framework is to make the tail-end niches $i \to L_\infty$ not just weakly interacting, but genuinely close to neutral. A distinction that makes almost no difference in either cost or benefit should surely

be negligible for all questions, including the "reconstitution test". To test this, we repeat the analysis for a model where we use the same declining sigmoid-shaped function $f(n)$ to scale down not only the trait interactions (cf. Fig 1A, B), but also the trait cost χ_i and benefit E_i. The result is shown in Fig. S2. As expected, we find that the L^* coarse-graining is now appropriate for all four of our example questions (this figure is to be compared with Fig. S1).

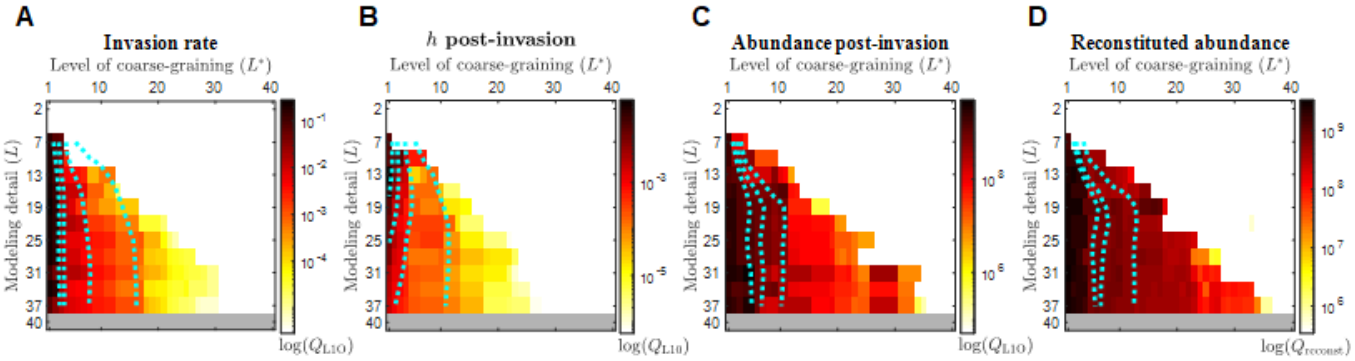

FIG. S2. Reproducing Fig. S1 for a model where tail-end niches are not just weakly interacting, but are also increasingly neutral (bring almost no benefit $E_i \to 0$ and incur almost no cost $\chi_i \to 0$); see text. As expected, such distinctions can be adequately coarse-grained away no matter the question (compare to Fig. S1).