nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-021-01462-8

Coexistence holes characterize the assembly and disassembly of multispecies systems

Marco Tulio Angulo[1]✉, Aaron Kelley[2], Luis Montejano[2], Chuliang Song[3,4,5,6] and Serguei Saavedra[3,6]✉

A central goal of ecological research has been to understand the limits on the maximum number of species that can coexist under given constraints. However, we know little about the assembly and disassembly processes under which a community can reach such a maximum number, or whether this number is in fact attainable in practice. This limitation is partly due to the challenge of performing experimental work and partly due to the lack of a formalism under which one can systematically study such processes. Here, we introduce a formalism based on algebraic topology and homology theory to study the space of species coexistence formed by a given pool of species. We show that this space is characterized by ubiquitous discontinuities that we call coexistence holes (that is, empty spaces surrounded by filled space). Using theoretical and experimental systems, we provide direct evidence showing that these coexistence holes do not occur arbitrarily—their diversity is constrained by the internal structure of species interactions and their frequency can be explained by the external factors acting on these systems. Our work suggests that the assembly and disassembly of ecological systems is a discontinuous process that tends to obey regularities.

Ecological systems are the product of various assembly and disassembly processes (for example, invasions and habitat fragmentation, respectively) that, together with the underlying ecological dynamics, determine which and how many species we observe in a given place and time[1–4]. Classic and more recent studies have focused on characterizing the limits on the maximum number of species that can coexist under given constraints[5–9]. Yet, we know little about the assembly and disassembly processes under which a community can reach such a maximum number, or whether this number is attainable in practice[10]. Specifically, can these assembly and disassembly processes operate smoothly until reaching the limits of coexistence, as assumed in classic coexistence studies[11,12]? Or is it possible that they find holes where coexistence abruptly breaks before reaching the limits, causing discontinuities in both processes[13]?

Indeed, our current understanding of the assembly and disassembly of ecological systems has two contrasting perspectives. The first perspective, grounded in classic works[14,15], suggests that these two processes are predictable following an axis of increasing ecological complexity. This predictability has been echoed by heuristic rules[16], mathematical formalisms[17] and conceptual frameworks[12] assuming that species coexistence is mostly additive. For instance, the mutual invasibility criterion establishes that, given a pool of S species, if all combinations of $(S-1)$ species coexist, the group of S species can also coexist, and vice versa. Consequently, under this first perspective, the assembly and disassembly processes are expected to operate smoothly until reaching the limits of coexistence (that is, reaching the maximum number of species that can coexist). The second perspective originates in evidence showing that the assembly and disassembly of ecological systems are very difficult to predict[18–20]. This difficulty comes from a complex interplay between many ecological and evolutionary processes, including species interactions[21], environmental conditions[22], dispersal[23], priority effects[1], phylogenetic relationships[24,25] and stochasticity[26]. Actually, while the mutual invisibility criterion can be a good rule of thumb, research has shown that it is unlikely to be observed in multispecies systems[27,28]. Consequently, under this second perspective, it is expected that coexistence breaks before reaching its limits, causing discontinuities in the assembly and disassembly processes. Importantly, the discrepancies between these two perspectives are partly due to the challenges of performing experimental work and partly due to the challenges of developing a formalism under which one can systematically study the multidimensional aspects of species coexistence[25].

Here, we introduce a new formalism to study the space of species coexistence formed by a given species pool. Our formalism captures coexistence using a hypergraph with vertices corresponding to species and hyperedges corresponding to species collections that coexist (one hyperedge per species collection that coexists). To study the space of species coexistence, we embed such coexistence hypergraphs in the Euclidean space. This embedding consists of associating to each vertex (that is, species) a point in the space and representing each hyperedge as a filled space between the corresponding vertices. In this way, filled spaces represent coexistence relationships between species groups, while empty spaces represent a lack of coexistence. Under this formalism, we show that discontinuities in the assembly and disassembly of ecological systems result in holes (empty spaces surrounded by filled space) of different dimensions in these coexistence hypergraphs, which we call coexistence holes. A coexistence hole occurs during the assembly process when a particular species collection is conceivable from a bottom-up perspective (that is, it can be assembled from sub-collections that coexist), but it is not realized (that is, it does not coexist). In turn, a coexistence hole occurs during the disassembly process when a particular species collection is realized (that is, it does coexist), but it cannot be disassembled into sub-collections

[1]CONACyT - Institute of Mathematics, Universidad Nacional Autónoma de México, Juriquilla, Mexico. [2]Institute of Mathematics, Universidad Nacional Autónoma de México, Juriquilla, Mexico. [3]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Department of Biology, McGill University, Montreal, Québec, Canada. [5]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada. [6]These authors contributed equally: Chuliang Song, Serguei Saavedra. ✉e-mail: mangulo@im.unam.mx; sersaa@mit.edu

(that is, its sub-collections cannot coexist). In this sense, coexistence holes indicate obstructions to assemble or disassemble a system. Detecting coexistence holes in ecological systems can help uncover hidden regularities in their assembly and disassembly, revealing what is possible and what is not.

We have organized this article as follows. First, we introduce formal definitions of coexistence holes, demonstrating their presence in small classic models of competition[6,29,30] and mutualism[31]. Next, we combine our formalism with a structuralist approach[10,13,32,33] to study the emergence of coexistence holes in systems with a larger number of species. Finally, we study coexistence holes in five experimental microbial systems. We conclude by discussing the possible applications and limitations of our formalism.

Discontinuities in assembly and disassembly processes

Coexistence holes. We consider ecological systems where individuals have been organized into S species (functional groups, taxa or other meaningful organizations). Following previous studies[34–36], we assume that every species collection either coexists or does not coexist (we discuss later how our framework can be adapted when coexistence depends on the initial species abundance). To study coexistence holes during the assembly process, we introduce the assembly hypergraph, H. This hypergraph has the isolated species $\{1,2,\cdots,S\}$ as vertices and it has one hyperedge, $h \subseteq \{1,\cdots,S\}$, for each different species collection that can coexist (Methods and Supplementary Note 1). Here, the dimension of a hyperedge $h \in H$ is $\dim(h)$ = 'the number of species in h' -1. The maximum $\max_{h\in H} \dim(h) + 1$ characterizes the limits of coexistence. Below such limits, the coexistence hypergraphs can have additional structure, as we discuss below.

We illustrate the concept of coexistence hypergraphs in the toy ecological system of $S=3$ species of Fig. 1a. In this hypothetical ecological system, species survive in isolation and coexist when assembled in pairs, but species will not coexist when assembled in a trio (Fig. 1b). The corresponding assembly hypergraph is $H = [\,[1],[2],[3],[1,2],[2,3],[3,1]\,]$ and the limit of coexistence is two species. By embedding the assembly hypergraph into a Euclidean space, discontinuities reveal themselves as holes in the hypergraph, which we call assembly holes. For example, embedding the assembly hypergraph H of the above toy ecological system into the plane reveals a one-dimensional assembly hole (Fig. 1c). The dimension of a hole is defined as the minimum dimension of its boundary. Zero-dimensional holes correspond to connected components of the hypergraph. We say a system has zero-dimensional holes if it contains two or more connected components. In our toy system, the hole's boundary, $\{[1,2],[2,3],[3,1]\}$, is one dimensional, making the hole itself one dimensional. A two-dimensional hole would appear, for example, in a hypergraph that looks like a tetrahedron with an empty interior. Thus, assembly holes occur during an assembly process to build a certain species collection (for example, to build $[1,2,3]$) if coexistence abruptly breaks at the end (Fig. 1d). Assembly holes characterize unexpected obstructions to build larger species collections from smaller ones—species collections that do not coexist, although they can be assembled from sub-collections that coexist.

To study coexistence holes during the disassembly process, we introduce the disassembly hypergraph, D. By definition, D contains all of the missing boundaries in the hyperedges of H, together with all species that survive in isolation (Methods and Supplementary Note 1). Therefore, except for zero-dimensional hyperedges (that is, isolated species), each hyperedge of D is a species collection that does not coexist despite it having been disassembled from a larger species collection that can coexist. To illustrate the disassembly hypergraph in an elementary case, consider the ecological system of $S=3$ species with coexistence outcomes as in Fig. 1e. Here, the trio of species coexists despite none of its pairs coexist. The assembly hypergraph for this hypothetical ecological system is $H = [\,[1],[1,2,3]\,]$ (see Fig. 1e). The corresponding disassembly hypergraph is $D = [\,[1],[2],[3],[1,2],[2,3],[3,1]\,]$. Thus, by embedding the disassembly hypergraph into a Euclidean space, discontinuities in the coexistence reveal themselves as disassembly holes. For example, embedding the disassembly hypergraph of the above toy ecological system into the plane uncovers a one-dimensional disassembly hole (Fig. 1f). Disassembly holes occur during a disassembly process starting at a certain species collection if coexistence abruptly breaks at the start (for example, starting at $[1,2,3]$ in Fig. 1g). Disassembly holes characterize unexpected obstructions to build smaller species collections from bigger ones—species collections that can be disassembled into sub-collections that do not coexist. In this sense, disassembly holes represent fragile species collections where removing one species will cause secondary extinctions.

Coexistence holes in small systems with classic dynamics. We illustrate assembly holes in the classic consumer–resource model of Tilman[21,29], considering $S=3$ species that compete for $M=2$ resources. Assuming that each species can reproduce using at least one resource, it follows that the three species always survive in isolation. An additional consequence of this model is the limits of coexistence[37]: with two resources, at most, two species can coexist (that is, the assembly hypergraph cannot have the interior $[1,2,3]$ of the triangle). However, the specific shape of the assembly hypergraph will depend on how the species consume the resources (Methods and Supplementary Note 2). If both resources are essential for the three species to grow, but one species is competitively superior to the other two, the assembly hypergraph will have two components characterized by two zero-dimensional assembly holes (Fig. 2a,b). Suppose now that the resources are complementary for one species (that is, species 2 has a positive reproductive rate even if one resource is absent; Fig. 2c). In that case, the assembly hypergraph has only one zero-dimensional assembly hole (Fig. 2d). Finally, suppose that species 1 consumes only resource 1, species 2 consumes both resources in a complementary way and species 3 consumes only resource 2 (Fig. 2e). Then, the assembly hypergraph has one one-dimensional assembly hole (Fig. 2f).

Disassembly holes occur, for example, in systems with mutualism or special types of competition between species (Methods). Specifically, two zero-dimensional disassembly holes occur when $S=2$ species survive through obligate mutualism only[31] (Fig. 2g,h). A one-dimensional disassembly hole occurs for $S=3$ species engaged in cyclic competition[38], where only species in isolation or in the trio coexist (Fig. 2i,j).

Uncovering coexistence holes via homology theory. In general, characterizing all discontinuities in the assembly and disassembly processes requires identifying all of the coexistence holes of different dimensions in the assembly and disassembly hypergraphs (see Methods for the formal definition of assembly and disassembly holes of arbitrary dimension). Identifying coexistence holes in systems with more than three species is challenging because their assembly/disassembly hypergraphs may not be adequately embedded into a two- or three-dimensional space. Consider the assembly hypergraph of Fig. 3a. In this example, the assembly holes are evident from the embedding into the plane. Figure 3b shows its associated disassembly hypergraph embedded in the plane. From this embedding, it is not evident that the species collection $[5,6,7,8,9]$ is a two-dimensional disassembly hole.

To overcome the above challenge, we constructed a novel homology theory to calculate the so-called Betti numbers of arbitrary hypergraphs (Methods and Supplementary Note 3). The Betti numbers are the vector $\beta = [\beta_0, \beta_1, \cdots]$, where β_k $(k \geq 0)$ counts the number of k-dimensional holes in the hypergraph. For example,

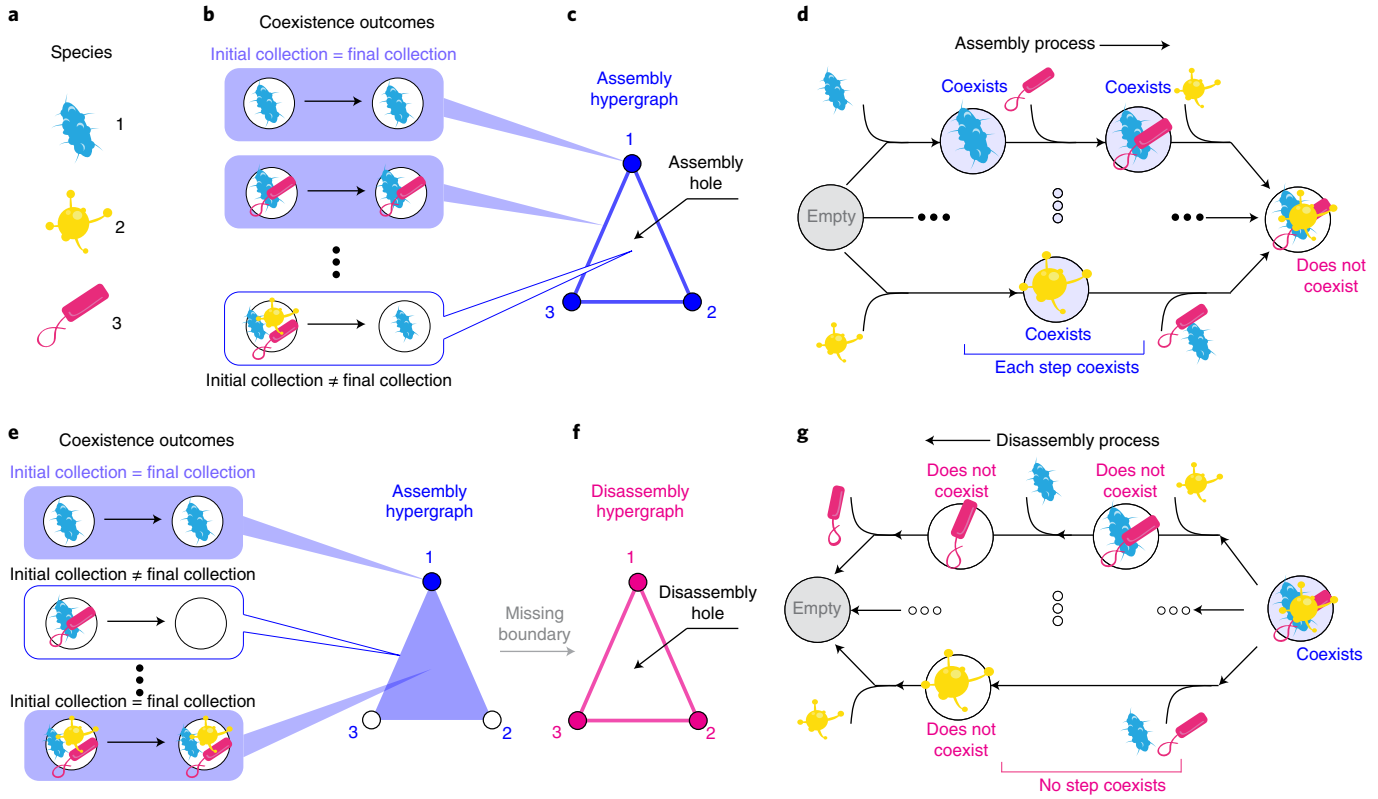

Fig. 1 | Coexistence holes characterize discontinuities in assembly and disassembly processes. a, A hypothetical pool of $S=3$ species. **b**, When each of the $2^S-1=7$ different species collection is assembled, it can either coexist (blue background) or not (white background). In this hypothetical example, species survive in isolation and coexist when assembled in pairs. However, the three species cannot coexist when assembled together. The corresponding assembly hypergraph is $H = [\,[1],[2],[3],[1,2],[2,3],[3,1]\,]$. **c**, When embedded into a two-dimensional space (that is, a plane), the assembly hypergraph reveals the assembly hole $h=[1,2,3]$. **d**, The assembly hole reveals that coexistence abruptly brakes: in all assembly processes to obtain [1, 2, 3], all of the intermediate species collections coexist, but in the final step coexistence does not occur. **e**, In these hypothetical coexistence outcomes, only species 1 survives in isolation and coexistence is possible only if the species are assembled in a trio. The corresponding assembly hypergraph is $H = [\,[1],[1,2,3]\,]$. **f**, The associated disassembly hypergraph is $D(H) = [\,[1],[2],[3],[1,2],[2,3],[3,1]\,]$, calculated from the missing boundary of H. Each hyperedge of D is a sub-community that does not coexist, despite it having been disassembled from the species collection [1, 2, 3] that coexists. When embedded into the plane, $D(H)$ uncovers the disassembly hole [1, 2, 3]. **g**, The disassembly hole reveals that coexistence abruptly brakes: despite [1, 2, 3] coexisting, in all disassembly processes starting from [1, 2, 3], not a single intermediate species collection with more than one species coexists.

the Betti numbers for the assembly hypergraph of Fig. 3a are $\beta(H)=[2,3,0,0,\cdots]$. The Betti numbers for the disassembly hypergraph of Fig. 3c are $\beta(D)=[4,1,1,0,\cdots]$. The Betti numbers distil the essential form of assembly/disassembly hypergraphs into their (homotopic) assembly/disassembly skeletons—namely, a minimal hypergraph with the same Betti numbers (Fig. 3b,d).

Results

A structuralist approach to explain coexistence holes. To study how coexistence holes emerge in ecological systems with a larger number of species, we adopt a structuralist approach[13,33] (Methods). We make this approach concrete by leveraging the mathematical tractability of the Lotka–Volterra model. Following previous works[32,39], we assume that the system's structure is phenomenologically captured by the interaction matrix $\boldsymbol{A} = \{a_{ij}\}_{i,j=1}^{S}$, with a_{ij} representing the effect of species j on the per-capita growth rate of species i. In turn, the context is phenomenologically captured by the intrinsic growth rates $\boldsymbol{r} = \{r_i\}_{i=1}^{S}$, with r_i representing how species i grows in isolation under given factors. We assume the context can change uniformly at random over the positive section of the unit sphere.

Under the above assumptions, from the universe of all possible assembly and disassembly skeletons, the interaction matrix $\boldsymbol{A}$ constrains which ones can be realized given the system's structure. From the set of all skeletons that can be realized, the intrinsic growth rates $\boldsymbol{r}$ determine which assembly and disassembly skeletons will be observed for that particular context. For example, consider the hypothetical system of $S=6$ species interacting as described by the interaction matrix $\boldsymbol{A}$ of Fig. 4a. From this interaction matrix, a fixed and finite set of possible assembly and disassembly skeletons emerge (Fig. 4b). Each one of these skeletons is compatible with a specific range of directions of $\boldsymbol{r}$, implying that each assembly and disassembly skeleton will be selected from the set of possible skeletons with a different probability. A large probability associated with a skeleton implies a larger fraction of external conditions (intrinsic growth rates) compatible with such a skeleton. Thus, the higher the probability of being selected, the higher the chances that a given skeleton occurs due to internal constraints (species interactions) alone. In our example, we have 4 possible assembly skeletons and 14 possible disassembly skeletons (Fig. 4c). Skeletons with assembly and disassembly holes are possible (white bars in Fig. 4c). Ecological systems with a higher strength or number of interspecific interactions can adopt a larger number of skeletons (Supplementary Note 5.1 and Supplementary Fig. 5).

Coexistence holes in random Lotka–Volterra systems. Figures 2i,j and 4a–c show that assembly and disassembly holes emerge in simple population dynamics, without sophisticated mechanisms such

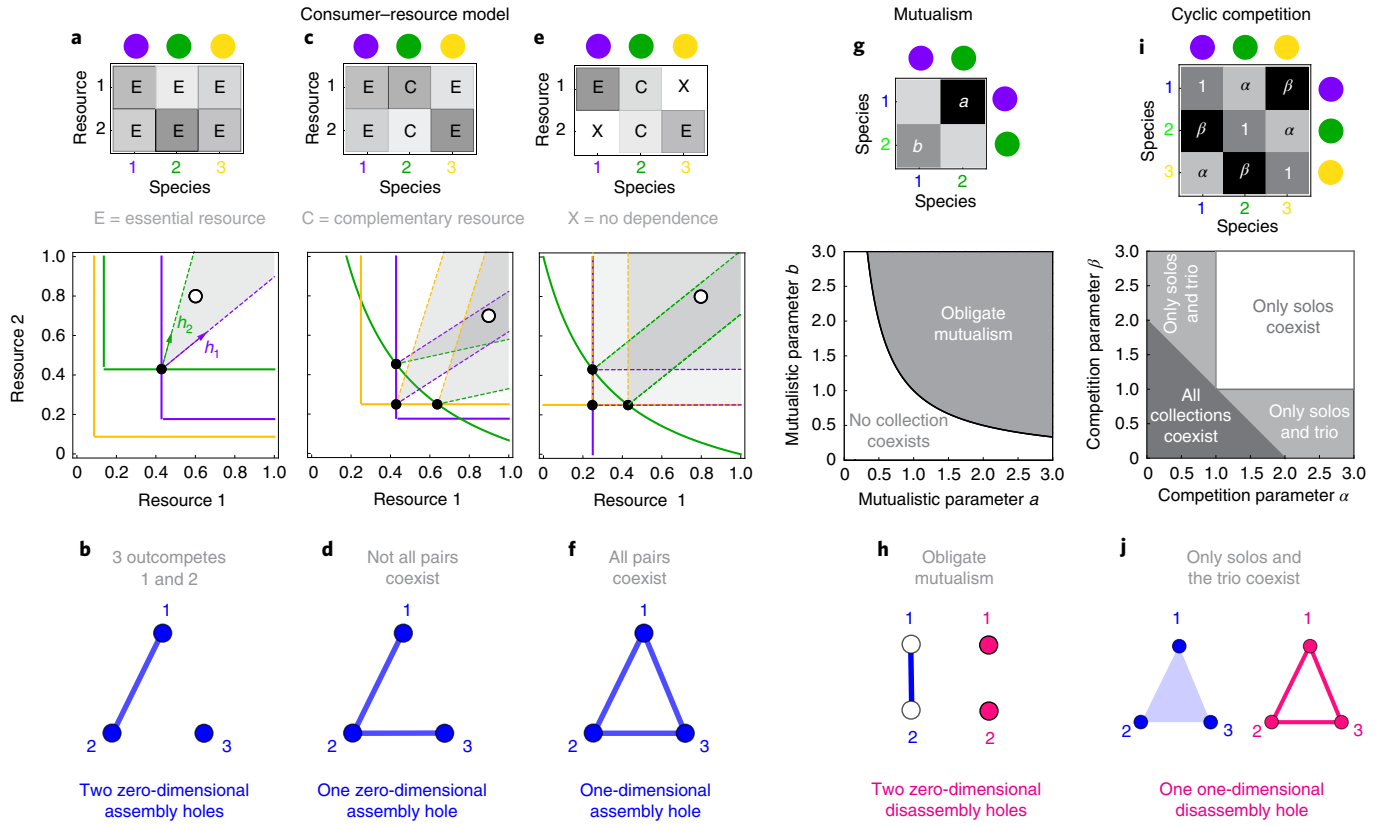

Fig. 2 | Assembly and disassembly holes in classic ecological models. a–**f**, Coexistence outcomes for $S=3$ species and $M=2$ resources using Tilman's consumer–resource model[29] under different conditions. In this model, the limit of coexistence implies that, at most, $M=2$ species can coexist. **a**, Top: impact matrix specifying how each species consumes the resources. In this example, all resources are (perfectly) essential for all species. Bottom: ZNGI for each species (coloured lines). An equilibrium occurs when the NZGIs intersect (black dot). The external supply rate (R_1°, R_2°) is shown as a white circle. The grey area shows the cone spanned by the impact vectors h_i of species 1 and 2 (columns of the impact matrix). If this cone contains the supply rate, the equilibrium is feasible (that is, it admits a positive abundance for each species). Note that this equilibrium can be stable or not. Under the conditions of the example, species 1 and 2 can coexist since their equilibrium is feasible and stable. However, species 3 will outcompete both of them. **b**, Assembly hypergraph corresponding to **a**. **c**, As in **a**, but now the two resources are complementary for species 2. Under these conditions, at most two of the three pairs can coexist. **d**, Assembly hypergraph corresponding to **c**. **e**, As in **c**, but now species 1 only consumes resource 1 and species 3 only consumes resource 2. In this scenario, all pairs of species can coexist. However, the limits of coexistence imply the trio of species cannot. **f**, Assembly hypergraph corresponding to **e**. **g**, Top: interactions between $S=2$ species in a classic model of mutualism[31], with the interaction strength characterized by two parameters: $a, b>0$. In this model, species die in isolation or if the mutualistic strength is low ($ab<1$; white region in bottom panel). If the mutualistic strength is strong, both species coexist in an obligate mutualism ($ab>1$; grey region in bottom panel). **h**, Assembly (blue) and disassembly (pink) hypergraphs corresponding to the obligate mutualism shown in **g**. **i**, Top: species interactions in the classic May–Leonard model of cyclic competition between $S=3$ species. Here, $\alpha, \beta>0$ characterizes the competition strength. Different coexistence outcomes are possible depending on the values of the competition strengths (bottom panel). **j**, Assembly and disassembly hypergraphs corresponding to **i**.

as complex functional responses[40] or higher-order interactions[41]. To investigate how general this conclusion is, we systematically analysed an ensemble of systems governed by Lotka–Volterra dynamics with random interaction matrices $\boldsymbol{A}$ (Methods and Supplementary Note 5.2). Two parameters of the interaction matrix characterize the ensemble: the connectance, $C_A \in [0,1]$, describing the probability that two species interact, and the typical interspecific interaction strength, $\sigma_A>0$. We found that as soon as the interaction matrix exceeds a small complexity threshold measured by $\sigma_A C_A$, the presence of assembly and disassembly holes is unavoidable (Fig. 4d and Supplementary Note 5.3). This result implies that discontinuities are the norm rather than the exception in the assembly and disassembly of random (unstructured) Lotka–Volterra systems.

We also found that coexistence holes in random Lotka–Volterra systems obey two general patterns. First, knowledge of the disassembly skeleton strongly determines which assembly skeleton the system can adopt, but not vice versa (Supplementary Note 5.4 and Supplementary Fig. 6). That is, there is an asymmetry in the information contained in the assembly and disassembly skeletons. Second, low-dimensional and high-dimensional assembly/disassembly holes are unlikely to co-occur. That is, when the numbers of low-dimensional (that is, one- or two-dimensional) and high-dimensional (that is, three- or four-dimensional) holes are used as coordinates in a plane, the assembly and disassembly skeletons generated by random Lotka–Volterra models avoid the upper-right corner (Fig. 4e). This pattern arises because the likelihood of observing assembly or disassembly holes depends on their dimensions and the particular interaction matrix (Supplementary Note 5.5 and Supplementary Fig. 7). Specifically, high-dimensional assembly holes are more likely at weak interspecific interactions, whereas low-dimensional holes are more likely at strong interspecific interactions. We also observe that the probability of finding an assembly or disassembly hole decreases with the hole's dimension, and that systems tend to generate more disassembly skeletons than assembly skeletons (Supplementary Fig. 7). Mathematically, it is possible to conceive hypergraphs that simultaneously have many

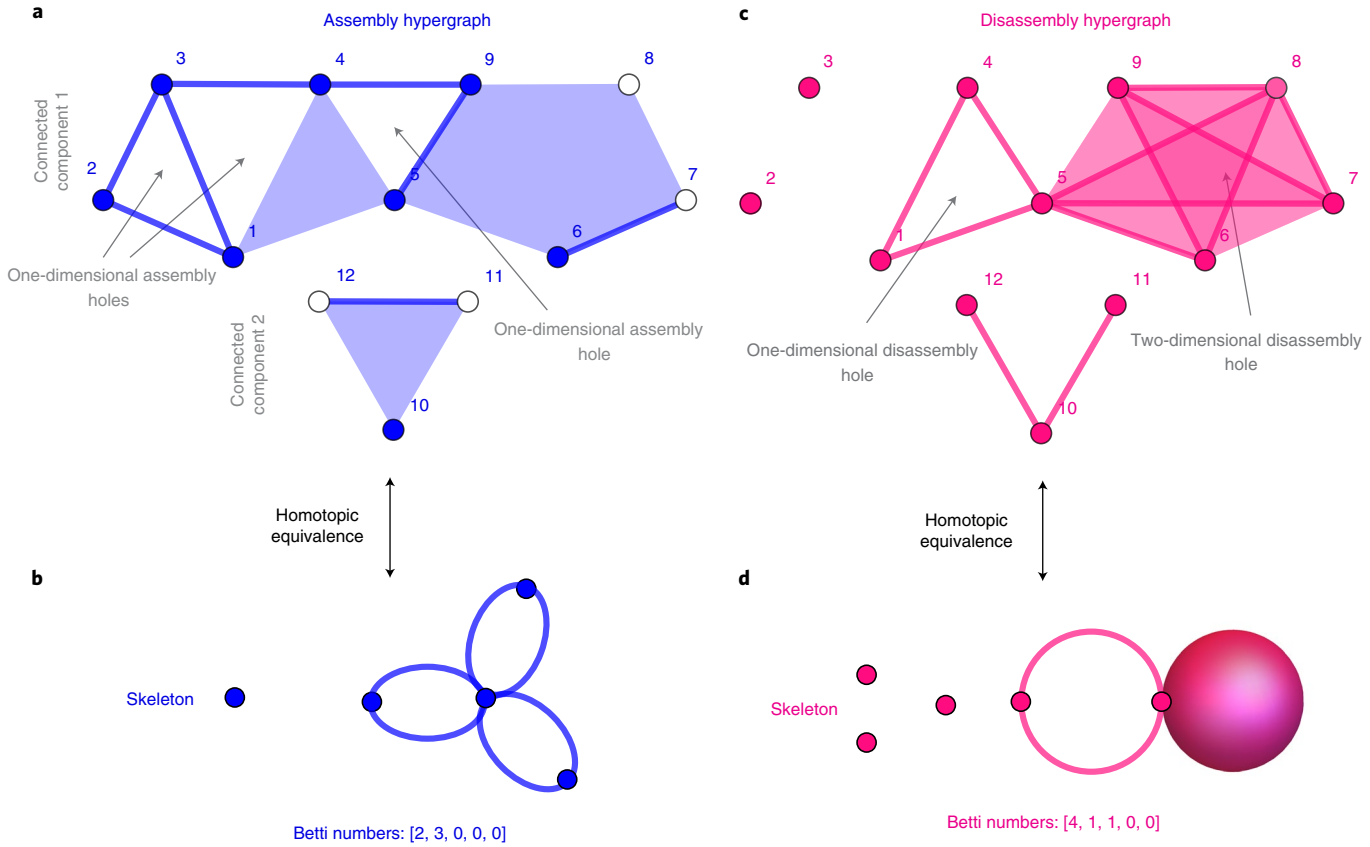

Fig. 3 | Homology reveals the essential structure of assembly and disassembly hypergraphs in terms of their skeletons. a, An assembly hypergraph for a hypothetical system of $S=12$ species. While the limit of coexistence in this example is five species, there exists additional structure within this limit. In particular, this assembly hypergraph contains two connected components (that is, [10, 11, 12] and the rest) and three one-dimensional assembly holes. **b**, Homotopic skeleton of the assembly hypergraph of **a**. The skeleton is quantified by its Betti numbers. **c**, Disassembly hypergraph corresponding to the assembly hypergraph of **a**. The disassembly hypergraph contains four connected components, one one-dimensional hole and one two-dimensional hole. Since this hypergraph is embedded in the plane, its two-dimensional hole is not obvious. **d**, Homotopic skeleton of the disassembly hypergraph of **c**. The skeleton is quantified by its Betti numbers.

low- and high-dimensional holes. Therefore, the absence of such skeletons shows how the internal constraints of random ecological systems (characterized by the interaction matrix) shape the likelihood of observing certain discontinuities.

Coexistence holes in empirical ecological systems. To study whether coexistence holes occur in empirical ecological systems, we analysed five experimental microbial communities: Vandermeer[42] ($S=4$), Friedman[16] ($S=8$), Stein[43] ($S=12$), Venturelli[44] ($S=14$) and MDSINE[45] ($S=16$). In each of these studies, a Lotka–Volterra model was parameterized and systematically validated using high-resolution experimental data, resulting in empirical estimations of the parameters $\boldsymbol{A}$ and $\boldsymbol{r}$ (Methods and Supplementary Note 6).

All five empirical interaction matrices can generate assembly and disassembly holes (Supplementary Fig. 8a–e). The numbers of possible assembly and disassembly skeletons differ across ecological systems, but most skeletons contain assembly and disassembly holes. In the two systems with smaller species pools, the empirical interaction matrices generate more assembly skeletons than disassembly skeletons (1.8 and 1.55 times more for Vandermeer's and Friedman's, respectively). The small number of disassembly skeletons in these systems could be explained by their size and simple assembly rules[16]. For the systems with larger species pools, the empirical interaction matrices generate between 2.14 (Venturelli's) and 221.5 (MDSINE) times more disassembly skeletons than assembly skeletons, as happens in random (unstructured) Lotka–Volterra systems. Importantly, these interaction matrices for larger species collections tend to have a larger proportion of skeletons with coexistence holes (that is, with discontinuities). We also confirm our expectations that low- and high-dimensional assembly/disassembly holes are unlikely to co-occur (grey points in Supplementary Fig. 8j–f). In the three systems with larger species pools, we also find that disassembly skeletons strongly influence assembly skeletons, but not vice versa (Supplementary Fig. 9). For example, in Stein's system, knowing which disassembly skeleton was adopted reduces the uncertainty of the adopted assembly skeleton by 91.68%. In contrast, knowing the assembly skeleton reduces the uncertainty of the adopted disassembly skeleton by 63.42%. In the MDSINE system, knowing the disassembly reduces the uncertainty of assembly by 74.54%, while assembly reduces the uncertainty of disassembly by <0.05%.

The role of internal and external constraints. Next, we focus on the five empirical assembly and disassembly hypergraphs obtained from the experimentally parameterized ($\boldsymbol{A}$, $\boldsymbol{r}$) (Fig. 5a–e). For Vandermeer's and Friedman's, we find that their assembly hypergraphs are a simplicial complex. A simplicial complex is a particular hypergraph where each hyperedge has all of its boundaries (Fig. 5a,b). Notice that this occurs if and only if the associated disassembly hypergraph consists only of the isolated vertices. A system whose assembly hypergraph is a simplicial complex follows the sim-

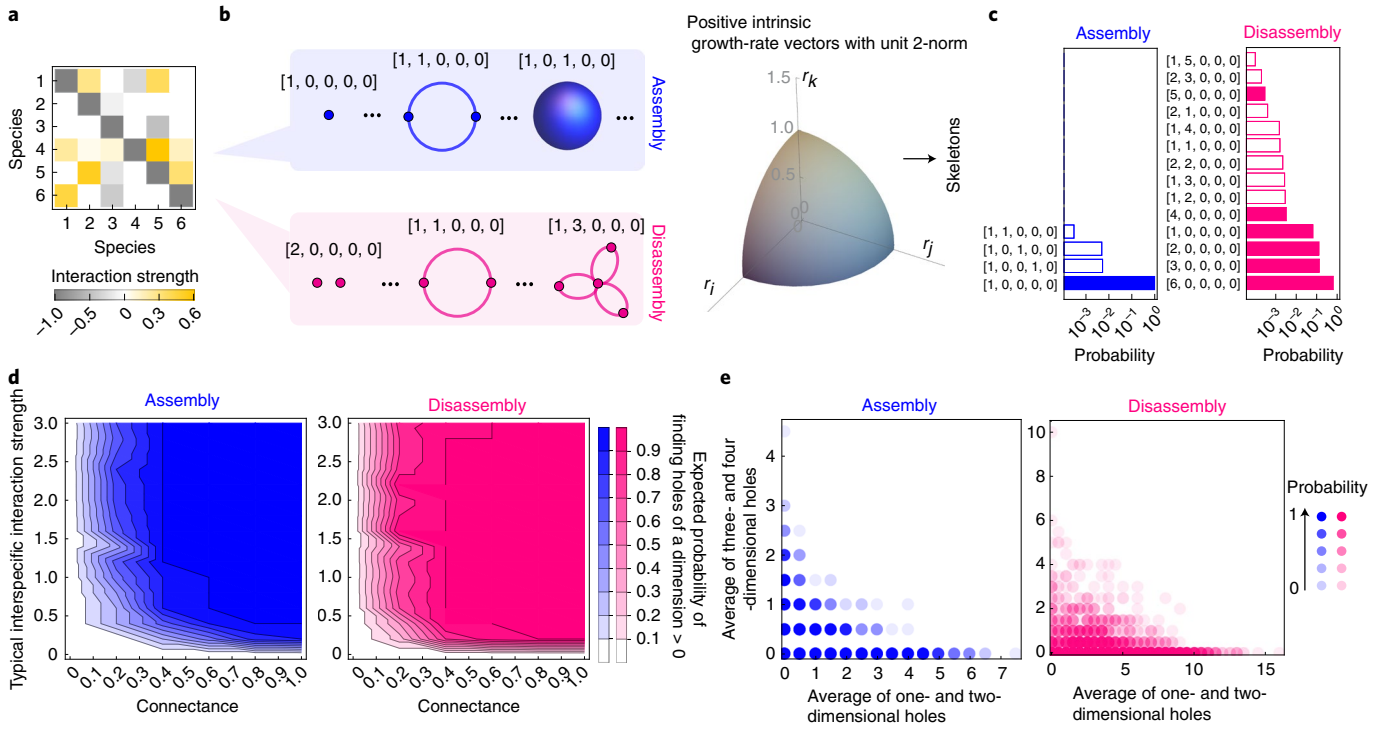

Fig. 4 | A structuralist approach to understanding the emergence of coexistence holes in the Lotka–Volterra model. a, A hypothetical interaction matrix, ***A***, describing the interactions between $S=6$ species (see values in Supplementary Equation (4)). **b**, Possible assembly and disassembly skeletons generated by the interaction matrix of **a**. **c**, By choosing **r** uniformly at random over the positive section of the unit 2-norm sphere $\{\mathbf{r} \in \mathbb{R}^S_{\geq 0} | r_1^2 + \cdots + r_S^2 = 1\}$ (left), some assembly/disassembly skeletons are more likely to occur (right). Note that the unit sphere is illustrated in three dimensions, although in reality it is a unit sphere of dimension $S=6$. Here, solid bars denote skeletons without holes and hollow bars denote skeletons with at least one hole. **d**, Expected probability of finding at least one assembly (blue) or disassembly hole (pink) in the ensemble of Lotka–Volterra models with random parameters and $S=8$ species. **e**, Not all possible assembly/disassembly skeletons are possible in ensembles of Lotka–Volterra models with random parameters. Namely, note that skeletons that simultaneously have a large number of low-dimensional holes (that is, one- or two-dimensional holes) and high-dimensional holes (that is, three- and four-dimensional holes), which occupy the top-right corners of the panels, are very unlikely.

ple closed-under-inclusion assembly rule: a species collection coexists if all of its sub-collections coexist. For Friedman's system, this result agrees with the existence of a simple assembly rule derived in their original work[16]. There is a large probability of observing such a simple assembly rule for Vandermeer's system due to internal constraints alone (among the random disassembly skeletons, the probability is $P=0.82$ for observing the empirical disassembly skeleton with only isolated vertices). For Friedman's system, such probability is lower ($P=0.25$), indicating that external factors (intrinsic growth rates) also play a role. Interestingly, it is possible to explain the lack of holes of dimension 1 or higher in Vandermeer's system by internal constraints alone ($P=0.48$). In contrast, explaining the presence of the two one-dimensional assembly holes in Friedman's system also requires considering external factors ($P=0.08$ for internal constraints alone).

Stein's and Venturelli's systems contain higher-dimensional assembly and disassembly holes (Fig. 5c,d). The empirical assembly and disassembly skeletons of Stein's system cannot be explained by internal constraints alone ($P=6.66\times10^{-4}$ and $P<3.33\times10^{-4}$, respectively). A similar situation occurs for Venturelli's ($P=0.13$ and $P=0.01$ for the empirical assembly and disassembly skeletons). For MDSINE (the system with the largest species pool), the dimension of its assembly and disassembly hypergraphs only allowed us to compute its assembly and disassembly holes up to dimension 2. This empirical assembly skeleton consists of a single connected component without holes, and internal constraints alone can explain it ($P=0.999$). However, its empirical disassembly skeleton has holes, and it is unlikely by internal constraints alone ($P=6\times10^{-3}$). These results confirm that internal constraints (species interactions) play an important role in the emergence of coexistence holes.

Finally, to understand what properties render an empirical skeleton more or less likely to be observed, we compared the number of empirical coexistence holes of different dimensions with their randomizations (Fig. 5f–j). We find cases of depletion or (strong) enrichment in the number of coexistence holes, depending on the particular system and the hole's dimension. In all empirical assembly and disassembly skeletons, low- and high-dimensional holes do not co-occur (colours in Supplementary Fig. 8), supporting our finding that coexistence holes obey general patterns. These results indicate that coexistence holes occur in empirical systems at specific dimensions due to a combination of internal constraints and external factors.

Discussion

Coexistence in multispecies systems has ubiquitous discontinuities or obstructions characterized by coexistence holes. Assembly holes do not exist if and only if $\beta(H)=[1,0,\cdots,0]$. In this particular case, it is possible to build each species collection that coexists from its sub-collections (that is, assembled from the bottom up), and an invasion analysis[46] can be applied with confidence. Our analysis of random (unstructured) and empirical ecological systems based on the Lotka–Volterra model suggests that the absence of coexistence holes is the exception rather than the norm. In particular, with random interactions, coexistence holes are more likely to occur when their complexity increases. This result aligns with May's complexity–stability trade-off[47,48], indicating that stability loss contributes

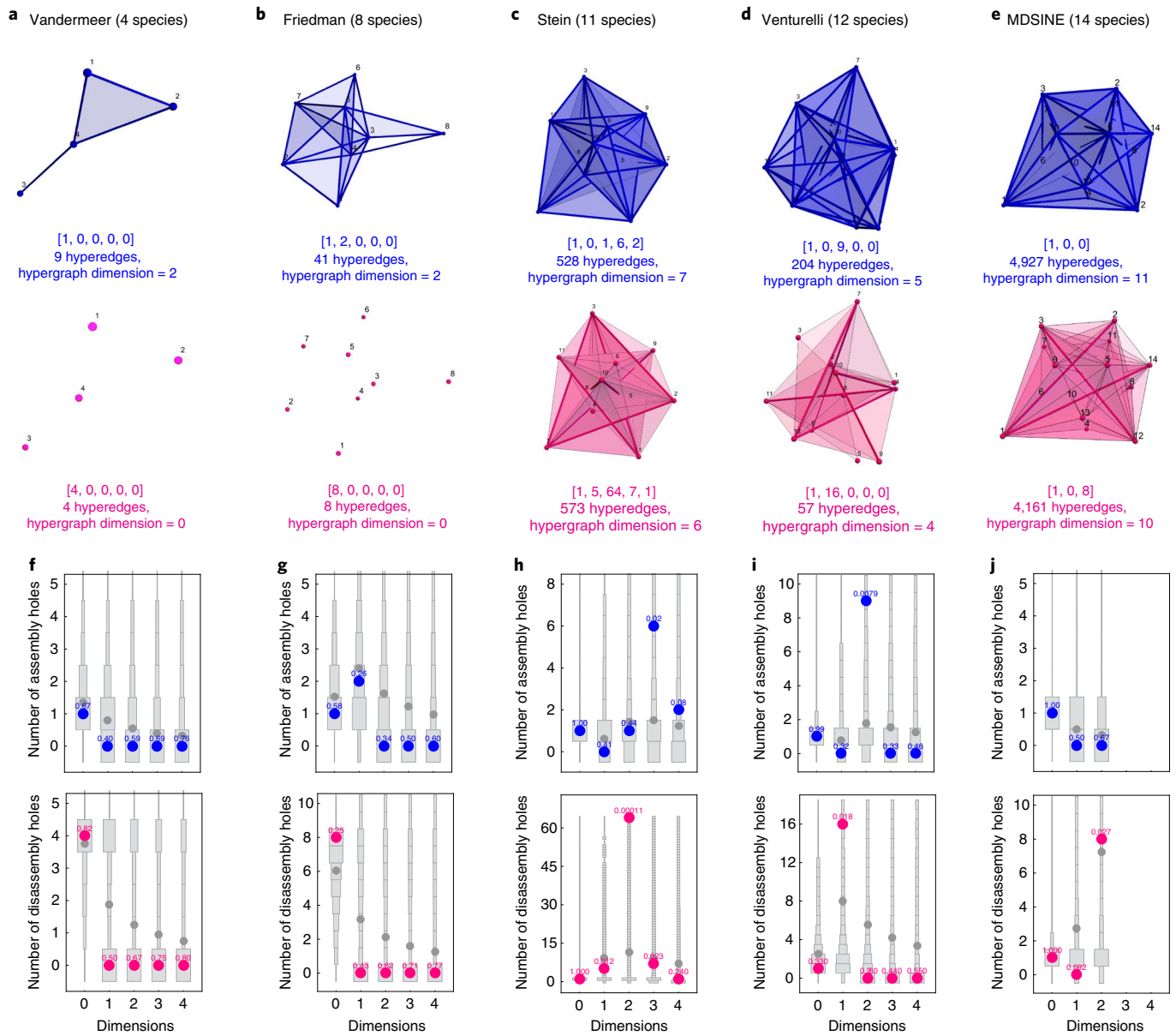

Fig. 5 | Assembly and disassembly holes in empirical ecological systems. a–e, Embedding of the assembly (blue) and disassembly (pink) hypergraphs for the empirical interaction matrix and empirical intrinsic growth-rate vector. **f–j**, Statistical analysis of the number of empirically observed assembly (blue) and disassembly (pink) holes with respect to the number of holes observed in the 3,000 randomizations of the intrinsic growth-rate vectors chosen uniformly at random over the positive sector of the unit sphere (grey). The grey dots denote the expected number of holes over the randomizations. The numbers above each coloured dot denote the probability of observing that number of holes among the 3,000 randomizations.

to generating discontinuities in the assembly and disassembly of ecological systems. It remains open to understand the prevalence of coexistence holes in more realistic situations, such as when species have specific interaction types[49], when the interactions are non-randomly organized[50] or when abiotic conditions affect species interactions[51].

Our analysis also suggests that coexistence holes obey general patterns. Knowing the disassembly skeleton of a system also informs about the assembly skeleton, but not vice versa. This result suggests that, in multispecies experiments, focusing on characterizing their disassembly is better as it contains more information about both processes. Understanding which ecological properties lead to this pattern is an important open question for future research. We also find that low- and high-dimensional coexistence holes are unlikely to co-occur. This result is expected for random systems. For example, high-dimensional assembly holes need larger species collections to coexist to form the hole's boundary. For random systems, this is more likely with weak interspecies interactions. In contrast, low-dimensional assembly holes can occur even when small species collections coexist, allowing for stronger interspecies interactions. Our analysis of empirically parametrized Lotka–Volterra models supports the hypothesis that species interactions play a crucial role in the emergence of coexistence holes.

Additional theoretical work is needed to understand the conditions leading to coexistence holes. Rough conditions can be derived from existing work. For example, additive coexistence[52] guarantees that the system's assembly is a complete hypergraph between all species that survive in isolation. In general, additive coexistence occurs if species occupy sufficiently different ecological niches. In the Lotka–Volterra model, some particular system structures, such

as food chains, imply the absence of coexistence holes[53] (see also refs. [25,54]). Our formalism can be adapted to cases when coexistence depends on the initial species abundance. Using a probability distribution for the initial species abundance, it is possible to calculate the corresponding Betti numbers' statistics.

One limitation of our study is that we calculated assembly hypergraphs using mathematical models. Ideally, assembly hypergraphs should be entirely constructed by experimentally testing each species collection's coexistence. This task is feasible for systems with a few species[55] or systems where massive automated co-culture experiments are possible[56]. We followed this approach to construct the assembly hypergraph of $S=5$ bacterial species in *Drosophila melanogaster* gut microbiota from in vivo experimental data[55] (Supplementary Note 7). The assembly hypergraph we obtained has one connected component and two one-dimensional assembly holes, supporting our hypothesis that coexistence holes are ubiquitous in ecological systems. In general, for systems with many species, the combinatorial explosion 2^S-1 in the number of different species collections makes it impossible to rely on experiments only. Circumventing this limitation requires parameterizing population dynamics models to predict the coexistence of some species collections (Supplementary Note 4.3). Future work should not overlook the limitations of the Lotka–Volterra model[57]. In general, we expect that coexistence holes become more likely in more detailed ecological models[58].

Our formalism allows comparing the assembly and disassembly of ecological systems using the corresponding hypergraphs' skeletons even if they have a different number of species. For example, regardless of its number of species or dynamics, a system follows the simple closed-under-inclusion assembly rule if and only if $\beta(D)=[S,0,\cdots,0]$. Indeed, comparing objects through the lenses of holes is far from a novel idea. Algebraic topology—the branch of mathematics roughly described as the study of holes—originated in the work of the great mathematician Henri Poincaré in the nineteenth century[59]. A discovery of algebraic topologists is that, even in low dimensions, holes can appear in many non-evident ways, such as in the infamous Klein bottle (the bottle has two holes, one which disappears if traversed twice). Recently, algebraic topology is finding more applications across the sciences[60–62]. Unfortunately, with the notable exception of Sugihara's work in the 1980s[63], algebraic topology has been an underused tool in ecology.

Our work shows how species coexistence is a complex interplay between filled spaces (that is, species collections that coexist) and empty spaces (that is, species collections that do not coexist). Some empty spaces give rise to assembly and disassembly holes that characterize cases when coexistence abruptly breaks, implying that coexistence (or biodiversity) is a discontinuous process[13]. Identifying these discontinuities can improve our understanding of which processes in nature are driven by internal constraints of design and not merely by randomness, as it can be naively perceived.

Methods

Definition of assembly and disassembly hypergraphs. Denote by $\mathrm{V}=\{1,2,\cdots,S\}$ the species pool and by 2^{V} its power set (that is, the collection of all subsets of V). We adopt as convention that sets do not contain repeated elements. Given a species collection $\Sigma\in 2^{\mathrm{V}}$, we assume its coexistence is a dichotomy. Thus, we formalize species coexistence as a function $c\colon 2^{\mathrm{V}}\to\{0,1\}$. For any species collection $\Sigma\in 2^{\mathrm{V}}$, we interpret the condition $c(\Sigma)=1$ as 'species in Σ coexist' and $c(\Sigma)=0$ as 'species in Σ do not coexist'. If Σ contains a single species, coexistence is interpreted as surviving in isolation. For mathematical completeness, for the empty set $\emptyset\in 2^{\mathrm{V}}$, we define $c(\emptyset)=0$.

To describe species coexistence in an ecological system, we need to encode all coexistence relationships that one species has with any other species collection. These relationships can be encoded using a hypergraph (Supplementary Note 1). Specifically, the assembly hypergraph $H=(\mathrm{V},\mathrm{H})$ captures all coexistence relationships, where V is its vertex set and $\mathrm{H}\subseteq 2^{\mathrm{V}}$ is its hyperedges given by:

$$\mathrm{H}:=\{\Sigma\in 2^{\mathrm{V}}|c(\Sigma)=1\}.$$

Note that in this hypergraph, a successful assembly process adding one species at a time consists of moving from an initial hyperedge of dimension d (representing an initial species collection that coexists) to some adjacent hyperedge of dimension $d+1$ (for example, moving from [1] to [1, 2] in Fig. 1a).

Next, we introduce the notion of disassembly hypergraph $D(H)$. For a hypergraph H, denote by $K(H)=\{\sigma\subset\mathrm{V}|\sigma\subset\tau$ for some $\tau\in H\}$ the minimal simplicial complex containing it. Recall that a simplicial complex is a hypergraph H such that if $h\in H$ is a hyperedge then $\tau\in H$ for all $\tau\subset h$. In words, in a simplicial complex, each hyperedge contains all of its boundaries. Define the missing boundary of a hypergraph H as $M(H):=K(H)\setminus H$. Then, the disassembly hypergraph $D(H)$ of the hypergraph $H=(\mathrm{V},\mathrm{H})$ is $D=(\mathrm{V},\mathrm{D}(H))$, where:

$$\mathrm{D}(H)=M(H)\cup\{h\in\mathrm{H}\text{ such that cardinality }(h)=1\}.$$

Assembly holes in Tilman's consumer–resource model. This classic model considers S species indirectly interacting by consuming M resources that are supplied to the system[29]. Denoting by $x_i(t)$ and $R_j(t)$ the abundance of the ith species and jth resource at time $t\geq 0$, the model takes the form:

$$\begin{aligned}\dot{x}_i &= x_i\,[\,f_i(R_1,\cdots,R_M)-m_i\,],\ i=1,\cdots,S,\\ \dot{R}_j &= (R_j^\circ-R_j)-\sum_{k=1}^{M}h_{jk}\,x_k f_k(R_1,\cdots,R_M),\ j=1,\cdots,M.\end{aligned}\tag{1}$$

Above, the symbol '˙' denotes a derivative with respect to time. The parameters of the model are: the mortality rate $m_i>0$ of the i-th species, the functional relationship $f_i:\mathbb{R}^M\mapsto\mathbb{R}$ between the availability of resources and the per-capita growth rate of species i, the supply rate $R_j^\circ\geq 0$ of the j-th resource and the impact vector $h_i=(h_{1i},\cdots,h_{Mi})^T$ of the i-th species (with $h_{ji}\geq 0$ describing the amount of resource j required to produce an individual of species i). The impact matrix is $H=(h_{ij})\in\mathbb{R}^{M\times S}$. Recall that, in this model, a feasible interior equilibrium exists for some species collection only if the supply rate $R^\circ=(R_1^\circ,\cdots,R_M^\circ)$ lies in the cone spanned by their impact vectors[29]. If a feasible equilibrium exists, the species collection coexists if such equilibrium is stable.

For our analysis, we studied $S=3$ species, $M=2$ resources and three different forms in which species consume the resources. First, in the system of Fig. 2a,b, both resources are essential for the three species (that is, the species' reproductive rate is zero if any resource is absent). This is modelled using

$$f_i(R_1,R_2)=\min\left\{\mu_i\frac{R_1}{R_1+K},\frac{R_2}{R_2+K}\right\}$$

and the impact matrix

$$H=\begin{pmatrix}2.49771 & 0.889986 & 1.959\\ 2.04922 & 3.22467 & 2.46907\end{pmatrix}.$$

We use the parameters $K=1$, $m_1=0.15$, $m_2=0.3$, $m_3=0.08$, $\mu_1=0.5$, $\mu_2=2.5$ and $\mu_3=1$ and the supply rate $R^*=(0.6,0.8)$. The bottom panel of Fig. 2a displays the zero net growth isoclines (ZNGIs) in the resource plane where the species' population does not change. The species collection [1,2] coexists since it has a feasible equilibrium (the supply rate is the cone spanned by the impact vectors). Furthermore, this equilibrium is stable (Supplementary Fig. 1a). The ZNGI for species 3 is below the ZNGIs of the other two species, indicating that species 3 is competitively superior and it cannot coexist with the other two species. Therefore, the assembly hypergraph has two zero-dimensional assembly holes represented by two connected components (Fig. 2b). More broadly, for essential resources, we find that at most one pair of species will coexist, even if the ZNGI of species 3 intersects the ZNGIs of species 1 and 2. Hence, the assembly hypergraph of the system always contains at least two zero-dimensional holes. Note that zero-dimensional assembly holes characterize groups where species coexist with some species of the same group, but they do not coexist with any species of other groups.

Second, in the system of Fig. 2c,d, resources are essential for species 1 and 2 but complementary for species 2 (that is, species 2 has a positive reproductive rate even if one resource is absent). Specifically, we keep the same $f_1(R_1R_2)$ and $f_3(R_1,R_2)$ as in the first system, but for species 2 we use:

$$f_2(R_1,R_2)=\frac{R_1}{R_1+K}+0.8\frac{R_2}{R_2+K},$$

with $K=1$. The impact matrix is chosen as:

$$H=\begin{pmatrix}1.49084 & 1.63248 & 0.531918\\ 0.960242 & 0.356332 & 1.69407\end{pmatrix}$$

and the supply rates are $R^*=(0.9,0.7)$. Geometrically, these choices result in a curved ZNGI for species 2 that intersects the ZNGIs of the other two species (bottom panel of Fig. 2c). For this particular example, a feasible equilibrium

exists for all pairs of species. However, only two of these three equilibria are stable (Supplementary Fig. 1b). Consequently, the assembly hypergraph looks like a broken triangle and consists of one connected component (Fig. 2d).

Finally, in the system of Fig. 2e,f, species 1 consumes only resource 1, species 2 consumes both resources in a complementary way and species 3 consumes only resource 2. We model this case using:

$$f_1(R_1,R_2)=0.5\frac{R_1}{R_1+K},\ f_3(R_1,R_2)=\frac{R_2}{R_2+K}.$$

In this sense, resource 1 is essential for species 1 and resource 2 is essential for species 3. Furthermore, we assume that species 2 consumes both resources, and that for this species both resources are complementary:

$$f_2(R_1,R_2)=\frac{R_1}{R_1+K}+\frac{R_2}{R_2+K}.$$

For the results of the text, we choose $K=1$ and the mortality rates $m_1=0.15$, $m_2=0.4$ and $m_3=0.3$. The impact matrix is:

$$H=\begin{pmatrix}3.93957 & 1.0809 & 0\\ 0 & 0.86745 & 3.33962\end{pmatrix}.$$

The supply rate is $R^*=(0.8,0.8)$. Under these conditions, a stable interior equilibrium exists for all pairs of species (Supplementary Fig. 1c). However, since the triangle's interior cannot be realized due to the limits of coexistence, the assembly hypergraph has one one-dimensional assembly hole (Fig. 2f).

Disassembly holes in classic models of mutualism and cyclic competition. We study disassembly holes in a classic phenomenological model of mutualism for $S=2$ species[31]. Denoting by $x_i(t)$ the abundance of the i-th species at time t, the model takes the form:

$$\dot{x}_1=x_1\left(r_1-\frac{r_1}{K(x_2;b)}x_1\right),$$
$$\dot{x}_2=x_2\left(r_2-\frac{r_2}{K(x_1;a)}x_2\right),$$

with $r_i>0$ the intrinsic growth rate of species i, and $K(x;c)=k_0\left(1-\exp\left(-\frac{cx}{k_0}\right)\right)$. The mutualistic interaction strength of species 2 on species 1 (respectively species 1 on species 2) is described by the parameter $a>0$ (respectively $b>0$) (see Fig. 2g). In this model, species cannot survive in isolation, but the two species can coexist if the mutualistic interaction strengths are large enough ($ab>1$). The parameter plane (a,b) has an obligate mutualism region where both species coexist (bottom panel of Fig. 2g). In this region, the coexistence hypergraph is $H=[\,[1,2]\,]$, producing a disassembly hypergraph $D=[\,[1],[2]\,]$ with two zero-dimensional disassembly holes (Fig. 2h). Higher-dimensional disassembly holes can appear in systems involving many species that coexist through obligate mutualism only. For the results presented in the main text, we choose $r_1=r_2=1$ and $k_0=1$.

We also study disassembly holes in the May–Leonard classic model of cyclic competition between $S=3$ species[30]. In this model, species directly compete with each other according to the Lotka–Volterra equations with the interaction matrix shown in Fig. 2i and equations given by

$$\dot{x}_1=x_1\left(1-x_1-\alpha x_2-\beta x_3\right),$$
$$\dot{x}_2=x_2\left(1-\beta x_1-x_2-\alpha x_3\right),$$
$$\dot{x}_3=x_3\left(1-\alpha x_1-\beta x_2-x_3\right).$$

The parameters $\alpha>0$ and $\beta>0$ describe the competitive strength between species. Species always survive in isolation, but their coexistence in pairs and the trio depend on the competition parameters (bottom panel of Fig. 2i). When the competition parameters are asymmetrical, only isolated species and the trio of species coexist, producing one one-dimensional disassembly hole (Fig. 2j).

A homology theory for hypergraphs. Here, we present a summary of our methods, referring to the Supplementary Notes for details and examples. Our formalism considers arbitrary hypergraphs H over the set of vertices $V=\{1,2,\cdots,S\}$. By construction, hypergraphs are combinatoric objects (that is, collections of subsets of V). To study their structure, and in particular to identify their holes, we need to endow hypergraphs with the structure of a topological space. We do this as usual by embedding the hypergraph into an S-dimensional Euclidean space $\mathbb{R}^S$. First, we associate to each vertex $i\in V$ the i-th unit vector of $\mathbb{R}^S$. Second, a hyperedge $h\in H$ is associated to the relative interior of the simplex spanned by the unit vectors associated to the vertices it contains—a process we denote as relint(h). In this form, the embedding or geometric realization $|H|\subseteq\mathbb{R}^S$ of the hypergraph H is:

$$|H|:=\bigcup_{h\in H}\mathrm{relint}(h).$$

Note that the location of the vertices in the space is arbitrary. Therefore, we focus only on the topological properties of the geometric realization of hypergraphs—that is, properties that remain invariant when the geometric realization of a hypergraph is transformed without glueing or cutting it (that is, invariant to homeomorphisms).

Since in principle the dimension S can be arbitrarily large, we study the topological properties of hypergraphs using algebraic topology. From a mathematical viewpoint, analysing the topological space generated by the geometric realization of a hypergraph is difficult because it is neither open nor closed in $\mathbb{R}^S$. Specifically, calculating the holes of $|H|$ requires using the notion of singular homology, which is in general very challenging to calculate. Indeed, there only exists a mature theory with efficient algorithms to analyse the topological properties of the special class of hypergraphs known as simplicial complexes—hypergraphs where each hyperedge contains all of its boundaries[64,65]. To the best of our knowledge, there is no efficient algorithm to calculate the homology (Betti numbers) for arbitrary hypergraphs.

To circumvent the above challenge, we built a homology theory for hypergraphs analogous to that for simplicial complexes. We build our homology by imagining we explode the hypergraph into its hyperedges (see illustration of Supplementary Fig. 1b,c). We will now connect back these hyperedges ensuring that we keep the original structure of the hypergraph. To characterize those connections that are allowed, we define the notion of k-simplex:

Definition 1. A k-simplex σ of a hypergraph H is a collection $\sigma=\{h_1,h_2,\cdots,h_{k+1}\}$ of $(k+1)$ hyperedges $h_i\in H$ such that $h_i\subseteq h_{i+1}$ for all i.

We can put together connections between groups of k hyperedges by adding together different k-simplices. This idea is formalized in the notion of a k-chain of the hypergraph. A k-chain is the formal sum

$$c=\sum_q n_q\sigma_q,$$

where each σ_q is a k-simplex of H, and the coefficients satisfy $n_q\in\{0,1\}$ with addition defined modulo 2 (that is, $1+1=0$). In words, a modulo 2 sum means that repeated hyperedges cancel each other out. From the above definition, we can construct the k-th chain group of the hypergraph H denoted by $\langle C_k(H),+\rangle$. This is the group of all k-chains, with '+' denoting addition modulo 2.

The final notion that we need is that of boundary, defined as follows:

Definition 2. The boundary of a k-simplex $\sigma=\{h_1,h_2,\cdots,h_{k+1}\}$ of H is the $(k-1)$-th chain:

$$\partial_k\sigma=\sum_{i=1}^{k+1}\sigma\setminus h_i=\sum_{i=1}^{k+1}\{h_1,h_2,\cdots,h_{k+1}\}\setminus h_i,\ k\geq 0.$$

For completeness, we define $C_{-1}(H)$ as the trivial group $\{0\}$, and the 0-boundary $\partial_0: C_0(H)\to C_{-1}(H)$ as the zero epimorphism. By linearity, we extend the definition of boundary to chains: if $c=\sum_q n_q\sigma_q$ is a k-chain, its boundary is defined as $\partial_k c:=\sum_q n_q\partial_k\sigma_q$. See Supplementary Fig. 1g–i for examples. The so-called fundamental boundary property states that $\partial_k\partial_{k+1}=0$. This property is satisfied using the above definition of boundary. In particular, this property implies that the boundary of a **cycle** (for example, a chain that starts and ends in the same hyperedge) is zero (example 8 in Supplementary Note 2).

The idea is now to use cycles to detect holes. Intuitively, a cycle could be the boundary of some chain, in which case it is a filled cycle. Cycles could also be the boundary of no chain—and in particular, of no hyperedge—implying that they are empty. Consequently, empty cycles characterize the boundary of a hole. Note that different cycles may encircle the same hole. Therefore, to count the number of different holes using cycles, it is necessary to construct an equivalence class of all cycles encircling the same hole. These ideas are made operative by characterizing two key subgroups of the k-th chain group $C_k(H)$ of a hypergraph H:

Definition 3.

(1) The k-th cycle group Z_k is:

$$Z_k=\ker\partial_k:=\{c\in C_k|\partial_k c=\emptyset\}.$$

(2) The k-th boundary group B_k is:

$$B_k=\mathrm{im}\partial_{k+1}:=\{c\in C_k|\exists d\in C_{k+1}\text{ such that }c=\partial_{k+1}d\}.$$

A chain $c\in B_k$ is the boundary of some higher-dimensional chain $d\in C_{k+1}$. Therefore, such a k-chain c is a k-boundary or a bounding cycle. Cycles that are not in B_k are non-bounding cycles. Therefore, bounding cycles bound higher-dimensional chains so that they are filled cycles. Non-bounding cycles are empty cycles.

It turns out that both subgroups Z_k and B_k are normal (because they are abelian), allowing the construction of quotient spaces. To illustrate the implications of this fact, consider the hypergraph of Supplementary Fig. 1b. Let $b\in B_1$ and

$z \in Z_1$ be bounding and non-bounding cycles of Supplementary Fig. 1b and Supplementary Fig. 1c, respectively. Glueing both cycles results in the cycle $z+b$ of Supplementary Fig. 1d. Importantly, note that $z+b$ is homologous to z. Namely, we can retract $z+b$ along the solid triangle $\{[2,4],[2,3,4],[4],[3,4]\}$ to obtain exactly z. In this sense, there exists only **one** cycle of the form $z+B_1$.

The above idea is formalized by the notion of quotient space Z_1/B_1. Extending this idea from $k=1$ to an arbitrary $k \geq 1$ leads to the definition of homology groups:

Definition 4. The k-th homology group $H_k = H_k(H)$ of the hypergraph H is $H_k = Z_k/B_k = \ker \partial_k / \mathrm{im}\partial_{k+1}$.

Thus, if $z_1 = z_2 + B_k$ for some $z_1, z_2 \in Z_k$, we say that the empty cycles z_1 and z_2 are homologous. The number of different empty cycles provide a characterization of the holes of the space. This is formalized by the Betti numbers, with the k-th Betti number corresponding to the number of k-dimensional holes of H:

Definition 5. The k-th Betti number $\beta_k(H)$ of the hypergraph H is $\beta_k(H) = \mathrm{rank}H_k$.

In Theorem 1 of Supplementary Note 3, we prove that the constructed homology exactly captures all holes of the geometric realization of a hypergraph. That is, any hole in the geometric realization of a hypergraph corresponds to an empty cycle family and every empty cycle family corresponds to a hole in the geometric realization of the hypergraph.

Definition of assembly and disassembly holes. Our homology theory for hypergraphs allows us to rigorously define assembly and disassembly holes as follows. Let $H_k(H)$ and $H_k(D)$ denote the k-th homology groups of the assembly and disassembly hypergraphs, respectively. Then, a k-dimensional assembly is a chain $c \in H_k(H)$, and a k-dimensional disassembly hole is a chain $c \in H_k(D)$.

Identifying coexistence holes by computing Betti numbers. To efficiently calculate the Betti numbers of hypergraphs, we showed that they can be calculated by first building an inclusion graph between the hyperedges of the hypergraph and then constructing the associated Vietoris–Rips complex (Proposition 1 of Supplementary Note 3). This result enabled us to compute the Betti numbers and hence the homology of hypergraphs using an efficient algorithm developed for Vietoris–Rips complexes[66]. An accompanying Julia package (with interface to R) provides all of the functionalities introduced in this paper. To efficiently calculate these Betti numbers, we represent the hypergraph as a simple graph between its hyperedges (Supplementary Note 3.3). However, such a graph is not the best representation to easily grasp the presence of holes (see example 10 in Supplementary Note 3). Also, note that analysing one by one each connected component of a hypergraph will considerably reduce the computational cost of calculating Betti numbers.

A structuralist approach to understand the emergence of coexistence holes. To study how coexistence holes emerge in ecological systems with a larger number of species, we adopt a structuralist approach[13,33]. This approach describes a system using its structure (that is, internal conditions that remain fixed, such as species' metabolic capacity to use some resource) and its context (that is, external conditions acting on the system that can change, such as temperature). It follows that, from the universe of all possible assembly and disassembly skeletons, the system's structure determines which ones the system can adopt. From such a set of skeletons that the system can adopt, the context will specify which one is observed. By using this structuralist approach with our formalism, we can quantify the possible assembly and disassembly skeletons that can be observed in a given system (that is, the forms that the ecological system can adopt) as the context changes (that is, as external conditions change).

To make the above approach operative, we leverage on the mathematical tractability of the Lotka–Volterra population dynamics model[52]:

$$\dot{x}_i = x_i \left[r_i + \sum_{j=1}^{S} a_{ij} x_j \right], \; i = 1, \ldots, S.$$

In this model, $x_i(t)$ represents the abundance of species i at time t. The parameter r_i represents the intrinsic growth rate of species i, and a_{ij} represents the effect that species j has on the per-capita growth rate of species i. The Lotka–Volterra model can be interpreted as a first-order approximation of the per-capita growth rate of species. Despite this simplicity, the Lotka–Volterra model successfully explains the dynamics of diverse ecological systems[33,42–44,67–69].

In the idealized governing laws described by the Lotka–Volterra model, internal conditions are phenomenologically captured by the interaction matrix $\boldsymbol{A} = \{a_{ij}\}_{i,j=1}^{S}$, describing how species interact with each other. External conditions are phenomenologically captured by the intrinsic growth rates $\boldsymbol{r} = \{r_i\}_{i=1}^{S}$, describing how species grow in isolation under given abiotic conditions. We use a probability density function $p(\boldsymbol{r})$ to describe how the context can change. Motivated by the fact that the feasibility of a species collection in the Lotka–Volterra model depends on the direction of $\boldsymbol{r}$ and not on its magnitude[32], we assume that $p(\boldsymbol{r})$ is some probability distribution over the unit sphere. Furthermore, because all species survive in isolation in the empirical microbial systems that we analyse, we consider that $p(\boldsymbol{r})$ is uniform over the positive section of the unit sphere. Assuming a uniform distribution conforms with the ergodicity hypothesis in dynamical systems theory[33]. Note that, in a general case, abiotic (that is, external) conditions may affect species interactions[51].

Constructing assembly hypergraphs for the Lotka–Volterra model. Using the Lotka–Volterra formalism, we constructed the associated assembly hypergraph H by computing whether each of the $2^S - 1$ different species collections coexists or not under model parameters $(\boldsymbol{A}, \boldsymbol{r})$. We defined coexistence following Jansen's permanence criterion[35,70] (Supplementary Note 4), which for the Lotka–Volterra model actually implies robust permanence[71].

To analyse Lotka–Volterra systems with random parameters, we constructed the interaction matrix $\boldsymbol{A} = \{a_{ij}\}_{i,j=1}^{S}$ by sampling interspecific interactions $i \neq j$ as $a_{ij} \sim \mathrm{Bernoulli}(C_A)\mathrm{Normal}(0, \sigma_A)$ and fixing the intra-specific interactions as $a_{ii} = -1$. Here, the connectance $C_A \in [0,1]$ describes the probability that two species interact, and $\sigma_A \geq 0$ describes the typical interaction strength.

Empirical ecological systems. To study whether coexistence holes occur in empirical ecological systems, we analysed five experimental microbial communities: a system of $S=4$ protozoa built by Vandermeer[42]; a system of $S=8$ soil bacterial species built by Friedman et al.[16]; a system of $S=11$ human gut bacterial species built by Stein et al.[43]; a system of $S=12$ human gut bacterial species built by Venturelli et al.[44]; and an in vivo system of $S=14$ gut bacterial species built for the MDSINE project[45] (see details in Supplementary Note 6). In each of these studies, a Lotka–Volterra model was parameterized and systematically validated using high-resolution experimental data, resulting in empirical estimations of the parameters $(\boldsymbol{A}, \boldsymbol{r})$. As mentioned before, in all of these experimental systems, species can survive in isolation—motivating our assumption that $p(\boldsymbol{r})$ is uniform over the positive section of the unit sphere. We choose these systems because the collected experimental data contain one or more measurements with the absolute abundance of species, which is a necessary condition for an adequate parameter inference[72].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All of the data analysed in this study are publicly available.

Code availability

The code supporting the results is archived in the GitHub repository at https://syntheticdynamics.github.io/CoexistenceHoles.jl.

Received: 24 September 2020; Accepted: 7 April 2021;
Published online: 27 May 2021

References

1. Fukami, T. Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annu. Rev. Ecol. Evol. Syst.* **46**, 1–23 (2015).
2. Tylianakis, J. M., Martínez-García, L. B., Richardson, S. J., Peltzer, D. A. & Dickie, I. A. Symmetric assembly and disassembly processes in an ecological network. *Ecol. Lett.* **21**, 896–904 (2018).
3. Chase, J. M., Blowes, S. A., Knight, T. M., Gerstner, K. & May, F. Ecosystem decay exacerbates biodiversity loss with habitat loss. *Nature* **584**, 238–243 (2020).
4. Vellend, M. *The Theory of Ecological Communities (MPB-57)* (Princeton Univ. Press, 2016).
5. Hutchinson, G. E. Homage to Santa Rosalia or why are there so many kinds of animals? *Am. Nat.* **93**, 145–159 (1959).
6. Tilman, D. *Resource Competition and Community Structure* (Princeton Univ. Press, 1982).
7. Barbier, M., Arnoldi, J.-F., Bunin, G. & Loreau, M. Generic assembly patterns in complex ecological communities. *Proc. Natl Acad. Sci. USA* **115**, 2156–2161 (2018).
8. Serván, C. A., Capitán, J. A., Grilli, J., Morrison, K. E. & Allesina, S. Coexistence of many species in random ecosystems. *Nat. Ecol. Evol.* **2**, 1237–1242 (2018).
9. MacArthur, R. Species packing and competitive equilibrium for many species. *Theor. Popul. Biol.* **1**, 1–11 (1970).
10. Medeiros, L. P., Boege, K., del Val, E., Zaldivar-Riverón, A. & Saavedra, S. Observed ecological communities are formed by species combinations that are among the most likely to persist under changing environments. *Am. Nat.* https://doi.org/10.1086/711663 (2020).
11. Barabás, G., D'Andrea, R. & Stump, S. M. Chesson's coexistence theory. *Ecol. Monogr.* **88**, 277–303 (2018).
12. Grainger, T. N. & Gilbert, J. M. L. B. The invasion criterion: a common currency for ecological research. *Trends Ecol. Evol.* **34**, 925–935 (2019).
13. Alberch, P. The logic of monsters: evidence for internal constraint in development and evolution. *Geobios* **22**, 21–57 (1989).
14. Clements, F. E. Nature and structure of the climax. *J. Ecol.* **24**, 252–284 (1936).
15. Odum, E. P. & Barrett, G. W. *Fundamentals of Ecology* 5th edn (Thomson Brooks/Cole, 2005).

16. Friedman, J., Higgins, L. M. & Gore, J. Community structure follows simple assembly rules in microbial microcosms. *Nat. Ecol. Evol.* **1**, 0109 (2017).
17. Chesson, P. Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.* **31**, 343–366 (2000).
18. Drake, J. A. Community-assembly mechanics and the structure of an experimental species ensemble. *Am. Nat.* **137**, 1–26 (1991).
19. Warren, P. H., Law, R. & Weatherby, A. J. Mapping the assembly of protist communities in microcosms. *Ecology* **84**, 1001–1011 (2003).
20. Schreiber, S. J. & Rittenhouse, S. From simple rules to cycling in community assembly. *Oikos* **105**, 349–358 (2004).
21. Chase, J. M. & Leibold, M. A. *Ecological Niches: Linking Classical and Contemporary Approaches* (Univ. Chicago Press, 2003).
22. Kraft, N. J. et al. Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.* **29**, 592–599 (2015).
23. Moore, R., Robinson, W., Lovette, I. & Robinson, T. Experimental evidence for extreme dispersal limitation in tropical forest birds. *Ecol. Lett.* **11**, 960–968 (2008).
24. Maherali, H. & Klironomos, J. N. Influence of phylogeny on fungal community assembly and ecosystem functioning. *Science* **316**, 1746–1748 (2007).
25. Serván, C. & Allesina, S. Tractable models of ecological assembly. *Ecol. Lett.* **24**, 1029–1037 (2021).
26. Rosindell, J., Hubbell, S. P. & Etienne, R. S. The unified neutral theory of biodiversity and biogeography at age ten. *Trends Ecol. Evol.* **26**, 340–348 (2011).
27. Case, T. J. Surprising behavior from a familiar model and implications for competition theory. *Am. Nat.* **146**, 961–966 (1995).
28. Saavedra, S. et al. A structural approach for understanding multispecies coexistence. *Ecol. Monogr.* **87**, 470–486 (2017).
29. Tilman, D. Resources: a graphical-mechanistic approach to competition and predation. *Am. Nat.* **116**, 362–393 (1980).
30. May, R. M. & Leonard, W. J. Nonlinear aspects of competition between three species. *SIAM J. Appl. Math.* **29**, 243–253 (1975).
31. Dean, A. M. A simple model of mutualism. *Am. Nat.* **121**, 409–417 (1983).
32. Song, C., Ahn, S. V., Rohr, R. P. & Saavedra, S. Towards a probabilistic understanding about the context-dependency of species interactions. *Trends Ecol. Evol.* **35**, 384–396 (2020).
33. Saavedra, S., Medeiros, L. P. & AlAdwani, M. Structural forecasting of species persistence under changing environments. *Ecol. Lett.* https://doi.org/10.1111/ele.13582 (2020).
34. Law, R. & Blackford, J. C. Self-assembling food webs: a global viewpoint of coexistence of species in Lotka–Volterra communities. *Ecology* **73**, 567–578 (1992).
35. Sigmuiud, K. Darwin's 'circles of complexity': assembling ecological communities. *Complexity* **1**, 40–44 (1995).
36. Law, R. & Morton, R. D. Permanence and the assembly of ecological communities. *Ecology* **77**, 762–775 (1996).
37. Wilson, J. B., Spijkerman, E. & Huisman, J. Is there really insufficient support for Tilman's *R** concept? A comment on Miller et al. *Am. Nat.* **169**, 700–706 (2007).
38. May, R. M. Simple mathematical models with very complicated dynamics. *Nature* **261**, 459–467 (1976).
39. Cenci, S., Song, C. & Saavedra, S. Rethinking the importance of the structure of ecological networks under an environment-dependent framework. *Ecol. Evol.* **8**, 6852–6859 (2018).
40. O'Dwyer, J. P. Whence Lotka-Volterra? *Theor. Ecol.* **11**, 441–452 (2018).
41. Levine, J. M., Bascompte, J., Adler, P. B. & Allesina, S.Beyond pairwise mechanisms of species coexistence in complex communities. *Nature* **546**, 56–64 (2017).
42. Vandermeer, J. H. The competitive structure of communities: an experimental approach with protozoa. *Ecology* **50**, 362–371 (1969).
43. Stein, R. R. et al. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS Comput. Biol.* **9**, e1003388 (2013).
44. Venturelli, O. S. et al. Deciphering microbial interactions in synthetic human gut microbiome communities. *Mol. Syst. Biol.* **14**, e8157 (2018).
45. Bucci, V. et al. MDSINE: Microbial Dynamical Systems Inference Engine for microbiome time-series analyses. *Genome Biol.* **17**, 121 (2016).
46. Turelli, M. A reexamination of stability in randomly varying versus deterministic environments with comments on the stochastic theory of limiting similarity. *Theor. Popul. Biol.* **13**, 244–267 (1978).
47. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, 2019).
48. Allesina, S. & Tang, S. The stability–complexity relationship at age 40: a random matrix perspective. *Popul. Ecol.* **57**, 63–75 (2015).
49. Allesina, S. & Tang, S. Stability criteria for complex ecosystems. *Nature* **483**, 205–208 (2012).
50. Grilli, J., Rogers, T. & Allesina, S. Modularity and stability in ecological communities. *Nat. Commun.* **7**, 12031 (2016).
51. Hoek, T. A. et al. Resource availability modulates the cooperative and competitive nature of a microbial cross-feeding mutualism. *PLoS Biol.* **14**, e1002540 (2016).
52. Case, T. J. *An Illustrated Guide to Theoretical Ecology* (Oxford Univ. Press, 2000).
53. Freedman, H. & So, J.-H. Global stability and persistence of simple food chains. *Math. Biosci.* **76**, 69–86 (1985).
54. Posfai, A., Taillefumier, T. & Wingreen, N. S. Metabolic trade-offs promote diversity in a model ecosystem. *Phys. Rev. Lett.* **118**, 028103 (2017).
55. Gould, A. L. et al. Microbiome interactions shape host fitness. *Proc. Natl Acad. Sci. USA* **115**, E11951–E11960 (2018).
56. Kehe, J. et al. Massively parallel screening of synthetic microbial communities. *Proc. Natl Acad. Sci. USA* **116**, 12804–12809 (2019).
57. Xiao, Y. et al. Mapping the ecological networks of microbial communities. *Nat. Commun.* **8**, 2042 (2017).
58. AlAdwani, M. & Saavedra, S. Is the addition of higher-order interactions in ecological models increasing the understanding of ecological dynamics? *Math. Biosci.* **315**, 108222 (2019).
59. Weibel, C. A. in *History of Topology* (ed. James, I.) 797–836 (North-Holland, 1999).
60. Carlsson, G. Topology and data. *Bull. Am. Math. Soc.* **46**, 255–308 (2009).
61. Rabadán, R. & Blumberg, A. J. *Topological Data Analysis for Genomics and Evolution: Topology in Biology* (Cambridge Univ. Press, 2019).
62. Sizemore, A. E., Phillips-Cremins, J. E., Ghrist, R. & Bassett, D. S. The importance of the whole: topological data analysis for the network neuroscientist. *Netw. Neurosci.* **3**, 656–673 (2019).
63. Sugihara, G. Graph theory, homology and food webs. In *Proc. Symposia in Applied Mathematics* **30**, 83–101 (American Mathematical Society, 1984).
64. Singh, G., Mémoli, F. & Carlsson, G. E. Topological methods for the analysis of high dimensional data sets and 3D object recognition. In *Symposium on Point Based Graphics* 91–100 (The Eurographics Association, 2007).
65. Giusti, C., Ghrist, R. & Bassett, D. S. Two's company, three (or more) is a simplex. *J. Comput. Neurosci.* **41**, 1–14 (2016).
66. Bauer, U. Ripser: efficient computation of Vietoris–Rips persistence barcodes. Preprint at https://arxiv.org/abs/1908.02518 (2019).
67. Fort, H. On predicting species yields in multispecies communities: quantifying the accuracy of the linear Lotka–Volterra generalized model. *Ecol. Model.* **387**, 154–162 (2018).
68. Halty, V., Valdés, M., Tejera, M., Picasso, V. & Fort, H. Modeling plant interspecific interactions from experiments with perennial crop mixtures to predict optimal combinations. *Ecol. Appl.* **27**, 2277–2289 (2017).
69. Tabi, A. et al. Species multidimensional effects explain idiosyncratic responses of communities to environmental change. *Nat. Ecol. Evol.* **4**, 1036–1043 (2020).
70. Jansen, W. A permanence theorem for replicator and Lotka–Volterra systems. *J. Math. Biol.* **25**, 411–422 (1987).
71. Schreiber, S. J. Criteria for C^r robust permanence. *J. Differ. Equ.* **162**, 400–426 (2000).
72. Angulo, M. T., Moreno, J. A., Lippner, G., Barabási, A.-L. & Liu, Y.-Y. Fundamental limitations of network reconstruction from temporal data. *J. R. Soc. Interface* **14**, 20160966 (2017).

Acknowledgements

We gratefully acknowledge financial support from CONACyT grant number A1-S-13909 (M.T.A.) and NSF grant number DEB-2024349 (S.S.).

Author contributions

M.T.A. conceived of the idea of coexistence holes. M.T.A., C.S. and S.S. designed and realized the study. M.T.A., L.M. and A.K. performed the theoretical analysis. M.T.A., C.S. and S.S. wrote the manuscript. A.K. and M.T.A. wrote the software package to identify coexistence holes. All authors revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-021-01462-8.

Correspondence and requests for materials should be addressed to M.T.A. or S.S.

Peer review information *Nature Ecology & Evolution* thanks Stefano Allesina, Andrew Letten and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Marco Tulio Angulo and Serguei Saavedra

Last updated by author(s): Mar 12, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No new data was generated for this study.
Data analysis	Analysis was made using Mathematica 12 and Julia v.1.3.1. Code supporting the results is archived in the Github repository https://syntheticdynamics.github.io/CoexistenceHoles.jl

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All the experimental datasets analyzed in this study are publicly available.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	No experiments were carried out for this work.
Research sample	N/A
Sampling strategy	N/A
Data collection	N/A
Timing and spatial scale	N/A
Data exclusions	N/A
Reproducibility	N/A
Randomization	N/A
Blinding	N/A

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging