Theoretical microbial ecology without species

Mikhail Tikhonov

Theoretical ecology without species

Mikhail Tikhonov*

Center of Mathematical Sciences and Applications

John A. Paulson School of Engineering and Applied Sciences and

Kavli Institute for Bionano Science and Technology, Harvard

University, 29 Oxford St, Cambridge, MA 02138, USA

Abstract

Ecosystems are commonly conceptualized as networks of interacting species. However, partitioning natural diversity of organisms into discrete units is notoriously problematic, and mounting experimental evidence raises the intriguing question whether this perspective is appropriate for the microbial world. Here, an alternative formalism is proposed that does not require postulating the existence of species as fundamental ecological variables, and provides a naturally hierarchical description of community dynamics. This formalism allows approaching the "species problem" from the opposite direction. While the classical models treat a world of imperfectly clustered organism types as a perturbation around well-clustered "species", the presented approach allows gradually adding structure to a fully disordered background. The relevance of this theoretical construct for describing highly diverse natural ecosystems is discussed.

---

* tikhonov@fas.harvard.edu

Although the basic unit participating in ecological interactions is an individual organism, constructing tractable theoretical models usually requires clustering individuals into discrete groups within which organisms are treated as identical; such classification can be based, for example, on taxonomy, development stage, or phenotype [1]. The resulting tension is a long-standing issue in community ecology: ever since Darwin [2], the difficulty of drawing sharp boundaries partitioning natural diversity into discrete categories [3] and the realization that the individual-level variation can be an important actor in ecological phenomena [4, 5] made the "partitioned community" assumption of well-delimited, uniform groups highly problematic [6–11].

The urgency of this issue has been highlighted over the last decade by the studies of microbial diversity in natural environments [12–17]. For microbes, the partitioning problem is intensified by the prevalence of asexual reproduction and horizontal gene transfer [18–20], and the fact that even genetically identical cells can assume different ecological roles, e.g. in a biofilm [21]. To cope with these issues, considerable attention has been devoted to identifying questions that can be asked and answered without explicitly specifying a partitioning [11], e.g. solely in terms of community metagenome [11, 22, 23]. These approaches adopt a convenient coarse-grained viewpoint that is appropriate for certain functional questions, e.g. comparing communities [22]. However, their ability to describe community dynamics is necessarily limited. For that purpose, the partitioned community assumption currently has no alternatives: although conceptually problematic, it is seen as an operational necessity [6, 24–26]. Short of resorting to individual-based modelling [4], it remains unclear how the dynamics of ecological communities could be described in a naturally hierarchical way [6, 9, 11, 29, 30]).

The hot debate surrounding the definition of the "unit" of ecological diversity prompts an intriguing question: is the partitioned community assumption an adequate description of the microbial world? The species-based intuition is incontestably useful, but might it be forcing onto our data a structure that it does not possess [20, 31, 32]? Ultimately, of course, this question should be settled by experiments, and reports are conflicting [33–37]. However, ideally we should be asking not whether a species-based picture is "adequate", but whether it is superior to alternatives. Can we imagine an ecology where no grouping of individuals is privileged, and how would we describe it? Regardless of our stance on the "species problem", we must recognize this question as an essential theoretical exercise [38]:

2

it is impossible to test a hypothesis unless we can imagine it to be false. Without an alternative, the partitioning assumption will continue to shape our questions, our models and even our data analysis in the form of "operational taxonomic units", an ill-defined proxy for a "bacterial species" extensively critiqued elsewhere [39].

Borrowing an idea from condensed matter physics, this work proposes a new way to tackle the species problem, approaching it "from the other side", as illustrated in Fig. 1. Consider a heterogeneous community where some individuals are more alike than others, but the clustering is imperfect, as is arguably the case in most natural situations [3, 19, 31, 39]. Currently, the only existing approach treats this scenario as a small deviation from the classic picture of well-clustered species. In physical terms, this is a perturbative expansion, where we start from a perfectly partitioned world (Fig. 1A) and attempt to recognize that species boundaries are "fuzzy" (Fig. 1B). This work proposes an explicit construct where the perturbative expansion can be performed around a different origin, starting from an unstructured phenotype background (Fig. 1C), and adding some structure.

Similar ideas have a long history and have been enormously productive in many areas of physics (cf. the classic theory of metals that treats electrons as a free gas). The contribution of this work is to port this idea into an eco-evolutionary context. In order to achieve this, it is necessary to first develop a formalism that avoids postulating a community partitioning. This work extends the notion of an "ecomode" [40, 41] to construct a naturally hierarchical, rank-free description of community dynamics in terms of population fluctuation eigenmodes.
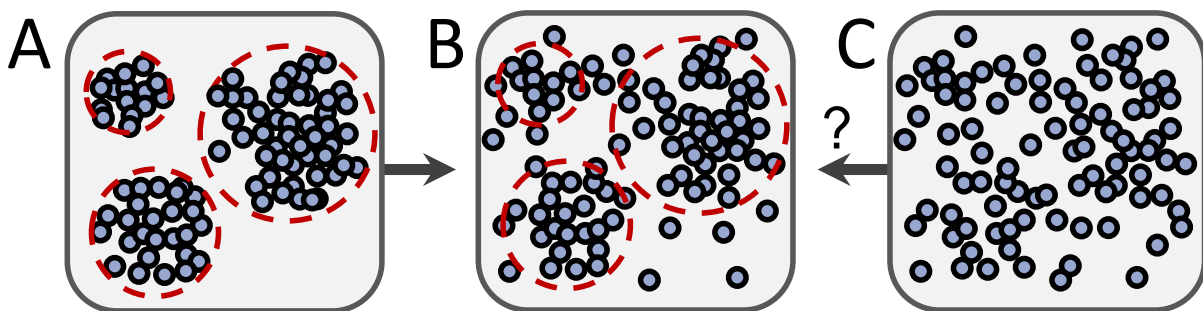

FIG. 1. Clustering organisms into discrete species is often problematic (**B**). Rather than describing the natural world as a deviation from a perfectly clustered case (**A**), this work proposes a theoretical construct where structure can be gradually added to a fully disordered "ecology without species" (**C**).
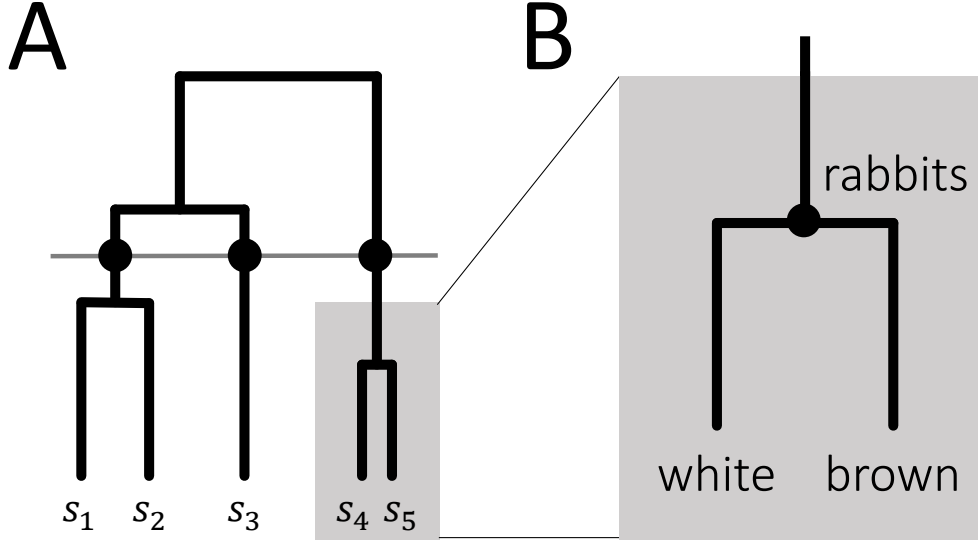
FIG. 2. **Coarse-graining degrees of freedom through clustering. A.** The most common approach to coarse-graining the description of an ecosystem is by merging branches of a phylogenetic tree. **B.** An elementary clustering operation where the abundances of two rabbit varieties are combined into a single degree of freedom.

This approach reduces to the familiar species-based perspective when it applies, but remains well-defined even when the species description breaks down. The basic idea behind this formalism is that species clustering can be reinterpreted as a special form of a more general operation, namely a change of basis in the compositional space.

I. REINTERPRETING SPECIES CLUSTERING AS A CHANGE OF BASIS

Describing an ecosystem is a problem of dimensionality reduction. Given a community of billions of individuals, or millions of strains, the challenge is to find a reduced set of degrees of freedom that could allow characterizing its properties or dynamics with an acceptable degree of accuracy.

The standard approach to this problem is through clustering. Most commonly, the microscopic degrees of freedom, such as abundance of individual strains, are assumed to be located at the leaves of a tree (based on phenotypic similarity, phylogenetics, etc.). A coarse-grained description is then constructed by choosing a transect cutting across this tree, and combining branches lying downstream of the transect (Fig. 2A). For concreteness,

4

consider an example of combining just two leaves. Imagine a community where a species
of rabbits comes in two varieties that differ only slightly: say, a "white" rabbit W and a
"brown" rabbit B (Fig. 2B). When describing their predator-prey dynamics with foxes, a
two- or three-variable description could be used, depending on the desired level of detail.
The comprehensive three-variable description might capture come subtle features, but for
many questions, the classic Lotka-Volterra model, whereby the two varieties are clustered
together into a single variable "rabbits" R, would provide a sensible approximation:

$$\{W, B\} \mapsto R. \tag{1}$$

This work proposes a different perspective on the clustering operation (1). To a first
approximation, most environmental perturbations will affect the population of rabbits as a
whole; e.g. an increase in predator efficiency, or habitat deterioration will depress the pop-
ulation of both varieties together. Upon careful examination, however, the small functional
differences will manifest themselves on a slow timescale: for example, if the predator is
better attuned to spotting a particular coat colour, one variety may eventually outcompete
the other. In this way, compositional changes naturally align not with the Lotka-Volterra
axes W and B, but with the rotated axes that provide an hierarchically-informed descrip-
tion: one variable $R = W + B$ ("rabbits") that changes on a fast timescale, and another
$C = W - B$ ("coat color") that changes on a slow timescale. This work proposes to reinter-
pret Fig. 2B not as clustering, but as a change of basis, followed by the decision to ignore a
slow-timescale variable. As we will see, an appropriately rotated basis can be constructed for
an arbitrary population of interacting individuals, and can serve as a naturally hierarchical
characterization of a community. Borrowing the term introduced in [40], one could call this
the "ecomode description".

II. THE EIGENMODE DESCRIPTION:
THE GENERAL DEFINITION

Let $\mathcal{C}$ denote a full individual-based description of a community: a list of all present
organisms (labelled by indices μ, ν...) and their functional characteristics. A general
model for ecological dynamics can be written as a rule that associates to each individual
μ a "rate of abundance change" r_μ: the probability per unit time with which it will either

5

generate another individual ($r_\mu > 0$) or die ($r_\mu < 0$). In a most general model, this rate can depend on any detail of the community state, as well as on external environmental parameters denoted by $\mathcal{E}$:

$$r_\mu = r_\mu(\mathcal{E}, \mathcal{C})$$

Within any such model, we can define a matrix $\mathcal{M}$ of interactions between individuals: $\mathcal{M}_{\mu\nu}$ is the effect that a hypothetical duplication of individual μ would have on the instantaneous division/death rate of ν.

$$\mathcal{M}_{\mu\nu} = r_\nu(\mathcal{E}, \{\mathcal{C} \text{ with } \mu \text{ duplicated}\}) - r_\nu(\mathcal{E}, \mathcal{C}) \tag{2}$$

Importantly, here μ and ν are *individuals*, not species, and so the definition of $\mathcal{M}$ does not require the partitioned community assumption. For the classic Lotka-Volterra scenario, the "functional characteristics" of an individual need only specify whether it is a fox or a rabbit, but we can allow this microscopic description to be arbitrarily detailed, down to the point of all individuals being unique. Note that the interaction matrix $\mathcal{M}$ is defined instantaneously; the number of individuals and thus the size of $\mathcal{M}$ can change in time.

The primary focus of this work is the eigensystem of the interaction matrix $\mathcal{M}$. For a population of $\mathcal{N}$ individuals, this $\mathcal{N} \times \mathcal{N}$ matrix will always have $\mathcal{N}$ eigenvalues. Note, however, that if individuals μ and μ' happen to be functionally identical, then $\mathcal{M}_{\mu\nu} = \mathcal{M}_{\mu'\nu}$ for all ν, and $\mathcal{M}$ has a zero eigenvalue. For a coarse model ascribing all individuals to only $K \ll \mathcal{N}$ distinct phenotypes, the matrix $\mathcal{M}$ will be highly degenerate with only K non-trivial eigenvalues; the remaining $\mathcal{N} - K$ will be strictly zero. A detailed functional description recognizing all individuals as unique will remove the degeneracy; however, if the partitioning into K categories were indeed a good approximation, $\mathcal{N} - K$ eigenvalues will remain small (here and everywhere, characterizing an eigenvalue as "small" refers to its absolute value). We are led to hypothesize that the spectrum of $\mathcal{M}$ may provide a naturally hierarchical description of ecological relationships in a community, with progressively smaller eigenvalues resolving finer and finer details. The grouping of individuals into discrete types, when it exists, is established as a result of this analysis; in this framework, it no longer constitutes a fundamental assumption.
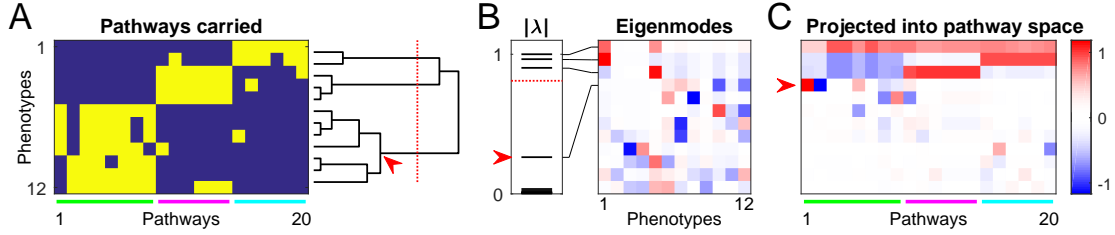
FIG. 3. For a well-clustered set of phenotypes, the eigenmode representation captures the hierarchy of phenotype similarity. **A:** Pathway presence matrix for 12 phenotypes, generated as variations on 3 "core phenotypes" (colored lines); phenotypes carrying a given pathway (shown in yellow) are able to metabolize the respective resource. The tree represents hierarchical clustering based on Hamming distance. Dotted transect corresponds to the clear grouping into 3 clusters. Under MacArthur's model, resource competition in this ecosystem leads to a unique stable equilibrium. **B:** The spectral structure of the ecosystem shown in A, in the vicinity of its equilibrium point. Left: the eigenvalues of the Jacobian. All eigenvalues are negative, corresponding to a stable equilibrium; shown is the absolute value. Right: the eigenvectors, ordered by decreasing $|\lambda|$; each is a linear combination of 12 phenotypes and is difficult to interpret directly. **C:** The same eigenvectors, projected into pathway expression space. It is now clear that the three dominant modes span the three "core phenotypes" (colored lines as in panel A). The fourth mode (arrowhead) corresponds to a subdivision of the first cluster (compare with panel A).

III. RESTRICTING GENERALITY: EIGENMODES IN LOTKA-VOLTERRA MODELS

In the interest of full generality, the definition of the previous section was explicitly individual-based. The existence of such a definition is a very important feature of the eigenmode framework; however, the inconvenience of dealing with discrete individuals is a technical nuisance that could obscure the larger points in the subsequent discussion. To build intuition about the eigenmode perspective, the best approach is to illustrate it on simple examples.

The most familiar setting is that of a generalized Lotka-Volterra model, with K continuous degrees of freedom. Its linearized dynamics in the vicinity of an equilibrium point take the

7

form:

$$\frac{d\,\delta n_a}{dt} = \sum_{b=1}^{K} M_{ab}\,\delta n_b. \tag{3}$$

Here δn denotes the deviation from equilibrium abundance, and the index a runs from 1 to K. For simplicity, this is the context that will be used for most of the discussion. Positing a finite number of K interacting phenotypes may seem to contradict the stated goal of avoiding the "species" partitioning. However, the reader is invited to think of these phenotypes not as a "species" label, but as a microscopic, sub-species description; e.g. interacting bacterial strains that may or may not group into species. In this context, the variable-size individual interaction matrix $\mathcal{M}$ has only K non-trivial eigenvalues and eigenvectors, which correspond to the eigensystem of the Lotka-Volterra interaction matrix M (see the Supplementary Material; SM). The eigenvalues have units of inverse time and correspond to the relaxation time scales of perturbations applied along each eigenmode.

In order to make the eigenmodes interpretable, rather than positing the interaction matrix M in some arbitrary manner, it is extremely helpful to construct it from a functional description of what these phenotypes do (recall the example above, where the $W + B$ axis could be conveniently interpreted as "rabbits" and $W - B$ as "coat color"). One simple way to do this is provided by the MacArthur model of resource competition [42–44], where phenotypes are characterized by the resources they are able to consume. This defines a particular instance of a competitive Lotka-Volterra model, while also giving us an intuition for phenotypic similarity, and a functional interpretation for the eigenmodes we will observe.

Specifically, following ref. [45], consider a well-mixed microbial community in a medium where a single limiting element (e.g. carbon) exists in N forms ("resources" $i \in \{1 \dots N\}$; in what follows $N = 20$). Each resource can be metabolized with a dedicated "pathway". The phenotypes are defined by two pieces of information: first, their requirement for the limiting element, and second, the pathways they carry. The former will be randomly drawn for simplicity, but the latter will be chosen strategically, by hand, in order to construct the most instructive examples. The set of K phenotypes defines a K-by-N binary matrix σ of pathway presence/absence ($\sigma_{ai} = 1$ if phenotype a carries the pathway enabling it to consume resource i), and this matrix σ determines the phenotype-phenotype interactions. Specifically, the competition between a pair of phenotypes (a, b) is the stronger, the larger the overlap in their resource preference $\sum_i \sigma_{ai}\sigma_{bi}$. Under MacArthur's model, such competition

8

always results in a unique and stable equilibirum [43, 45]. The mathematical details are provided in the SM, but for the discussion that follows, this description should already be sufficient.

IV. THE EFFECTIVE NUMBER OF SPECIES IS DEFINED BY A SPECTRAL GAP

For the first example, consider a community composed of phenotypes depicted in Fig. 3A, competing for 20 resources supplied in equal abundance. These 12 phenotypes were purposefully generated to form three clear clusters (see SM for the exact procedure; MATLAB scripts (Mathworks, Inc.) reproducing all examples and figures are available as Supplementary file 1). This community can be interpreted as harbouring three species; within each species, phenotypes share a core set of pathways (indicated) and differ only in a small subset, similar to the "core" and "accessory genome" of a species [46]. How is this scenario seen from the eigenmode perspective?

Resource competition between these phenotypes defines a 12-by-12 interaction matrix. Its eigenmode spectrum is presented in Fig. 3B. Each eigenvector is a linear combination of phenotypes, and is difficult to interpret. However, since each phenotype is defined by its resource consumption, their linear combination can be conveniently characterized in the same way: the pathway presence/absence matrix σ can be used to project each eigenvector into the pathway expression space (Fig. 3C). This provides a functional interpretation for the eigenmodes: each row in the matrix shown in Fig. 3C indicates how the community-level pathway expression changes if its composition is perturbed in the direction of a given eigenmode.

We observe that the spectrum is dominated by 3 modes, whose projections in the pathway expression space are linear combinations of the "core genomes" of the three phenotype clusters (Fig. 3C, top 3 rows). This indicates that at the short timescale, the response of the system to a perturbation can be well approximated using only 3 degrees of freedom. Above, characterizing this habitat as "harbouring 3 species" was an interpretation that relied on a subjective judgement of Fig. 3A by a human observer. The eigenmode spectrum contains this information, and more, in an objective format, decomposing community dynamics over a range of time scales.
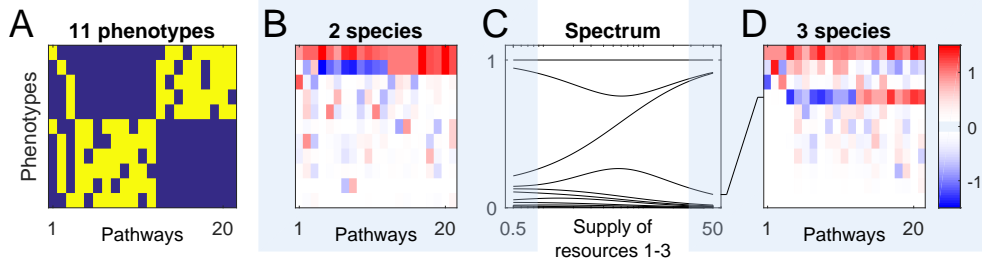
9

FIG. 4. A smooth interpolation between two unrelated clustering patterns into, respectively, two and three effective species. **A:** Pathway presence matrix (same format as in Fig. 3A) for 11 phenotypes. Their preference for resources 1–3 induces a grouping into three clusters, but the remaining pathways induce a conflicting clustering pattern (with only two groups). **B:** If the supply of resource 1–3 is low, community dynamics exhibits two dominant modes, reflecting the 2-cluster structure. As the supply of these resources is increased (panel **C**), the preference for resources 1–3 becomes the more relevant clustering criterion for describing the dynamics. **D:** The same 11 phenotypes now behave as a community of 3 effective species. Unlike a speciation event, where the uncertainty in defining the number of species is merely a matter of the desired level of detail (*cf.* Fig. 3), the intermediate regime of panel C interpolates between unrelated clustering patterns (not nested).

The first eigenvector is a collective mode, with each entry close to 1. This corresponds to the statement that, if we insisted on reducing the community to a single degree of freedom, our best choice would be to say that all resources are being consumed at a certain overall rate. However, this is a poor approximation; capturing the dominant fast dynamics requires 3 degrees of freedom. Far below the dominating eigenvalues, a fourth one becomes apparent (arrowhead). Examining the fourth row in Fig. 3C and comparing it to panel A reveals that this mode corresponds to a finer structure within the first cluster, where coexisting variants differ in their ability to consume resources 1 and 2. The eigenmode decomposition provides a way to describe dynamics with increasing level of detail; smaller differences manifest themselves as smaller eigenvalues. Thus, the eigenmode description of community structure is naturally hierarchical.

The key property of the eigenmode spectrum of Fig. 3B is the existence of a noticeable *gap* in its spectrum, which provides a justifiable cutoff point for the number of degrees of freedom to include into a low-dimensional description (note that in some contexts, the term

10

"spectral gap" is used in a narrower sense to describe the difference between specifically the first and second eigenvalues; this is not the meaning adopted here). Whenever an ecosystem exhibits a spectral gap, one could define the effective number of species as the number of dominant eigenvalues. Importantly, however, the spectral gap need not be unique, and is a fluid property: under a continuous change of parameters, gaps can move, appear and disappear. The next example explores the implications of this fact.

V. MOVING THE SPECTRAL GAP: "TWO AND A HALF SPECIES"?

Consider now a community of 11 phenotypes of Fig. 4A. For equiabundant resources, we observe the expected "2-species" spectrum (Fig. 4B), the dominant eigenmodes reflecting the two "core genomes" that correspond to the dominating binary clustering pattern.

However, by construction (see SM for details), each of these phenotypes specializes in only one of the first three resources. If the supply of these resources is strongly increased, this preference will become the more relevant clustering criterion; the same 11 phenotypes now behave as a community of 3 effective species (Fig. 4D). Note that in this regime, the eigenmode corresponding to the binary clustering (which used to be mode #2) is again part of the spectrum (now as mode #4), but is characterized by a much smaller eigenvalue, signaling much slower dynamics than those governed by the dominant three resources.

Both these extremes were easy to interpret, but how should we think of the regime lying in between (Fig. 4C)? A smooth process leading to a change in the number of species may seem familiar from the study of speciation. One cluster of phenotypes can be smoothly separated into two, and if we insisted on counting species, then somewhere in between there must be a gray area of "more than one, but less than two". This inability to draw a hard threshold of exactly when two close varieties should be considered distinct species is not, however, a significant challenge to the species paradigm. Whether a finer subdivision is warranted is merely a question of the desired level of detail. This was already illustrated in the previous example: the spectrum of Fig. 3B exhibits two significant gaps, leaving us the choice of characterizing it with either three or four degrees of freedom. And just as in Fig. 4, subjecting the community of Fig. 3A to an increased abundance of resources 1 and 2 would reduce the first spectral gap, making the finer (4-species) description increasingly

more justifiable. In this scenario, the origin of our inability to define an exact number of species is well understood.

However, Fig. 4 shows that the problem is more general. The somewhat artificial example of Fig. 4A was purposefully constructed to show that a smooth change of parameters can interpolate between two clustering patterns even when they are entirely unrelated (not nested): the same 11 phenotypes are grouped as {1–5}, {6–11} in one case (panel B), and {1, 6}, {2, 7–10}, {3–5, 11} in the other (panel D). The "gray area" in between is of a different nature than the uncertainty about whether to further subdivide a particular cluster. This suggests that the intermediate regime of Fig. 4C should be interpreted not as "two and a half species", but rather as a situation where a gap-less spectrum means that the effective number of species cannot be defined at all.

Characterizing such a "regime without species" is indeed our ultimate destination. Unfortunately, this particular example suffers from an obvious problem: it was constructed within a Lotka-Volterra framework with 11 degrees of freedom. Whether or not they can be grouped into clusters, such a model can always be seen as describing an interacting system of 11 species.

This issue is a consequence of the decision, made early on, to work within the familiar Lotka-Volterra context. No simple Lotka-Volterra model can generate a truly gap-less spectrum: as discussed above, a model grouping individuals into K discrete categories always has a gap separating the K non-trivial eigenvalues from the vast number of those that are strictly zero, as a consequence of the partitioning assumption. As a direct consequence of the partitioned community assumption, *virtually all classical ecological models describe systems whose eigenvalue spectrum exhibits a prominent gap.* What reasons do we have to believe that natural ecosystems, e.g. the microbial communities inhabiting our intestines, possess this property? Absent such evidence, we must acknowledge that spectral gaps, whose existence is a prerequisite for a low-dimensional "species-based" description to be adequate, could be an artifact of familiar theoretical models, and not necessarily the best approximation of ecological reality. If one could build an ecosystem model whose spectrum really did exhibit no gaps, such a scenario fundamentally could not be conceptualized as a system of interacting species. One signature of this regime is that any attempt to approximate it with a K-dimensional Lotka-Volterra model will appear to contain exactly K species, for any K. The next section will construct one concrete example exhibiting this behavior.
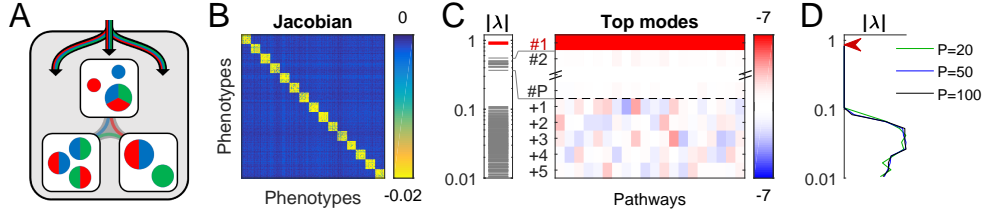
FIG. 5. Constructing a model case of an ecology without species. **A:** A multi-patch extension of MacArthur's resource competition model. Externally supplied resources are split between P patches harboring separate communities (here, $P = 3$). Patch boundaries are assumed to be (weakly) permeable to resources, but not organisms (see SM for details). In this cartoon, phenotypes are represented by circles; size signifies abundance, while colors correspond to resources consumed. **B:** The phenotype-phenotype interaction matrix for a particular realization of the model with $N = 20$ resources and $P = 15$ patches; the 298 coexisting phenotypes are ordered by patch. Organisms in the same patch interact more strongly than across patches, leading to a block-diagonal structure. **C:** The top P eigenmodes reflect the P-patch structure of the habitat. Modes 2 through P have vanishingly small projection into the pathway expression space: they represent inter-patch dynamics and are hidden from an external observer who cannot resolve individual patches. **D:** The shape of the eigenvalue distribution is independent of P (left); modes "hidden" from external observer (2 through P) are omitted. The unique dominant mode is the collective mode (arrowhead). The remainder of the spectrum follows a dense distribution with no gap, reflecting the unstructured background of interacting phenotypes.

VI. AN ECOLOGY WITHOUT SPECIES

Constructing an ecosystem with a dense spectrum requires a model where the number of distinct phenotypes can be arbitrarily large. It would be convenient to also maintain two useful features of the previous examples: a functional characterization of phenotypes, and the existence of a stable equilibrium. Finally, to underscore the interpretation of this scenario as a regime "without species", we will set an additional objective, requiring the set of coexisting phenotypes to be unstructured, so that no grouping is privileged. Ideally, then, we would like to have a model where an arbitrarily large, unstructured set of functionally defined phenotypes can co-exist at a stable equilibrium.

In the classic (well-mixed) framework of MacArthur, N resources can stably sustain at

most N phenotypes [44]. However, imagine a spatially structured habitat with P patches, whose boundaries are permeable to resources, but not to organisms (Fig. 5A). The mathematical details of this model are provided in the supplement; however, qualitatively, the behavior is intuitive. At infinite permeability (resources are exchanged freely), the P patches are effectively a single patch, and we are back to the classic MacArthur model. If permeability is zero, the P isolated patches can sustain up to PN phenotypes, but these have no interactions across patch boundaries. At intermediate permeability, we find the desired situation where the interactions are nontrivial, and yet the total number of coexisting phenotypes can be arbitrarily large. Although the number of distinct binary vectors is bounded by 2^N, two phenotypes located in different patches count as distinct even if their pathway content is identical, because they interact differently with other phenotypes (for example, of the two "red-blue" phenotypes in the cartoon in Fig. 5A, one has a strong competitor in the same patch, but the other does not).

For each patch, define an independent pool of local phenotypes that is fully random (each phenotype carries or not a given pathway with probability $1/2$). After the system equilibrates, the total number of coexisting phenotypes is of order PN. Fig. 5B shows the complete interaction matrix for 298 phenotypes in $P = 15$ patches (see SM for details). Predictably, phenotypes group by patch: at weak resource permeability, organisms interact more strongly within a patch than they do across patches. However, the pathway content of these phenotypes is, by construction, completely unstructured.

These observations are reflected in the eigenmode structure. P eigenvalues clearly stand out in the spectrum (Fig. 5C), reflecting the patch structure: at the very top, we recognize the collective mode; the remaining $P-1$ modes are pairwise differences between the collective modes of the individual patches. These dominant modes are followed by a dense spectrum of eigenvalues, whose distribution exhibits no gap (Fig. 5D): aside from the patch structure, no grouping of phenotypes is in any way privileged, offering no possibility for a coarse-grained description.

In order to take the limit $P \to \infty$, consider now the projection of the eigenvectors into the N-dimensional expression space. This corresponds to adopting the perspective of an observer who can measure the community-wide expression of metabolic pathways, but cannot resolve individual patches. Such an observer can see the dominant collective mode, but modes 2 through P are effectively hidden: their projection into the pathway expression space has

14

vanishingly small components (panel C). To understand this, consider a perturbation where all phenotypes in patch 1 experience an increase in abundance, and all phenotypes in patch 2, a decrease. Such a perturbation will relax quickly, corresponding to one of the dominant (strongly negative) eigenvalues. However, the internal dynamics represented by this mode is effectively hidden from an outside observer whose measurements do not resolve individual patches. Such an observer can only see a unique collective mode over a fully unstructured background (modes from $P+1$ onwards): a dense spectrum, whose shape is independent of P.

This characterization allows sending P to infinity. This limit is proposed as a tractable model case of an "ecology without species". In this regime, the number of interacting phenotypes becomes infinite, their pathway content is unstructured, and the spectrum is dense; yet this construct continues to describe a valid ecology at equilibrium whose properties could be investigated. For instance, if the supply of some resource is increased, this system will have some well-defined response, which our hypothetical observer could measure and could attempt to predict.

What is the utility of this regime? The extreme, fully unstructured case is certainly a poor approximation of any natural setting. However, it provides a baseline that can be perturbed to describe novel scenarios of "weakly structured" ecologies, as illustrated in the next and final example.

VII.   DESCRIBING WEAKLY STRUCTURED ECOLOGIES

Fig. 6 repeats the analysis of Fig. 5, with one exception: the first and last pathway are forbidden from ever appearing within the same organism. Specifically, for each patch, the procedure generating the local phenotype pool was modified as follows: each entry in the pathway presence matrix σ was independently set to 1 with probability $1/2$ as before, but whenever this resulted in a phenotype carrying the forbidden pathway pair, this phenotype was removed from the pool. This kind of structure is not implausible: a detrimental cross-talk between certain pathways can make their simultaneous activity in the same organism difficult; for example, the enzyme nitrogenase is inactivated by oxygen and thus must be kept separate from oxygen-utilizing pathways.

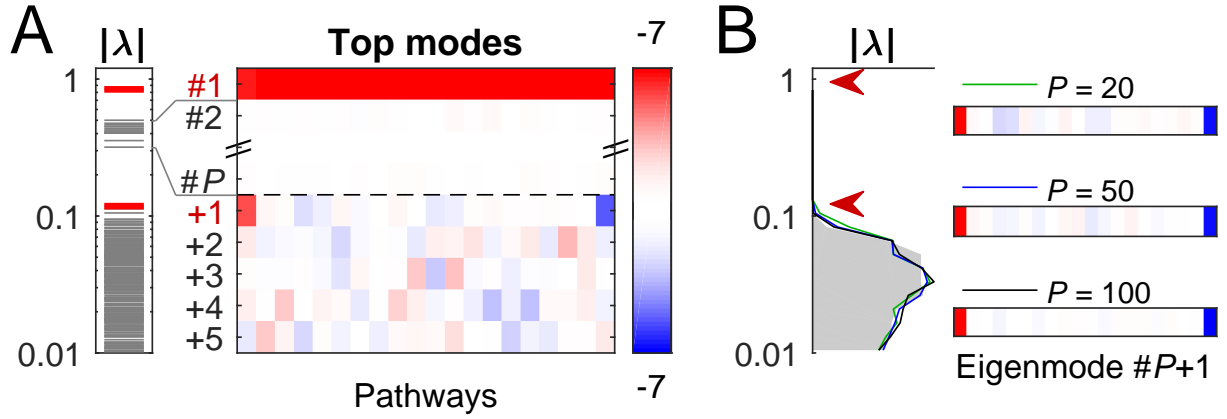The spectrum of this ecosystem, projected into the pathway expression space, is shown

15

FIG. 6. A weakly structured ecology as a perturbation of the no-species regime. **A:** When the phenotype pool is modified to introduce an anticorrelation between first and last pathway, the spectrum acquires a mode reflecting this structure (highlighted). This new mode is the dominant observable mode (after the collective mode). **C:** The leading observable eigenmodes remain at the same position in the spectrum (arrrowheads) and retain their structure (right). The rest of the spectrum follows the same distribution as in the unstructured regime (shaded; compare with Fig. 5D).

in Fig. 6A. Comparing to Fig. 5C, we see that there are now two modes that stand out from the gap-less background. In addition to mode #1 (the collective mode), there is now also mode #$(P+1)$ that represents the only perturbation imposed on an otherwise unstructured phenotype background. The spectrum in Fig. 6A was computed for one particular random realization of a system with $P = 15$ patches; however, as P is increased, the projection of this mode into pathway space and its position in the spectrum both remain invariant (Fig. 6B). This example demonstrates that the fully unstructured regime constructed here can indeed be meaningfully perturbed to describe the dynamics of "weakly structured" ecologies, an interesting theoretical scenario that was previously out of reach.

To maintain consistency with the previous examples, the approach taken here focused exclusively on resource competition. The "hidden" modes corresponding to unobservable inter-patch dynamics is the price to pay for this simplicity. Other approaches are possible; one might, for instance, imagine constructing a dense-spectrum model by means of a fully individual-based description, or through a sequence of iterative subdivisions of a Lotka-

16

Volterra model (similar to the approach taken in Ref. [48], but repeated an infinite number of times). One of these methods might yield a more elegant construction; however, the multi-patch example constructed here has the advantage of being easily interpretable, and potentially even analytically tractable with methods developed in Ref. [49].

VIII. DISCUSSION

The mathematics of the "eigenmode approach" described here is very simple. Anyone familiar with the Principle Component Analysis (PCA) has encountered the problem of choosing the number of components on which to focus. In rare circumstances, a few components are sufficient to capture most of the variance; the generic situation, however, is that the spectrum shows no obvious gap, offering no natural truncation point [50]. Normally (e.g., when investigating stability [1, 51]), one thinks of the PCA modes as linear combinations of the fundamental degrees of freedom, namely the species abundances. Here, the novelty is to ask whether the eigenmodes could be used as the fundamental degrees of freedom themselves.

One advantage of this approach is that degrees of freedom are inferred, rather than postulated. The eigenmode formalism admits the possibility that the most relevant variables may take a different form than simply the combined abundance of some taxa. As a simplest example, imagine a two-species consortium S_1, S_2 that runs the same metabolic reaction that species S_3 performs alone. In this scenario, the combination $S_1 + S_2 - S_3$ might be a functionally "negligible" (slow) eigenmode, but the reduction of dimensionality this observation affords does note reduce to simple branch-merging on a phylogenetic tree. Situations when one diverse set of species is essentially interchangeable with another diverse set are likely the rule, rather than an exception, as evidenced by the dramatic compositional variations of human-associated microbiota across healthy individuals, at all taxonomic levels [16]. This suggests that the number of negligible eigenmodes may indeed be large, and recognizing them could simplify the task of modelling these communities.

Whether the eigenmode idea could indeed provide a practical recipe for constructing low-dimensional descriptions of real ecosystems is a question that can only be settled by an empirical study. Measuring the species-species interaction matrix is difficult, but has been attempted [51–53], and the empirically determined ecomodes were previously shown to be

interpretable for low-diversity communities [40]. An alternative definition of an ecomode, easier to access experimentally, might rely on the spectral decomposition not of the interaction matrix, but of the correlation matrix, computed across an ensemble of communities [41]. The two definitions are not equivalent [41], and further work is required to investigate which approach would be more successful in practice.

Perhaps more profoundly, however, the framework developed here motivates new questions that could be asked about ecosystems. Are there general laws governing the structure and evolution of community spectra? What ecological mechanisms might cause a given eigenvector to change in a predictable way, or a cause a given eigenvalue to become more prominent? How is the spectrum affected by evolutionary mechanisms? One may, for example, expect that purifying selection encourages the development of spectral gaps, while diversity-creating mutations have the opposite effect. A major advance in classical physics came with the development of the renormalization group, a framework explaining why many physical systems, as they become large, do in fact develop a gap in the spectrum, enabling some low-dimensional "effective theory" to become an excellent description [54]. Ecology and evolution provide a new context for asking similar questions, which are particularly relevant today with the study of highly diverse microbial ecosystems of medical and environmental interest [16, 17].

In contrast to questions like *"how many species are present?"*, or *"what happens if species X is introduced?"*, those formulated above do not require postulating the existence of species as fundamental ecological variables. To emphasize this, here it was shown that even an extreme scenario of an "ecology without species" can now be conceptualized. The classic Lotka-Volterra model investigates an ecology of perfectly defined species, a regime that has been studied extensively (Fig. 1a). This work presented a theoretical construct that can be seen as the polar opposite, one where the phenotypes are completely unstructured, and yet form a valid ecology (Fig. 1c).

In any real setting, organisms sharing a recent common ancestor are functionally similar, and the approximation "every single organism is unlike any other" (Fig. 1c) is even more extreme than the converse assumption of perfect clustering into a small number of species (Fig. 1a). Nevertheless, a theoretical investigation of this unstructured regime will likely be a valuable step towards understanding the middle ground occupied by the real-life communities; in physical terms, its role would be to provide a different origin for a perturbative

expansion (Fig. 1). For example, in condensed matter physics it was recently proposed that a granular medium at the brink of a jamming transition can be thought of as a "maximally disordered" solid, the opposite extreme of a perfectly ordered crystal [55]. Importantly, it was shown that certain properties of real-life solids, even at relatively low disorder, can be more adequately described as a perturbation of this "maximally disordered" regime, rather than as a perturbation of a perfect crystal [55]. Similarly, it is intriguing to hypothesize that some behaviors of highly diverse natural ecosystems may be better described as a perturbation of a fully unstructured no-species regime (adding some tendency of phenotypes to cluster), rather than as a small deviation from the classic picture of well-clustered species (where we attempt to recognize that species boundaries are "fuzzy"). This approach would provide a systematic framework to investigate the implications of "structured" and "unstructured" variation imposed at a range of scales, effects whose importance at the intraspecific level was recently highlighted [56].

The eigenmode formalism is by no means limited to resource competition models; MacArthur's model merely provided a convenient setting for constructing easily interpretable examples. The framework described here does, however, suffer from a number of other limitations: for example, only deterministic dynamics were considered, and the discussion was restricted to small fluctuations around some ecological state. Nevertheless, the question raised here was whether our description of ecosystems could move beyond the inherently discrete concept of a "species". In light of this challenge, the simplifications above provide a reasonable starting point.

---

[1] *Theoretical Ecology: Principles and Applications* eds. May R, McLean A (Oxford University Press, 2007), 272pp.

[2] Ereshefsky M (2010) Darwin's solution to the species problem. Synthese 175:405–425.

[3] Hey J (2006) On the failure of modern species concepts. *Trends Ecol Evol* 21:447–450.

[4] DeAngelis DL, Mooij WM (2005) Individual-based modeling of ecological and evolutionary processes. *Annu Rev Ecol Evol Syst* 36: 147–168.

[5] Bolnick DI, Amarasekare P, Araujo MS, Burger R, Levine JM, Novak M, Rudolf VH, Schreiber SJ, Urban MC and Vasseur DA (2011) Why intraspecific trait variation matters in community ecology. *Trends ecol evol* 26(4): 183–192.

[6] Shapiro BJ, Polz MF (2014) Ordering microbial diversity into ecologically and genetically cohesive units. *Trends microbiol*, 22(5): 235–247.

[7] De Queiroz K (2007) Species concepts and species delimitation. *Syst Biol* 56: 879–886.

[8] Achtman M, Wagner M (2008) Microbial diversity and the genetic nature of microbial species. *Nature Rev Microbiol* 6: 431–440.

[9] Hart MW (2011) The species concept as an emergent property of population biology. *Evolution* 65(3): 613-616.

[10] Cordero OX, Polz MF (2014) Explaining microbial genomic diversity in light of evolutionary ecology. *Nature Rev Microbiol* 12: 263–273.

[11] Konopka A (2009) What is microbial community ecology? *ISME J* 3: 1223–1230.

[12] Gill SR et al. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312(5778): 1355–1359.

[13] Turnbaugh PJ et al. (2009) A core gut microbiome in obese and lean twins. *Nature* 457(7228): 480–484.

[14] Caporaso JG et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* 108 Suppl 1: 4516–4522.

[15] Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R (2012) Diversity, stability and resilience of the human gut microbiota. *Nature* 489: 220–230.

[16] Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486(7402): 207–214.

[17] Gilbert JA, Jansson JK, Knight R (2014) The Earth Microbiome project: successes and aspirations. *BMC Biology* 12: 69.

[18] Staley JT (2006) The bacterial species dilemma and the genomicphylogenetic species concept. *Philos Trans R Soc Lond B Biol Sci* 361(1475): 1899–1909.

[19] Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP (2009) The bacterial species challenge: making sense of genetic and ecological diversity. *Science* 323(5915): 741–746.

[20] Zhaxybayeva O, Doolittle WF (2011) Lateral gene transfer. *Curr Biol* 21(7): R242–R246.

[21] Lyons NA, Kolter R (2015) On the evolution of bacterial multicellularity. *Curr Opin Microbiol* 24: 21–28.

[22] Tringe SG et al. (2005) Comparative metagenomics of microbial communities. *Science* 308(5721): 554–557.

[23] Vieites JM, Guazzaroni ME, Beloqui A, Golyshin PN, Ferrer M (2009) Metagenomics approaches in systems microbiology. *FEMS Microbiol Rev* 33(1): 236–255.

[24] Sites JW Jr, Marshall JC (2004) Operational criteria for delimiting species *Annu Rev Ecol Evol Syst* 35: 199–227.

[25] Cohan FM (2006) Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Philos Trans R Soc B* 361(1475): 1985–96.

[26] Cohan FM, Perry EB (2007) A systematics for discovering the fundamental units of bacterial diversity. *Curr Biol* 17(10): R373–386.

[27] McGill BJ, Enquist BJ, Weiher E, Westoby M (2006) Rebuilding community ecology from functional traits. *Trends ecol evol* 21(4): 178–185.

[28] Litchman E, Klausmeier CA (2008) Trait-based community ecology of phytoplankton. *Annu Rev Ecol Evol Syst* 39: 615–639.

[29] Ereshefsky M (2001) *The poverty of Linnaean hierarchy* (Cambridge University Press, Cambridge).

[30] Okasha S (2011) Biological ontogeny and hierarchical organization: a defense of rank freedom. *The major transitions in evolution revisited*, eds. Calcott B, Sterelny K (MIT Press, Cambridge, MA), pp53–63.

[31] Hendry AP, Vamosi SM, Latham SJ, Heilbuth JC, Day T (2000) Questioning species realities. *Conservation Genetics* 2000, 1(1): 67–76.

[32] Doolittle WF (2012) Population genomics: how bacterial species form and why they don't exist. *Curr Biol* 22(11): R451–3

[33] Acinas SG et al. (2004) Fine-scale phylogenetic architecture of a complex bacterial community *Nature* 430: 551-554.

[34] Hunt DA, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF (2008) Resource partitioning

and sympatric differentiation among closely related bacterioplankton. *Science* 320: 1081–1085.

[35] Shapiro DJ et al. (2012) Population genomics of early events in the ecological differentiation of bacteria. *Science*, 336: 48–51.

[36] Kashtan N et. al. (2014) Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* 344(6182): 416–420.

[37] Biller SJ, Berube PM, Lindell D, Chisholm SW (2015) *Prochlorococcus:* the structure and function of collective diversity. *Nature Rev Microbiol* 13: 13-27.

[38] Prosser JI, et al. (2007) Essay — The role of ecological theory in microbial ecology. *Nature Rev Microbiol* 5(5): 384–392.

[39] Schmidt TSB, Rodrigues JFM, von Mering C (2015) Limits to robustness and reproducibility in the demarcation of operational taxonomic units. *Environ Microbiol* 17(5): 1689–1706.

[40] Hekstra DR, Leibler S (2012) Contingency and statistical laws in replicate microbial closed ecosystems. *Cell* 149(5): 11640–73.

[41] Frentz Z, Kuehn S, Leibler S (2015) Strongly deterministic population dynamics in closed microbial communities *Phys. Rev. X* 5, 041014

[42] MacArthur R, Levins R (1967) The limiting similarity, convergence, and divergence of coexisting species. *American Naturalist* 101(921): 377–385.

[43] MacArthur R (1969) Species packing, and what interspecies competition minimizes. *Proc Natl Acad Sci U S A* 64(4): 1369–71.

[44] MacArthur R (1970) Species packing and competitive equilibrium for many species. *Theor popul biol* 1(1):1–11.

[45] Tikhonov M (2016) Community-level cohesion without cooperation. *eLife* doi:10.7554/eLife.15747

[46] Medini D et al. (2008) Microbiology in the post-genomic era. *Nature Rev Microbiol* 6: 419–430.

[47] Schleuter D, Daufresne M, Massol F, Argillier C (2010) A user's guide to functional diversity indices. *Ecological monographs* 80 (3), 469–484.

[48] Shtilerman E, Kessler DA, Shnerb NM (2015) Emergence of structured communities through evolutionary dynamics. *J Theor Biol* 383: 138–144.

[49] Tikhonov M, Monasson R (2016) A collective phase in resource competition in a highly diverse ecosystem. arXiv:1609.01270.

[50] Bradde S, Bialek W (2016) PCA meets RG. arXiv:1610.09733.

22

[51] Ives AR, Dennis B, Cottingham KL, Carpenter SR (2003) Estimating community stability and ecological interactions from time-series data. *Ecol. Monographs* 73(2): 301–330.

[52] Stein RR, Bucci V, Toussaint NC, Buffie CG, Rätsch G, Pamer EG, Sander C, Xavier JB (2013) Ecological modeling from time-series inference: Insight into dynamics and stability of intestinal microbiota. *PLoS Comput Biol* 9(12): e1003388.

[53] Fisher CK and Mehta P (2014) Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PloS one*, 9(7): e102451.

[54] Machta BB, Chachra R, Transtrum MK, Sethna JP (2013) Parameter space compression underlies emergent theories and predictive models. *Science*, 342(6158): 604–607.

[55] Goodrich CP, Liu AJ, Nagel SR (2014) Solids between the mechanical extremes of order and disorder. *Nature Physics*, 10: 578-581.

[56] Moran EV, Hartig F, Bell DM (2016) Intraspecific trait variation across scales: implications for understanding global change responses. *Glob Change Biol*, 22: 137-150.