

# Top-down and bottom-up cohesiveness in microbial community coalescence

Juan Diaz-Colunga<sup>1\*</sup>, Nanxi Lu<sup>1\*</sup>, Alicia Sanchez-Gorostiaga<sup>1,2\*</sup>, Chang-Yu Chang<sup>1</sup>,  
Helen S. Cai<sup>1</sup>, Joshua E. Goldford<sup>3</sup>, Mikhail Tikhonov<sup>4</sup>, and Álvaro Sánchez<sup>1✉</sup>

<sup>1</sup>Department of Ecology & Evolutionary Biology and Microbial Sciences Institute, Yale University, New Haven, CT, USA

<sup>2</sup>Department of Microbial Biotechnology, Centro Nacional de Biotecnología (CNB-CSIC), Cantoblanco, Madrid, Spain

<sup>3</sup>Physics of Living Systems, Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>4</sup>Department of Physics, Center for Science & Engineering of Living Systems, Washington University in St. Louis, St. Louis, MO, USA

✉alvaro.sanchez@yale.edu

\*These authors contributed equally

## Abstract

Microbial communities frequently invade one another as a whole, a phenomenon known as *community coalescence*. Despite its potential importance for the assembly, dynamics and stability of microbial consortia, as well as its prospective utility for microbiome engineering, our understanding of the processes that govern it is still very limited. Theory has suggested that microbial communities may exhibit cohesiveness in the face of invasions emerging from collective metabolic interactions across microbes and their environment. This cohesiveness may lead to correlated invasional outcomes, where the fate of a given taxon is determined by that of other members of its community –a hypothesis known as *ecological co-selection*. Here, we have performed over a hundred invasion and coalescence experiments with microbial communities of various origins assembled in two different synthetic environments. We show that the dominant members of the primary communities can recruit their rarer partners during coalescence (top-down co-selection) and also be recruited by them (bottom-up co-selection). With the aid of a consumer-resource model, we found that the emergence of top-down or bottom-up cohesiveness is modulated by the structure of the underlying cross-feeding networks that sustain the coalesced communities. The model also predicts that these two forms of ecological co-selection cannot co-occur under our conditions, and we have experimentally confirmed that one can only be strong when the other is weak. Our results provide direct evidence that collective invasions can be generically expected to produce ecological co-selection as a result of cross-feeding interactions at the community level.

## Introduction

Microbial communities often invade one another. This has been observed, for instance, in river courses where terrestrial microbes mix with aquatic microorganisms [1–3], or in soil communities being invaded as a result of tillage and outplanting [4–6] or by aerially dispersed bacteria and fungi [7]. Gut microbiomes can invade external communities through the host's secretions [8], and the skin microbiota is also subject to invasions when it makes contact with environmental sources of microbes [9].

The phenomenon by which entire microbiomes invade one another has been termed *community coalescence* [10]. Ecologists have long contemplated the idea that interactions between multiple co-invading species can produce correlated invasional outcomes [10–20]. However, and in spite of its clear potential importance, the role of coalescence in microbiome assembly is only beginning to be addressed and little is known about the mechanisms that govern it and its potential implications or applications [21]. Early mathematical models of community-community invasions [11, 22] as well as more recent work [23–26] suggest that high-order invasion effects are common during community coalescence. Communities that have a previous history of coexistence may exhibit an emergent “cohesiveness” which produces correlated invasional outcomes among species from the same community [17, 27]. The situation where ecological partners in the invading community recruit each other into the final coalesced community has been called *ecological co-selection* [27, 28].

The mechanisms of ecological co-selection during community coalescence are still poorly understood. Do a few key species recruit everyone else, or are collective interactions among all species (including the rarer members of the community) relevant for coalescence outcomes? While it is reasonable to expect species with larger population sizes to have a proportionally oversized effect, natural communities tend to be highly diverse [29] and the

49 role played by the less abundant community members has long been subject to debate [30]. Laboratory cultures  
50 have also been found to contain uneven distributions of multiple strains that feed off the metabolic secretions of  
51 the dominant species [31, 32]. The fate of these sub-dominant taxa may be dependent on the invasion success of  
52 their dominant species, or, alternatively, the dominant itself may owe its ability to invade the resident community  
53 (at least in part) to cross-feeding or other forms of facilitation from the rarer members of its native, invasive com-  
54 munity. We refer to these two opposite scenarios as the “top-down” (i.e. when the dominant invader co-selects  
55 other sub-dominant taxa into the final community during coalescence) or “bottom-up” (i.e. when it is co-selected  
56 by the cohort of rare species in its initial community) forms of community cohesiveness, respectively. Either of  
57 these forms of co-selection could, in principle, be positive (i.e. recruitment) or negative (antagonism), as illus-  
58 trated in [Figure 1e](#), and it is even plausible that both top-down and bottom-up co-selection may be present at the  
59 same time, i.e., that the dominant species and its cohort of rarer species co-select one another during microbial  
60 community coalescence. Which of these potential scenarios are typically found in nature? Previous theoretical  
61 and computational studies suggest that the answer is determined by the type and strength of the interactions of  
62 the community members with one another and with the environment [23, 25, 26], but addressing this question has  
63 been experimentally challenging [27, 28].

64 In previous work, we have shown that diverse, multi-species enrichment communities self-assemble *ex situ*  
65 in synthetic minimal environments with a single supplied limiting carbon source under serial growth-dilution  
66 cycles [32] ([Figure 1a-b](#)). After serially passaging our communities over 7-8 growth-dilution cycles, these multi-  
67 species communities reach a state of equilibrium, where stable coexistence is sustained by dense cross-feeding  
68 facilitation networks [32, 33]. In addition, and similar to natural communities, species abundance distributions in  
69 these enrichment communities are generally long-tailed and uneven ([Figure 1d](#) and [Figure S1](#)), with the dominant  
70 (most abundant) species typically comprising most of the biomass (median = 46%, [Figure S1](#)). Because these  
71 communities are easy to manipulate and grow in high throughput, they represent good test cases to investigate  
72 ecological co-selection during community coalescence. Here we focus on the dominants and ask whether they  
73 can co-select or be co-selected by the sub-dominant species in their communities (henceforth referred to as their  
74 cohorts, [Figure 1c](#)).

75 Our experimental results indicate that co-selection is positive under our conditions: the success of the co-  
76 horts after community coalescence were positively correlated with that of their dominants in pairwise competition,  
77 and we did not find a single community coalescence experiment where the cohorts had a negative effect on their  
78 dominant taxa. We also found that in the community coalescence experiments where bottom-up co-selection  
79 was observed, top-down co-selection was absent. Conversely, in the set of community coalescence events where  
80 bottom-up co-selection was absent, we observed clear signatures of top-down co-selection. Our experiments clus-  
81 tered in either one of these two limiting scenarios, while others (e.g. both top-down and bottom-up co-selection  
82 being present) were conspicuously absent. To rationalize these findings, we turned to a Microbial Consumer-  
83 Resource Model (MicroCRM) [32, 34, 35] that has been previously found to be able to capture the dynamics of  
84 microbial communities dominated by metabolic interactions, as is the case for the ones assembled in our exper-  
85 imental conditions [32, 33]. We show that the trends we observed in our experiments are all reproduced with  
86 minimal model assumptions, and that the recurrence of top-down and bottom-up co-selection is determined by the  
87 configuration of the cross-feeding networks in the MicroCRM. The good agreement between the MicroCRM and  
88 our experiments emphasizes the usefulness of this model to explain and potentially predict the fate of microbial  
89 communities during coalescence.

## 90 Results & Discussion

91 We collected eight natural microbiomes from different soil and plant environmental samples ([Figure 1a](#)) and used  
92 them to inoculate eight identical habitats containing minimal media with either glutamine or citrate as the only sup-  
93 plied carbon source. We chose these two carbon sources because they are metabolized through different pathways  
94 in bacteria [36, 37], and we hypothesize that communities assembled in either resource will be supported by cross-  
95 feeding networks of distinct sets of metabolites [32, 33], thus leading to potentially variable degrees of community  
96 cohesiveness and coalescence outcomes [20, 23, 24, 26]. After inoculation, all communities were serially passaged  
97 for 12 transfers (84 generations), with an incubation time of 48 hours and a dilution factor of 1:100. ([Figure 1b](#),  
98 [Methods: Stabilization of environmental communities in simple synthetic environments](#)). In previous work we  
99 have shown that under these conditions, 12 transfers allow communities to approach a state of “generational equi-  
100 librium”, where the community composition at the end of one batch incubation will be the same as in consecutive  
101 incubations. We isolated the dominant species of every community ([Methods: Isolation of dominant species](#)) and  
102 identified them by Sanger-sequencing their 16S rRNA gene ([Methods: Determination of community composition](#)  
103 [by 16S sequencing](#)), which correctly matched the dominant Exact Sequence Variant (ESV) [38, 39] found through  
104 community-level 16S Illumina sequencing ([Figure S1](#)). These dominants remained at high frequency after seven

105 additional transfers with the exception of two of the citrate communities and one of the glutamine communities  
106 (where the dominants were presumably a transiently dominating species) that were excluded from further analysis  
107 ([Figure S1](#)). Similarly, pairs of communities where the dominants shared a same 16S sequence and had similar  
108 colony morphology were excluded ([Figure S1](#)).

## 109 Top-down ecological co-selection

110 One form of cohesiveness may arise when the sub-dominant members of the community depend on the dominant  
111 species. This can occur, for instance, when the dominant provides resources (or stressors) that select for the sub-  
112 dominant taxa ([Figure 1e](#), left panels). If communities being coalesced exhibit positive cohesiveness from the  
113 top-down, the fate of the sub-dominant community members will be tied to their dominant: if a dominant gets  
114 excluded, the cohort will be likely to fall with it, whereas if the dominant thrives after coalescence, its cohort will  
115 be likely to follow suit. In this scenario, we would expect the outcome of community coalescence to be predicted  
116 by which of the two dominants is most competitive in pairwise competition. Likewise, competition between  
117 dominants should be affected only weakly by the presence or absence of sub-dominant species, which would play a  
118 passive role under top-down co-selection. To test this hypothesis, we performed all pairwise competitions between  
119 dominant species in either the glutamine or citrate environments by mixing them 1:1 on their native media and  
120 propagating the cultures for seven serial transfers, roughly 42 generations ([Methods: Coalescence, competition and](#)  
121 [invasion experiments](#)). We then carried out all possible pairwise community coalescence experiments by mixing  
122 equal volumes of the communities and propagating the resulting cultures for seven extra transfers ([Figure 1f](#)). Note  
123 that this setup is symmetric, but for the purposes of later analysis, it is convenient to think of one community (the  
124 *resident*) being invaded by the other (the *invasive*). The frequencies of all species in both community-community  
125 and dominant-dominant competitions were determined by 16S Illumina sequencing ([Methods: Determination of](#)  
126 [community composition by 16S sequencing](#)).

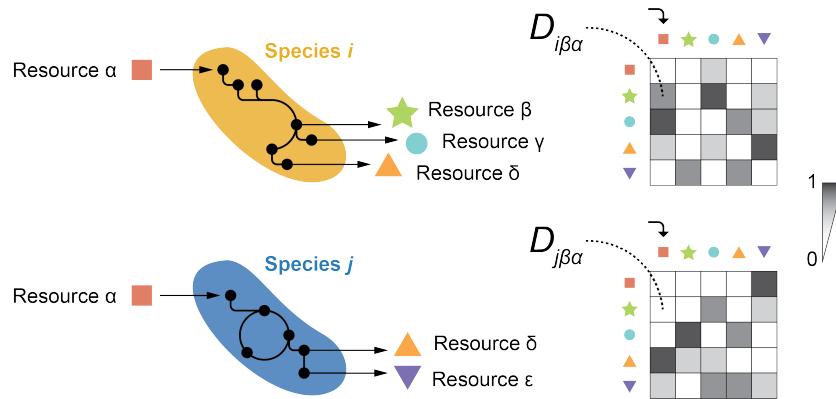
127 To test the effects of top-down co-selection at the community level, we quantified the distances between the  
128 invasive and coalesced communities using the relative Bray-Curtis similarity ([Methods: Metrics of community](#)  
129 [distance](#)) and compared them to the outcomes of the pairwise competitions between dominants alone ([Figure 2a](#)).  
130 We noticed a difference between communities assembled in the glutamine and citrate environments: for the latter,  
131 the structure of the coalesced communities tends to be strongly dictated by the result of the dominant-dominant  
132 competition ([Figure 2b](#) right panel,  $R^2 = 0.57$ ,  $p < 10^{-4}$ ,  $N = 22$ ). For the former, the pairwise competitive ability  
133 of an invasive dominant is only weakly predictive of the performance of the invasive community in coalescence  
134 ([Figure 2b](#) left panel,  $R^2 = 0.15$ ,  $p < 0.05$ ,  $N = 34$ ). In both cases, the data is consistent with positive, rather  
135 than negative top-down co-selection ([Figure 2a](#)). Alternative quantification of the distance between the coalesced  
136 and invasive communities yielded similar results, with weaker effects when the metric used accounts only for the  
137 presence/absence of specific species and not for their relative abundance ([Figure S2](#)). All these metrics include  
138 the presence of the dominant species themselves. To better disentangle the effect that these dominants have on the  
139 other members of their communities, we repeated the analysis this time excluding the dominant species from the  
140 compositional data, finding that our results still hold ([Figure S3](#)). We then examined whether, as predicted by the  
141 top-down cohesiveness hypothesis, the cohorts would play a passive role on the competition between dominant  
142 species. We found that, for communities assembled in the citrate environments, the relative frequency of a domi-  
143 nant against another in head-to-head pairwise competition is highly predictive of its relative frequency against that  
144 same other dominant when the cohorts are present too, i.e. during community coalescence ([Figure 2c](#) blue dots,  
145  $R^2 = 0.83$ ,  $p < 10^{-8}$ ,  $N = 22$ ). This is not the case for the glutamine communities ([Figure 2c](#) red dots,  $R^2 = 0.04$ ,  
146  $p > 0.05$ ,  $N = 34$ ). This suggests that, in the glutamine environments, head-to-head competition between domi-  
147 nants is strongly affected by interactions between those dominants and the rarer taxa of the communities. On the  
148 other hand, the cohorts of sub-dominant taxa seem to play a more passive role in the citrate environments. To-  
149 gether, these observations indicate that communities stabilized with citrate as the primary supplied resource display  
150 a strong degree of top-down cohesiveness, with the fates of the sub-dominant species determined to a large extent  
151 by dominant-dominant pairwise competition. This competition is, in turn, only weakly affected by the presence  
152 of the cohorts. For glutamine communities, although some level of top-down co-selection is consistent with our  
153 data, the cohorts do not appear to just be passively responding to their dominants but rather playing an active role  
154 in community coalescence.

155 To investigate the determinants of top-down co-selection and the factors modulating its strength, we ran a set  
156 of simulations of community coalescence. We used a Microbial Consumer-Resource Model (MicroCRM) [32, 34]  
157 as implemented in the Community Simulator package for Python [35] ([Box 1](#)). We chose this modeling framework  
158 because communities assembled under our experimental conditions (natural microbiomes re-assembled into multi-  
159 species communities through serial growth-dilution cycles in synthetic minimal media with a single carbon source)  
160 have been shown to be sustained by dense metabolic cross-feeding networks [32, 33] for which the MicroCRM  
161 provides a good description. We and others have previously found a strong concordance between the behavior of

162 laboratory and natural microbial communities and the behavior of the MicroCRM [32, 34, 35, 40, 41]. To repro-  
163 duce our experimental protocol *in silico*, we first generated a library of resources and two non-overlapping pools  
164 of species. Each pool was used to seed a collection of 100 invasive and 100 resident communities respectively, by  
165 randomly choosing 50 species and allowing them to stabilize through 20 growth-dilution cycles. We then mixed  
166 these stable communities in pairs to simulate our coalescence and dominant-dominant competition experiments  
167 ([Methods: Simulations](#)). We found that the MicroCRM simulations naturally exhibit the observed correlation  
168 between the head-to-head pairwise competition of dominants and the outcome of community coalescence ([Fig-](#)  
169 [ure 2d](#)), further supporting the idea that top-down ecological co-selection consistently emerges from metabolic  
170 interactions across species. Moreover, we found that top-down co-selection is observed under a wide range of  
171 different simulation conditions and cross-feeding networks ([Figure S5](#)), indicating that it is a robust phenomenon.

172 **Box 1: A Microbial Consumer-Resource Model for community coalescence**

173 The Microbial Consumer-Resource Model (MicroCRM) [32, 34, 35] is a modeling framework based on the  
 174 classic MacArthur's consumer resource model [42]. It encodes the dynamics of a system with  $S$  species and  
 175  $M$  resources in terms of a consumer preference matrix  $\mathbf{c}$  and a metabolic matrix  $\mathbf{D}$ , with an additional set of  
 176 parameters controlling the species maintenance costs ( $m_i$  for species  $i$ ), the resource energy densities ( $w_\alpha$  for  
 177 resource  $\alpha$ ), the energy to growth rate conversion factor ( $g_i$  for species  $i$ ) and the leakage fraction, i.e. the  
 178 amount of energy lost as byproducts when a resource is consumed ( $l_\alpha$  for resource  $\alpha$ ). The element  $c_{ia}$  of the  
 179 consumer preference matrix represents the uptake rate of resource  $\alpha$  by species  $i$  (although the relationship  
 180 between  $c_{ia}$  and the uptake rate can be more complex in modeling scenarios that are not considered here, see  
 181 [32, 34, 35]). Experimental evidence suggests that individual species can secrete different sets of metabolites  
 182 to the environment when growing on a same primary resource [33, 43, 44]. Thus, we define  $\mathbf{D}$  as a three-  
 183 dimensional matrix where the element  $D_{i\beta\alpha}$  represents the energy flux in the form of resource  $\beta$  that is secreted  
 184 by species  $i$  when it metabolizes resource  $\alpha$ . Note that  $D_{i\beta\alpha}$  need not be equal to  $D_{j\beta\alpha}$  if  $i \neq j$  (see illustration  
 185 below).



187 The following equations describe the kinetics of the abundances of the  $i$ -th species (denoted as  $N_i$ ) and  
 188 the  $\alpha$ -th resource (denoted as  $R_\alpha$ ):

$$189 \frac{dN_i}{dt} = g_i N_i \left[ \sum_{\alpha} (1 - l_{\alpha}) w_{\alpha} c_{i\alpha} R_{\alpha} - m_i \right] \quad (1)$$

$$190 \frac{dR_{\alpha}}{dt} = - \sum_j N_j c_{j\alpha} R_{\alpha} + \sum_j \sum_{\beta} N_j c_{j\beta} R_{\beta} \left[ l_{\beta} D_{j\beta\alpha} \frac{w_{\beta}}{w_{\alpha}} \right] \quad (2)$$

191 These equations can take slightly different forms in certain cases, e.g. if the primary resource is supplied  
 192 continuously instead of at the beginning of each growth cycle [34, 35]. They represent a good approxima-  
 193 tion for the community dynamics between consecutive serial dilutions in our setup. Here, we assembled *in*  
 194 *silico* communities by randomly sampling a set of species from a pool, then integrating equations 1 and 2,  
 195 diluting the final abundances, replenishing the primary resource, and repeating the process until generational  
 196 equilibrium was achieved (Methods: Simulations). Coalescence simulations were carried out following the  
 197 same logic, this time seeding the coalesced communities by mixing the invasive and resident ones instead of  
 198 sampling from a species pool.

205 **Bottom-up co-selection during community coalescence**

206 Our data indicates that the primary resource supplied to the communities can modulate the effect that the cohorts  
 207 have in the dominants pairwise competition ([Figure 2c](#)) and the strength of top-down co-selection ([Figure 2b](#)). The  
 208 fact that our model captures these trends suggests that this might be a result of the metabolic interactions between  
 209 community members, including the rarer taxa. To investigate the potential role of the cohorts in coalescence,  
 210 i.e. whether the dominants may be co-selected for or against by them ([Figure 1e](#), right panels), we ran a new  
 211 set of simulations this time invading resident communities with the dominants alone ([Methods: Simulations](#)). We  
 212 compared the invasion success of the dominants in isolation with respect to our previous simulations where they  
 213 invaded accompanied by their cohorts. The invasion success of the dominants was quantified by their relative  
 214 abundance in the final stabilized communities ([Figure 3a](#)). Whenever positive bottom-up ecological co-selection  
 215 is strong, we expect to see dominants reaching higher invasion success with their cohorts than by themselves, with  
 216 the strongest instances occurring when dominants are unable to invade on their own but reach high densities when  
 217 invading together with their cohorts ([Figure 3a](#), green shaded region). Alternatively, a high degree of bottom-  
 218 up antagonism would result in dominants invading more effectively alone than in the presence of their cohorts  
 219 ([Figure 3a](#), red shaded region). Finally, if bottom-up co-selection is weak, we would see a similar invasion success  
 220 regardless of the presence or absence of the cohort ([Figure 3a](#), gray shaded region).

221 In our simulations of the MicroCRM, we found no instances of bottom-up antagonism but multiple such in-  
 222 stances of positive bottom-up co-selection as well as of no (or weak) bottom-up coselection ([Figure 3b](#)). Many  
 223 dominant members of our *in silico* communities could not invade another community on their own (or could only  
 224 do so at very low final relative abundances, below 0.1) but were able to reach high frequencies when they were  
 225 accompanied by their cohorts in community coalescence. Notably, this behavior was contingent on the metabolic  
 226 matrix being sparse and different for different families (i.e.  $D_{j\beta\alpha}$  need not be equal to  $D_{j\beta\alpha}$  for any two species  $i$  and  
 227  $j$  and resources  $\alpha$  and  $\beta$ , see [Box 1](#)), as experiments suggest is the case in natural settings [33, 43, 44] ([Figure S4](#)).  
 228 Thus, theory indicates that positive bottom-up co-selection is frequent and potentially very strong, while negative  
 229 bottom-up co-selection is far more uncommon. Interestingly, our simulations suggest that strong bottom-up co-  
 230 selection should only be observed in communities where top-down co-selection is weak, while top down co-selection  
 231 is only seen when bottom-up co-selection is weak. To better illustrate this prediction, we divided our simulations  
 232 into two subsets: the first one was comprised of the instances where positive bottom-up co-selection was strong  
 233 (i.e. dots in the green shaded region of [Figure 3b](#)), the second set included all other cases (dots near the diagonal of  
 234 [Figure 3b](#)). We reexamined our original simulations and found that when bottom-up positive co-selection is  
 235 strong, the pairwise competition of dominants is not predictive of coalescence outcomes ([Figure 3c](#), left panel;  
 236  $R^2 = 0.00, p > 0.05, N = 21$ ) indicating that top-down co-selection is weak. At the same time, when considering  
 237 only those coalesced communities in the diagonal of [Figure 3b](#) (where bottom-up co-selection is weak), our model  
 238 predicts that the fates of the sub-dominant community members after coalescence are more strongly determined  
 239 by the head-to-head competition between dominants in isolation ( $R^2 = 0.34, p < 10^{-6}, N = 79$  for instances where  
 240 bottom-up co-selection is weak, [Figure 3c](#) right panel;  $R^2 = 0.22, p < 10^{-5}, N = 100$  when all instances are  
 241 considered, [Figure 2d](#)).

242 We then asked whether this behavior predicted by the model was also observed in our experimental commu-  
 243 nities. To address this question, we carried out a new round of experiments where we invaded the resident com-  
 244 munities with the invasive dominants alone ([Methods: Coalescence, competition and invasion experiments](#)). After  
 245 stabilization ([Methods: Stabilization of environmental communities in simple synthetic environments](#)), we quanti-  
 246 fied species abundance through 16S Illumina sequencing ([Methods: Determination of community composition by](#)  
 247 [16S sequencing](#)). Consistent with the behavior of our model, we observed that whenever bottom-up co-selection  
 248 is seen, it is always positive and we do not see any instances of antagonistic co-selection ([Figure 3d](#)). Interestingly,  
 249 bottom-up recruitment appears to be more frequent in the glutamine environments, where top-down co-selection  
 250 was weak, than in the citrate ones, where top-down co-selection was strong ([Figure 2](#)). We then repeated our anal-  
 251 ysis in [Figure 3c](#), this time splitting our data according to the observed strength of bottom-up co-selection instead  
 252 of the primary carbon source as we had done in [Figure 2b](#). Our findings were in line with the model prediction:  
 253 pairwise competition between dominants is only predictive of coalescence outcomes if bottom-up co-selection is  
 254 weak ([Figure 3e](#),  $R^2 = 0.07, p > 0.05, N = 14$  when bottom-up co-selection is strong;  $R^2 = 0.37, p < 10^{-4},$   
 255  $N = 42$  when bottom-up co-selection is weak). Once the bottom-up communities are removed, both the glutamine  
 256 and citrate communities display similar degrees of top-down cohesiveness ([Figure 3e](#), right panel). This suggests  
 257 that the main difference between citrate and glutamine habitats from the standpoint of community coalescence is  
 258 that the latter is richer in communities exhibiting bottom-up cohesiveness than the former. When this difference is  
 259 factored out, both behave similarly.

260 **Understanding the mechanisms of ecological co-selection: a minimal model of community coalescence**

261 In view of the success of our model in reproducing the experimentally observed trends in ecological co-selection,  
262 we set out to better understand the mechanisms for its emergence. In our experimental conditions and in the  
263 MicroCRM simulations, communities are sustained by dense cross-feeding facilitation networks. These networks  
264 can have a very vertical, top-down structure if a single species (the dominant) cross-feeds the rarer members of  
265 the community but these do not cross feed the dominant in return. Alternatively, if the dominant is strongly  
266 cross-fed by its cohort the network structure would be more horizontal. In the latter scenario, positive bottom-up  
267 co-selection of a dominant can take place if cross-feeding from its cohort allows it to persist in the final community  
268 after coalescence –even if it cannot invade successfully in isolation.

269 We found it useful to study a minimal model of community coalescence to test these ideas ([Methods: Minimal](#)  
270 [model](#)). This model is comprised of two communities (resident and invasive) with only two species each as  
271 illustrated in [Figure 4a](#). Within each community, the dominant species ( $S_1$  and  $s_1$  respectively) are able to utilize  
272 the single externally supplied resource ( $R_1$ ). They secrete a single byproduct ( $R_2$  and  $r_2$  respectively) off which  
273 the sub-dominants ( $S_2$  and  $s_2$  respectively) can feed. Finally, these sub-dominants secrete an additional resource  
274 ( $R_3$  and  $r_3$  respectively). The dominants' ability to utilize their sub-dominants' metabolic byproducts determines  
275 whether the structure of the cross-feeding networks of these minimal communities is vertical (if the dominants  
276 cannot utilize the cohort secretions and thus are not cross-fed by them) or horizontal (in the opposite scenario).  
277 The model parameters controlling how effectively the dominants can metabolize said byproducts modulate the  
278 direction of the cross-feeding networks ([Figure 4a](#)).

279 In the limit case when the cross-feeding networks of both the invasive and resident communities are strictly  
280 vertical (that is, the sub-dominants are passively sustained by the dominants but do not cross-feed them), but  
281 also different in the resources each secrete, it is straightforward that the outcome of community coalescence will  
282 depend on the competitive ability of the dominants to grow on the single externally supplied resource. The most  
283 competitive dominant will co-select its sub-dominant (i.e. top-down co-selection) through the secretion of specific  
284 metabolic byproducts that it can consume (but which the sub-dominant of the other community ([Figure 4b](#)). If  
285 the resident community is maintained by a more horizontal cross-feeding network, it can display further resistance  
286 to invasion by the vertical invasive community. In this scenario, even if the resident dominant is less competitive  
287 for the externally supplied resource than the invasive dominant, cross-feeding from the resident cohort can in  
288 principle still allow it to persist in coalescence. The stronger the metabolic flux from the resident cohort towards  
289 the dominant, the more prominent this effect can be ([Figure 4c](#)). The situation could become more interesting when  
290 both resident and invasive communities exhibit a horizontal cross-feeding network. In this scenario, we reasoned  
291 that the invasive dominant may not be able to invade the resident community by itself if it is less competitive for the  
292 externally supplied resource than the resident dominant ([Figure 4d](#)), but also if, despite being more competitive,  
293 cross-feeding from the resident cohort towards the resident dominant favors the success of the latter ([Figure 4e](#)).  
294 Even in this case, we reason that the invasive community could still dominate in coalescence, provided that cross-  
295 feeding from the invasive cohort towards the invasive dominant is strong enough to overcome the competitive  
296 disadvantage that said dominant may have in isolation.

297 In summary, thinking through our minimal model tells us that coalescence outcomes should be contingent on  
298 the direction of the cross-feeding networks sustaining the communities in this simple setting. To verify our intui-  
299 tive reasoning above, we ran simulations of all scenarios described above with our minimal model of community  
300 coalescence implemented in the MicroCRM framework ([Methods: Minimal model](#)). In line with our initial propo-  
301 sition, simulations indicate that bottom-up co-selection of a dominant that is unable to invade by itself is possible  
302 if said dominant is strongly cross-fed by its cohort ([Figure 4](#)).

303 **Community hierarchy regulates the strength of bottom-up co-selection**

304 How do the ideas above scale to more complex and diverse communities? In natural microbiomes and in our  
305 laboratory cultures, a large number of species can coexist and cross-feed each other, giving rise to facilitation  
306 networks that are far more dense than the ones in our minimal model. To generalize the intuition gained in  
307 [Figure 4](#) to communities with more than two species, we introduce a hierarchy index  $h$  that quantifies how vertical  
308 a cross-feeding network is:

$$h = \frac{\Delta N_{\text{dom}}^{\text{R1}}}{\Delta N_{\text{dom}}} \quad (3)$$

309 where  $\Delta N_{\text{dom}}$  represents the overall increase in dominant biomass within a single batch incubation for a genera-  
310 tionally stable community, and  $\Delta N_{\text{dom}}^{\text{R1}}$  represents the increase in said biomass resulting from the metabolism of  
311 the primary resource ( $R_1$ ) only. If the dominant was just utilizing the primary resource, the cross-feeding network  
312 would be very vertical ( $h \sim 1$ ), whereas if it was growing mostly on the secretions of other taxa, it would be more

313 horizontal ( $h \ll 1$ ). We quantified the hierarchies of the resident and invasive communities in the MicroCRM  
314 simulations shown in [Figure 3b](#), finding that  $h$  follows a bimodal distribution ([Figure 5a](#)). We therefore divided  
315 our simulations into four groups according to whether the cross-feeding networks of both resident and invasive  
316 communities were vertical (high  $h$ ) or horizontal (low  $h$ ) as shown in [Figure 5b](#). For each group, we evaluated the  
317 frequency of instances of bottom-up co-selection, i.e. the fraction of cases where a dominant that could not invade  
318 in isolation was successful when accompanied by its cohort (green area of [Figure 3b](#)). We found that bottom-up  
319 ecological co-selection is significantly more frequent when the invasive community is non-hierarchical ([Figure 5c](#)),  
320 in line with what the minimal model anticipated ([Figure 4d-e](#)).

## 321 Conclusions

322 Understanding the mechanisms underlying the responses of microbial communities to invasions is an essential but  
323 poorly understood question in microbial ecology [10]. Theory has suggested that communities may exhibit an  
324 emergent cohesiveness [11, 17, 23, 24], leading to members of the same community recruiting one another during  
325 community-community invasions. Our results provide direct experimental evidence of ecological co-selection in  
326 a large number of community coalescence experiments, and highlight the critical role that may be played by the  
327 rarer, sub-dominant species in the generation of community cohesiveness.

328 Our simulations suggest that the strength and direction of ecological co-selection may be modulated by the  
329 underlying cross-feeding networks that shape the structure of communities in synthetic minimal environments  
330 [32, 33]. This idea is supported by the observation that our Microbial Consumer-Resource Model captures the  
331 trends observed experimentally when we enable a large variation in the metabolic fluxes across species. The model  
332 predicts a trade-off between the strength of bottom-up co-selection and the ability of dominant-dominant pairwise  
333 competition to dictate coalescence outcomes, which we have confirmed experimentally. It also suggests that rarer  
334 taxa may play a more prominent role in co-selecting dominant species when the cross-feeding interactions across  
335 community members are horizontal rather than hierarchical. Testing this theoretical prediction would require one  
336 to map the cross-feeding networks of all of our communities. Keeping track of every nutrient secreted by every  
337 species in co-culture and by which species they are uptaken is still a low throughput process that is both labor  
338 intensive and expensive, but recent progress in metabolomic tools promise to help us test this hypothesis in future  
339 work. Our findings, together with previous results in different systems [27] as well as theoretical predictions  
340 [11, 22–26], suggest that collective interactions of microbes with one another and with the environment should be  
341 generically expected to produce ecological co-selection during community coalescence.

342 **Methods**

343 **Stabilization of environmental communities in simple synthetic environments**

344 Communities were stabilized *ex situ* as described in [32]. In short, environmental samples (soil, leaves...) within  
345 one meter radius in eight different geographical locations were collected with sterile tweezers or spatulas into 50mL  
346 sterile tubes (**Figure 1a**). One gram of each sample was allowed to sit at room temperature in 10mL of phosphate  
347 buffered saline (1×PBS) containing 200µg/mL cycloheximide to suppress eukaryotic growth. After 48h, samples  
348 were mixed 1:1 with 80% glycerol and kept frozen at -80°C. Starting microbial communities were prepared by  
349 scraping the frozen stocks into 200µL of 1×PBS and adding a volume of 4µL to 500µL of synthetic minimal media  
350 (1×M9) supplemented with 200µg/mL cycloheximide and 0.07 C-mol/L glutamine or sodium citrate as the carbon  
351 source in 96 deep-well plates (1.2mL; VWR). Cultures were then incubated still at 30°C to allow for re-growth.  
352 After 48h, samples were fully homogenized and biomass increase was followed by measuring the optical density  
353 (620nm) of 100µL of the cultures in a Multiskan FC plate reader (Thermo Scientific). Communities were stabilized  
354 [32] by passaging 4µL of the cultures into 500µL of fresh media (1×M9 with the carbon source) every 48h for  
355 a total of 12 transfers at a dilution factor of 1:100, roughly equivalent to 80 generations per culture (**Figure 1b**).  
356 Cycloheximide was not added to the media after the first two transfers.

357 **Isolation of dominant species**

358 For each community, the most abundant colony morphotype at the end of the ninth transfer was selected (**Figure 1c**),  
359 resuspended in 100µL 1×PBS and serially diluted (1:10). Next, 20µL of the cells diluted to 10<sup>-6</sup> were plated in the  
360 corresponding synthetic minimal media and allowed to regrow at 30°C for 48h. Dominants were then identified,  
361 inoculated into 500µL of fresh media and incubated still at 30°C for 48h. After this period, the communities  
362 stabilized for eleven transfers and the isolated dominants were ready for the competition experiments at the onset  
363 of the twelfth transfer.

364 **Coalescence, competition and invasion experiments**

365 All possible pairwise dominant-dominant and community-community competition experiments were performed  
366 by mixing equal volumes (4µL) of each of the eight communities or eight dominants at the onset of the twelfth  
367 transfer. Competitions were set up in their native media, i.e. in 500µL of 1×M9 supplemented with 0.07 C-mol/L  
368 of either glutamine or citrate in 96 deep-well plates. Plates were incubated at 30°C for 48h. Pairwise competitions  
369 were further propagated for seven serial transfers (roughly 42 generations, **Figure 1f**) by transferring 8µL of each  
370 culture to fresh media (500µL).

371 **Determination of community composition by 16S sequencing**

372 The sequencing protocol was identical to that described in [32]. Community samples were collected by spinning  
373 down at 3500rpm for 25min in a bench-top centrifuge at room temperature; cell pellets were stored at -80°C  
374 before processing. To maximize Gram-positive bacteria cell wall lysis, the cell pellets were re-suspended and  
375 incubated at 37°C for 30min in enzymatic lysis buffer (20mM Tris-HCl, 2mM sodium EDTA, 1.2% Triton X-100)  
376 and 20mg/mL of lysozyme from chicken egg white (Sigma-Aldrich). After cell lysis, the DNA extraction and  
377 purification was performed using the DNeasy 96 protocol for animal tissues (Qiagen). The clean DNA in 100µL  
378 elution buffer of 10mM Tris-HCl, 0.5mM EDTA at pH 9.0 was quantified using Quan-iT PicoGreen dsDNA Assay  
379 Kit (Molecular Probes, Inc.) and normalized to 5ng/µL in nuclease-free water (Qiagen) for subsequent 16S rRNA  
380 Illumina sequencing. 16S rRNA amplicon library preparation was performed following a dual-index paired-end  
381 approach [45]. Briefly, PCR amplicon libraries of V4 regions of the 16S rRNA were prepared sing dual-index  
382 primers (F515/R805), then pooled and sequenced using the Illumina MiSeq chemistry and platform. Each sample  
383 went through a 30-cycle PCR in duplicate of 20µL reaction volumes using 5ng of DNA each, dual index primers,  
384 and AccuPrime Pfx SuperMix (Invitrogen). The thermocycling procedure includes a 2min initial denaturation step  
385 at 95°C, and 30 cycles of the following PCR scheme: (a) 20-second denaturation at 95°C, (b) 15-second annealing  
386 at 55°C, and (c) 5-minute extension at 72°C. The duplicate PCR products of each sample were pooled, purified,  
387 and normalized using SequalPrep PCR cleanup and normalization kit (Invitrogen). Barcoded amplicon libraries  
388 were then pooled and sequenced using Illumina Miseq v2 reagent kit, which generated 2×250bp paired-end reads  
389 at the Yale Center for Genome Analysis (YCGA). The sequencing reads were demultiplexed on QIIME 1.9.0 [46].  
390 The barcodes, indexes, and primers were removed from raw reads, producing FASTQ files with both the forward  
391 and reverse reads for each sample, ready for DADA2 analysis [39]. DADA2 version 1.1.6 was used to infer unique  
392 biological exact sequence variants (ESVs) for each sample and naïve Bayes was used to assign taxonomy using  
393 the SILVA version 123 database [47, 48].

394 **Metrics of community distance**

395 Beta-diversity indexes between the invasive and coalesced communities or the resident and coalesced communities  
 396 were computed using various similarity metrics. For two arbitrary communities with ESV abundances represented  
 397 by the vectors  $\mathbf{x} = (x_1, x_2, \dots, x_S)$  and  $\mathbf{y} = (y_1, y_2, \dots, y_S)$  (where  $x_i$  and  $y_i$  represent the relative abundance of the  
 398  $i$ th ESV in each community respectively and  $S$  is the total number of ESVs), the Bray-Curtis similarity  $BC(\mathbf{x}, \mathbf{y})$   
 399 is calculated as [49]

$$BC(\mathbf{x}, \mathbf{y}) = \sum_i \min(x_i, y_i) \quad (4)$$

400 The Jensen-Shannon similarity  $JS(\mathbf{x}, \mathbf{y})$  is defined as one minus the Jensen-Shannon distance (which is, in turn,  
 401 the square root of the Jensen-Shannon divergence [50])

$$JS(\mathbf{x}, \mathbf{y}) = 1 - \sqrt{\frac{1}{2}KL(\mathbf{x}, \mathbf{m}) + \frac{1}{2}KL(\mathbf{y}, \mathbf{m})} \quad (5)$$

402 where  $\mathbf{m} = (\mathbf{x} + \mathbf{y}) / 2$  and  $KL$  denotes the Kullback-Leibler divergence [51]

$$KL(\mathbf{x}, \mathbf{y}) = \sum_i x_i \log_2 \left( \frac{x_i}{y_i} \right) \quad (6)$$

403 Using base-two logarithms ensures that the metric is bounded between 0 and 1. The Jaccard similarity is given by  
 404  $J(\mathbf{x}, \mathbf{y})$  [52]

$$J(\mathbf{x}, \mathbf{y}) = \frac{|\mathbf{x} \cap \mathbf{y}|}{|\mathbf{x} \cup \mathbf{y}|} \quad (7)$$

405 Additionally, we quantified coalescence outcomes by examining the fraction of the endemic cohort of the original  
 406 communities that persists in the coalesced one. We call  $E(\mathbf{x}, \mathbf{y})$  to the fraction of endemic species of  $\mathbf{x}$  that are also  
 407 found in  $\mathbf{y}$ .

408 For all the metrics above, we quantified the relative similarity between the invasive and the coalesced communi-  
 409 ties using relative metrics (denoted as  $Q$ ):

$$Q(\mathbf{x}_I, \mathbf{x}_R, \mathbf{x}_C) = \frac{F(\mathbf{x}_I, \mathbf{x}_C)}{F(\mathbf{x}_I, \mathbf{x}_C) + F(\mathbf{x}_R, \mathbf{x}_C)} \quad (8)$$

410 where the subindices I, R and C correspond to the invasive, resident and coalesced communities respectively,  
 411 and  $F$  represents one of  $BC$  (Bray-Curtis similarity),  $JS$  (Jensen-Shannon similarity),  $J$  (Jaccard similarity) or  $E$   
 412 (endemic survival) defined above.

413 **Simulations**

414 We used the Community Simulator package [35] and included new features for our simulations. In the package,  
 415 species are characterized by their resource uptake rates ( $c_{ia}$  for species  $i$  and resource  $\alpha$ ), and they all share a  
 416 common metabolic matrix  $\mathbf{D}$ . The element  $D_{\alpha\beta}$  of this matrix represents the fraction of energy in the form of  
 417 resource  $\alpha$  secreted when resource  $\beta$  is consumed. Here we implemented a new operation mode in which species  
 418 can secrete different metabolites (and/or in different abundances) when consuming a same resource. We call  $D_{i\alpha\beta}$  to  
 419 the fraction of energy in the form of resource  $\alpha$  secreted by species  $i$  when consuming resource  $\beta$ . In the Community  
 420 Simulator underlying Microbial Consumer-Resource Model, this means that the energy flux  $J_{i\beta}^{\text{out}}$  [32, 34] now takes  
 421 the form

$$J_{i\beta}^{\text{out}} = \sum_{\alpha} D_{i\beta\alpha} l_{\alpha} J_{i\alpha}^{\text{in}} \quad (9)$$

422 The documentation for the Community Simulator contains detailed descriptions of the model formulation, param-  
 423 eters and package use. For the updated package with the new functionality, see [Data & code availability](#).

424 For our simulations, we first generated a library of 2640 species divided into three specialist families of 800  
 425 members each and a generalist family of 240 members. We split this library into two non-overlapping pools of  
 426 1320 species each. We randomly sampled 50 species from each pool in equal ratios to seed 100 resident and 100  
 427 invasive communities respectively. We then let grow and diluted the communities serially, replenishing the primary  
 428 resource after each dilution. We repeated the process 20 times to ensure generational equilibrium was achieved  
 429 [32]. We then performed the *in silico* experiments by using the generationally stable communities to seed 100  
 430 coalesced communities that were again stabilized as described previously. Similarly, we identified the dominant

431 (most abundant) species of every resident and invasive community to carry out pairwise competition and single  
 432 invasion simulations.

433 Most other parameters were set to the defaults of the original Community Simulator package, with the only  
 434 exception of the maintenance costs ( $m$ ) which are set to zero for all species (equivalent to assuming cell death is  
 435 negligible through the duration of our growth cycles) and the sparsity of the metabolic matrices ( $s$ ) which is set to  
 436 0.9 to generate significant variability in the secretion fluxes across different species (see main text).

#### 437 Minimal model

438 Our minimal model is set within the same MicroCRM framework that we used for the previous simulations. As  
 439 described in the main text, the model contains two communities of two species each ( $S_1$  and  $S_2$  in the resident  
 440 community,  $r_1$  and  $r_2$  in the invasive community), with five resources in total, out of which the first one ( $R_1$ ) is  
 441 replenished externally at the beginning of each growth cycle and the rest correspond to the species' metabolic  
 442 byproducts. Each species secretes a unique byproduct, meaning that the metabolic matrix  $\mathbf{D}$  is binary in this case.  
 443 The specific structure of  $\mathbf{D}$  is displayed below –because it is a 3-dimensional matrix in our framework, we have  
 444 “sliced” it into the four 2-dimensional matrices corresponding to our four species.

$$\mathbf{D}_1 = \begin{array}{ccccc} & R_1 & R_2 & R_3 & r_2 & r_3 \\ \begin{matrix} R_1 \\ R_2 \\ R_3 \\ r_2 \\ r_3 \end{matrix} & \left( \begin{matrix} 0 & * & 0 & * & * \\ 1 & * & 0 & * & * \\ 0 & * & 1 & * & * \\ 0 & * & 0 & * & * \\ 0 & * & 0 & * & * \end{matrix} \right) \end{array}$$

$$\mathbf{D}_2 = \begin{array}{ccccc} & R_1 & R_2 & R_3 & r_2 & r_3 \\ \begin{matrix} R_1 \\ R_2 \\ R_3 \\ r_2 \\ r_3 \end{matrix} & \left( \begin{matrix} * & 0 & * & * & * \\ * & 0 & * & * & * \\ * & 1 & * & * & * \\ * & 0 & * & * & * \\ * & 0 & * & * & * \end{matrix} \right) \end{array}$$

$$\mathbf{D}_3 = \begin{array}{ccccc} & R_1 & R_2 & R_3 & r_2 & r_3 \\ \begin{matrix} R_1 \\ R_2 \\ R_3 \\ r_2 \\ r_3 \end{matrix} & \left( \begin{matrix} 0 & * & * & * & 0 \\ 0 & * & * & * & 0 \\ 0 & * & * & * & 0 \\ 1 & * & * & * & 0 \\ 0 & * & * & * & 1 \end{matrix} \right) \end{array}$$

$$\mathbf{D}_4 = \begin{array}{ccccc} & R_1 & R_2 & R_3 & r_2 & r_3 \\ \begin{matrix} R_1 \\ R_2 \\ R_3 \\ r_2 \\ r_3 \end{matrix} & \left( \begin{matrix} * & * & * & 0 & * \\ * & * & * & 0 & * \\ * & * & * & 0 & * \\ * & * & * & 0 & * \\ * & * & * & 1 & * \end{matrix} \right) \end{array}$$

445 Asterisks indicate values of  $D_{i\beta\alpha}$  that are irrelevant because species  $i$  cannot utilize resource  $\alpha$  (and so the metabolic  
 446 flux from  $\alpha$  to  $\beta$  corresponding to that species will always be zero regardless of the value of  $D_{i\alpha\beta}$ ). The consumer  
 447 preference matrix  $\mathbf{c}$  takes the following form:

$$\mathbf{c} = \begin{array}{ccccc} & R_1 & R_2 & R_3 & r_2 & r_3 \\ \begin{matrix} S_1 \\ S_2 \\ s_1 \\ s_2 \end{matrix} & \left( \begin{matrix} C_{11} & 0 & C_{13} & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 \\ c_{11} & 0 & 0 & 0 & c_{13} \\ 0 & 0 & 0 & 100 & 0 \end{matrix} \right) \end{array}$$

448 Where we made the sub-dominants equally strong consumers of their dominants' secretions ( $C_{22} = c_{22} = 100$ ),  
 449 and we varied all other uptake rates depending on the scenario we were considering (see main text). Whenever we  
 450 were interested in the ratio between two rates (e.g.  $c_{13}/C_{13}$  in Figure 4e) we gave the one in the denominator a  
 451 fixed value of 1 and let one in the numerator range within the specified limits.

452 **Data & code availability**

453 Experimental data and code for the analysis, as well as code for the simulations and the updated Community  
454 Simulator package with instructions for enabling the new features are in [github.com/jdiazc9/coalescence](https://github.com/jdiazc9/coalescence).

455 **Author contributions**

456 AS, MT conceived the study and designed the experiments. NL and ASG performed experiments. NL, JEG, and  
457 JDC processed and analyzed experimental data. HSC, CYC and JDC wrote simulations. JDC analyzed simulations  
458 results. AS and JDC interpreted results. JDC, NL, ASG, MT and AS wrote the paper.

459 **Funding**

460 Work in the Sanchez lab is supported by NIH grant 1R35 GM133467-01, and by a Packard Fellowship from the  
461 David and Lucille Packard Foundation. The funding for this work partly results from a Scialog Program spon-  
462 sored jointly by the Research Corporation for Science Advancement and the Gordon and Betty Moore Foundation  
463 through grants to Yale University by the Research Corporation and the Simons Foundation.

464 **Acknowledgements**

465 The authors wish to thank Pankaj Mehta, Wenping Cui, Robert Marsland and all members of the Sanchez labora-  
466 tory for many helpful discussions. We also wish to express our gratitude to the Goodman laboratory at Yale for  
467 technical help during the early stages of this project.

## 468 References

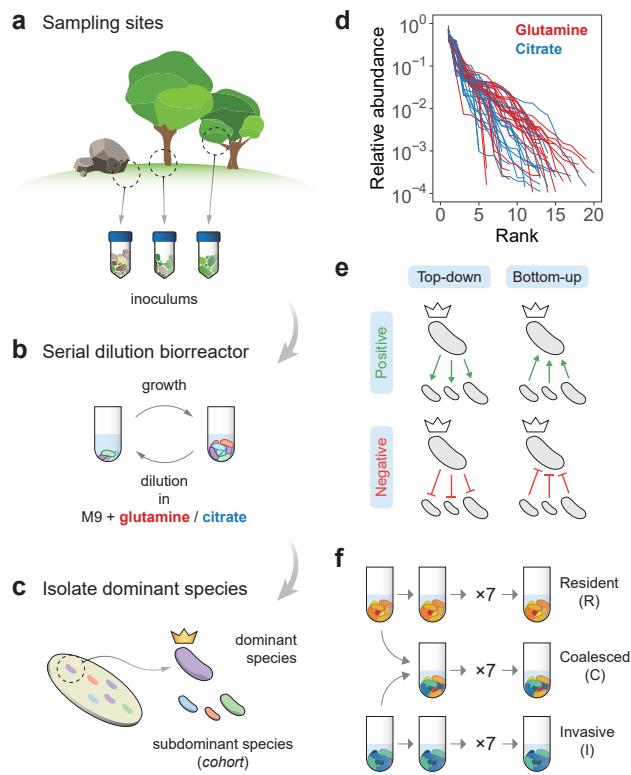
- 469 1. Mansour I, Heppell CM, Ryo M and Rillig MC (2018). Application of the microbial community coalescence  
470 concept to riverine networks. *Biological Reviews* **93**(4):1832–1845
- 471 2. Luo X, Xiang X, Yang Y, Huang G, Fu K, Che R and Chen L (2020). Seasonal effects of river flow on microbial  
472 community coalescence and diversity in a riverine network. *FEMS Microbiology Ecology* **96**(8):fiaa132
- 473 3. Vass M, Székely AJ, Lindström ES, Osman OA and Langenheder S (2021). Warming mediates the resistance  
474 of aquatic bacteria to invasion during community coalescence. *Molecular Ecology* **30**(5):1345–1356
- 475 4. Rillig MC, Lehmann A, Aguilar-Trigueros CA, Antonovics J, Caruso T, Hempel S, Lehmann J, Valyi K,  
476 Verbruggen E et al. (2016). Soil microbes and community coalescence. *Pedobiologia* **59**(1-2):37–40
- 477 5. Ramoneda J, Le Roux J, Stadelmann S, Frossard E, Frey B and Gamper HA (2020). Coalescence of rhizo-  
478 bial communities in soil interacts with fertilization and determines the assembly of rhizobia in root nodules.  
479 *bioRxiv*
- 480 6. Rochefort A, Simonin M, Marais C, Guillerm-Erckelboudt AY, Barret M and Sarniguet A (2020). Asymmetric  
481 outcome of community coalescence of seed and soil microbiota during early seedling growth. *bioRxiv*
- 482 7. Evans SE, Bell-Dereske LP, Dougherty KM and Kittredge HA (2019). Dispersal alters soil microbial commu-  
483 nity response to drought. *Environmental Microbiology* **22**(3):905–916
- 484 8. Dutton CL, Subalusky AL, Sanchez A, Estrela S, Lu N, Hamilton SK, Njoroge L, Rosi EJ and Post DM  
485 (2021). The meta-gut: Hippo inputs lead to community coalescence of animal and environmental micro-  
486 biomes. *biorXiv*
- 487 9. Vandegrift R, Fahimipour AK, Muscarella M, Bateman AC, Wymelenberg KVD and Bohannan BJ (2019).  
488 Moving microbes: the dynamics of transient microbial residence on human skin. *biorXiv*
- 489 10. Rillig MC, Antonovics J, Caruso T, Lehmann A, Powell JR, Veresoglou SD and Verbruggen E (2015). Inter-  
490 change of entire communities: microbial community coalescence. *Trends in Ecology & Evolution* **30**(8):470–  
491 476
- 492 11. Gilpin M (1994). Community-level competition: asymmetrical dominance. *Proceedings of the National  
493 Academy of Sciences* **91**(8):3252–3254
- 494 12. Simberloff D and Holle BV (1999). Positive Interactions of Nonindigenous Species: Invasional Meltdown?  
495 *Biological Invasions* **1**(1):21–32
- 496 13. Grosholz ED (2005). Recent biological invasion may hasten invasional meltdown by accelerating historical  
497 introductions. *Proceedings of the National Academy of Sciences* **102**(4):1088–1091
- 498 14. Simberloff D (2006). Invasional meltdown 6 years later: important phenomenon, unfortunate metaphor, or  
499 both? *Ecology Letters* **9**(8):912–919
- 500 15. Gurevitch J (2006). Commentary on Simberloff (2006): Meltdowns, snowballs and positive feedbacks. *Ecol-  
501 ogy Letters* **9**(8):919–921
- 502 16. Green PT, O'Dowd DJ, Abbott KL, Jeffery M, Retallick K and Nally RM (2011). Invasional meltdown:  
503 Invader-invader mutualism facilitates a secondary invasion. *Ecology* **92**(9):1758–1768
- 504 17. Livingston G, Jiang Y, Fox JW and Leibold MA (2013). The dynamics of community assembly under sudden  
505 mixing in experimental microcosms. *Ecology* **94**(12):2898–2906
- 506 18. Prior KM, Robinson JM, Dunphy SAM and Frederickson ME (2015). Mutualism between co-introduced  
507 species facilitates invasion and alters plant community structure. *Proceedings of the Royal Society B: Biolog-  
508 ical Sciences* **282**(1800):20142846
- 509 19. O'Loughlin LS and Green PT (2017). Secondary invasion: When invasion success is contingent on other  
510 invaders altering the properties of recipient ecosystems. *Ecology and Evolution* **7**(19):7628–7637
- 511 20. Castledine M, Sierociński P, Padfield D and Buckling A (2020). Community coalescence: an eco-evolutionary  
512 perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**(1798):20190252

- 513 21. Rillig MC, Tsang A and Roy J (2016). Microbial Community Coalescence for Microbiome Engineering.  
514 *Frontiers in Microbiology* **7**:1967
- 515 22. Toquenaga Y (1997). Historicity of a Simple Competition Model. *Journal of Theoretical Biology* **187**(2):175–  
516 181
- 517 23. Tikhonov M (2016). Community-level cohesion without cooperation. *eLife* **5**:e15747
- 518 24. Tikhonov M and Monasson R (2017). Collective Phase in Resource Competition in a Highly Diverse Ecosys-  
519 tem. *Physical Review Letters* **118**(4):048103
- 520 25. Vila JCC, Jones ML, Patel M, Bell T and Rosindell J (2019). Uncovering the rules of microbial community  
521 invasions. *Nature Ecology & Evolution* **3**(8):1162–1171
- 522 26. Lechón P, Clegg T, Cook J, Smith TP and Pawar S (2021). The role of competition versus cooperation in  
523 microbial community coalescence. *biorXiv*
- 524 27. Sierociński P, Milferstedt K, Bayer F, Großkopf T, Alston M, Bastkowski S, Swarbreck D, Hobbs PJ, Soyer OS  
525 et al. (2017). A Single Community Dominates Structure and Function of a Mixture of Multiple Methanogenic  
526 Communities. *Current Biology* **27**(21):3390–3395.e4
- 527 28. Rillig MC and Mansour I (2017). Microbial Ecology: Community Coalescence Stirs Things Up. *Current  
528 Biology* **27**(23):R1280–R1282
- 529 29. Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DS, Parfrey LW, Farjalla VF and Doebeli M (2016).  
530 High taxonomic variability despite stable functional structure across microbial communities. *Nature Ecology  
531 & Evolution* **1**(1):0015
- 532 30. Winfree R, Fox JW, Williams NM, Reilly JR and Cariveau DP (2015). Abundance of common species, not  
533 species richness, drives delivery of a real-world ecosystem service. *Ecology Letters* **18**(7):626–635
- 534 31. Rosenzweig RF, Sharp RR, Treves DS and Adams J (1994). Microbial evolution in a simple unstructured  
535 environment: genetic differentiation in *Escherichia coli*. *Genetics* **137**(4):903–917
- 536 32. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, Segrè D, Mehta P and Sanchez A  
537 (2018). Emergent simplicity in microbial community assembly. *Science* **361**(6401):469–474
- 538 33. Estrela S, Vila JCC, Lu N, Bajic D, Rebollo-Gomez M, Chang CY and Sanchez A (2020). Metabolic rules  
539 of microbial community assembly. *biorXiv*
- 540 34. Marsland III R, Cui W, Goldford J, Sanchez A, Korolev K and Mehta P (2019). Available energy fluxes drive  
541 a transition in the diversity, stability, and functional structure of microbial communities. *PLoS Computational  
542 Biology* **15**(2):e1006793
- 543 35. Marsland R, Cui W, Goldford J and Mehta P (2020). The Community Simulator: A Python package for  
544 microbial ecology. *PLoS ONE* **15**(3):e0230430
- 545 36. Dimroth P (2004). Molecular Basis for Bacterial Growth on Citrate or Malonate. *EcoSal Plus* **1**(1)
- 546 37. Forchhammer K (2007). Glutamine signalling in bacteria. *Frontiers in Bioscience* **12**(1):358
- 547 38. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA and Holmes SP (2016). DADA2: High-  
548 resolution sample inference from Illumina amplicon data. *Nature Methods* **13**(7):581–583
- 549 39. Callahan BJ, McMurdie PJ and Holmes SP (2017). Exact sequence variants should replace operational taxo-  
550 nomic units in marker-gene data analysis. *The ISME Journal* **11**:2639–2643
- 551 40. Marsland R, Cui W and Mehta P (2020). A minimal model for microbial biodiversity can reproduce experi-  
552 mentally observed ecological patterns. *Scientific Reports* **10**:3308
- 553 41. Estrela S, Sanchez-Gorostiaga A, Vila JC and Sanchez A (2021). Nutrient dominance governs the assembly  
554 of microbial communities in mixed nutrient environments. *eLife* **10**
- 555 42. MacArthur R (1970). Species packing and competitive equilibrium for many species. *Theoretical Population  
556 Biology* **1**(1):1–11

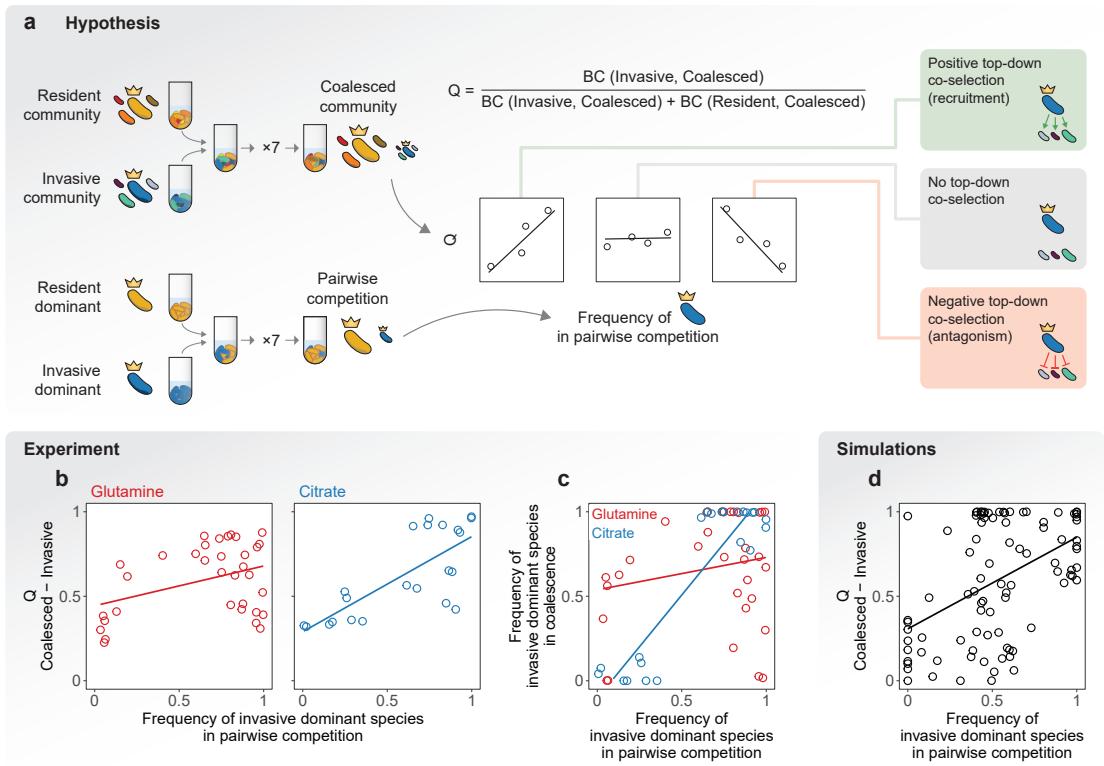
- 557 43. Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, Bonilla G, Kar A, Leiby N et al.  
558 (2014). Metabolic Resource Allocation in Individual Microbes Determines Ecosystem Interactions and Spatial  
559 Dynamics. *Cell Reports* **7**(4):1104–1115
- 560 44. Pinu FR, Granucci N, Daniell J, Han TL, Carneiro S, Rocha I, Nielsen J and Villas-Boas SG (2018). Metabolite  
561 secretion in microorganisms: the theory of metabolic overflow put to the test. *Metabolomics* **14**(4)
- 562 45. Kozich JJ, Westcott SL, Baxter NT, Highlander SK and Schloss PD (2013). Development of a Dual-Index  
563 Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina  
564 Sequencing Platform. *Applied and Environmental Microbiology* **79**(17):5112–5120
- 565 46. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich  
566 JK et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*  
567 **7**:335–336
- 568 47. Wang Q, Garrity GM, Tiedje JM and Cole JR (2007). Naïve Bayesian Classifier for Rapid Assignment of  
569 rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology* **73**(16):5261–  
570 5267
- 571 48. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J and Glöckner FO (2013). The SILVA  
572 ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*  
573 **41**(D1):D590–D596
- 574 49. Curtis JT and Bray JR (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin.  
575 *Ecological Monographs* **27**(4):325–349
- 576 50. Lin J (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*  
577 **37**(1):145–151
- 578 51. Kullback S and Leibler RA (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*  
579 **22**(1):79–86
- 580 52. Jaccard P (1912). The distribution of the flora in the alpine zone. *New Phytologist* **11**(2):37–50

581 **Figures**

582

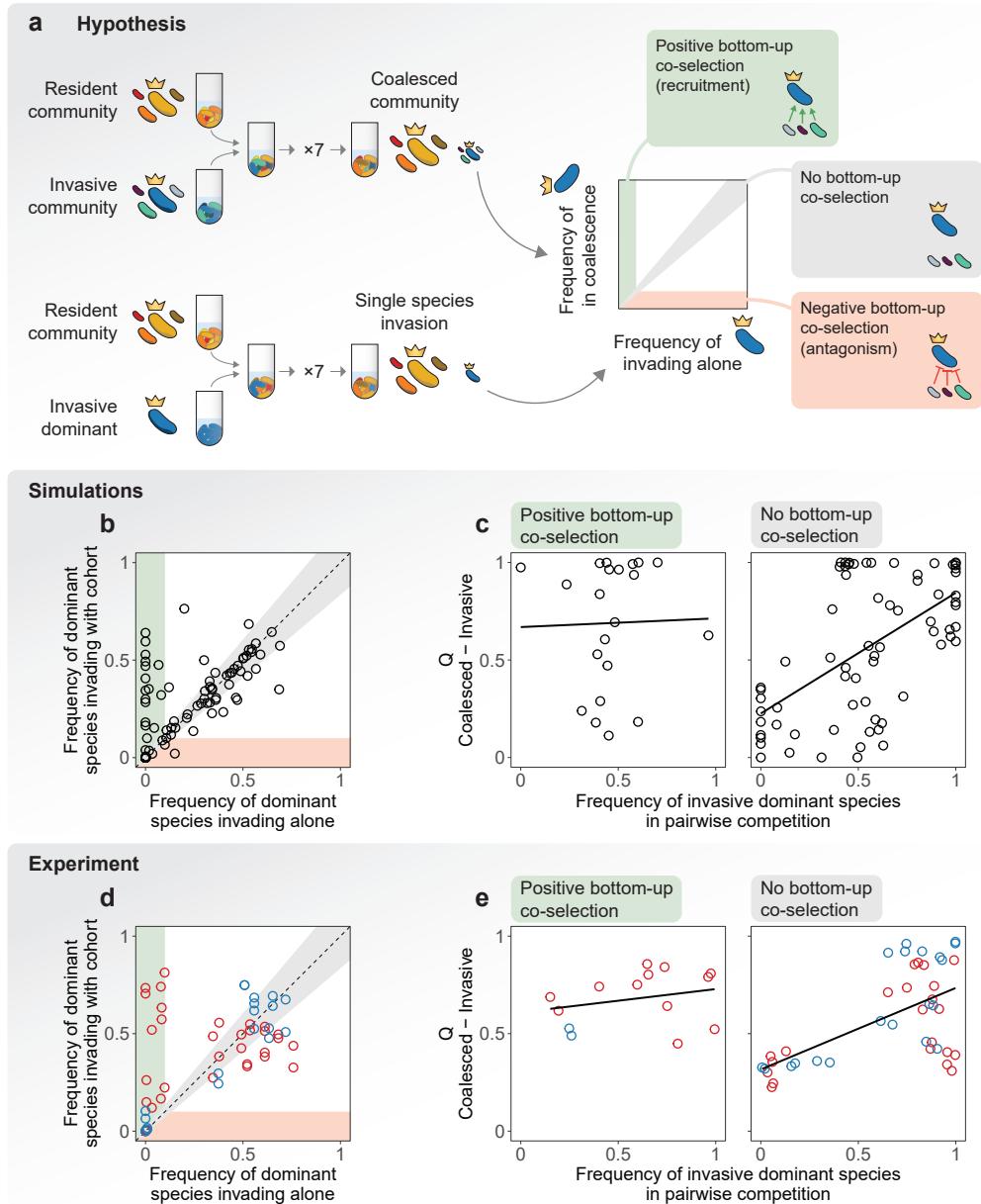


583 **Figure 1. Overview of the experimental protocol.** **a.** Environmental samples collected from eight different locations were used to inoculate our communities. **b.** Communities were stabilized in serial batch culture bioreactors in minimal synthetic media with glutamine or citrate as the only supplied carbon source. **c.** Communities were plated in minimal media agar plates and the most abundant species (the “dominants”) from each community were isolated. We refer to the set of sub-dominant species as the “cohorts”. **d.** Rank-frequency distributions of the eight communities stabilized in either glutamine (red) or citrate (blue), sequenced at a depth of  $10^4$  reads. Three biological replicates per community are shown. Community compositions are skewed and long-tailed. **e.** Our hypothesis is that ecological co-selection can take place from the top-down, i.e. the dominant co-selecting the cohort, or from the bottom-up, i.e. the cohort co-selecting the dominant. Both forms of co-selection can be positive (recruitment) or negative (antagonism). **f.** Illustration of the protocol of our coalescence experiments. All pairs of communities were inoculated into fresh minimal media supplemented with the same carbon source where communities had been previously stabilized. The coalesced (C) and original resident (R) and invasive (I) communities were then serially diluted and allowed to grow for seven additional transfers.



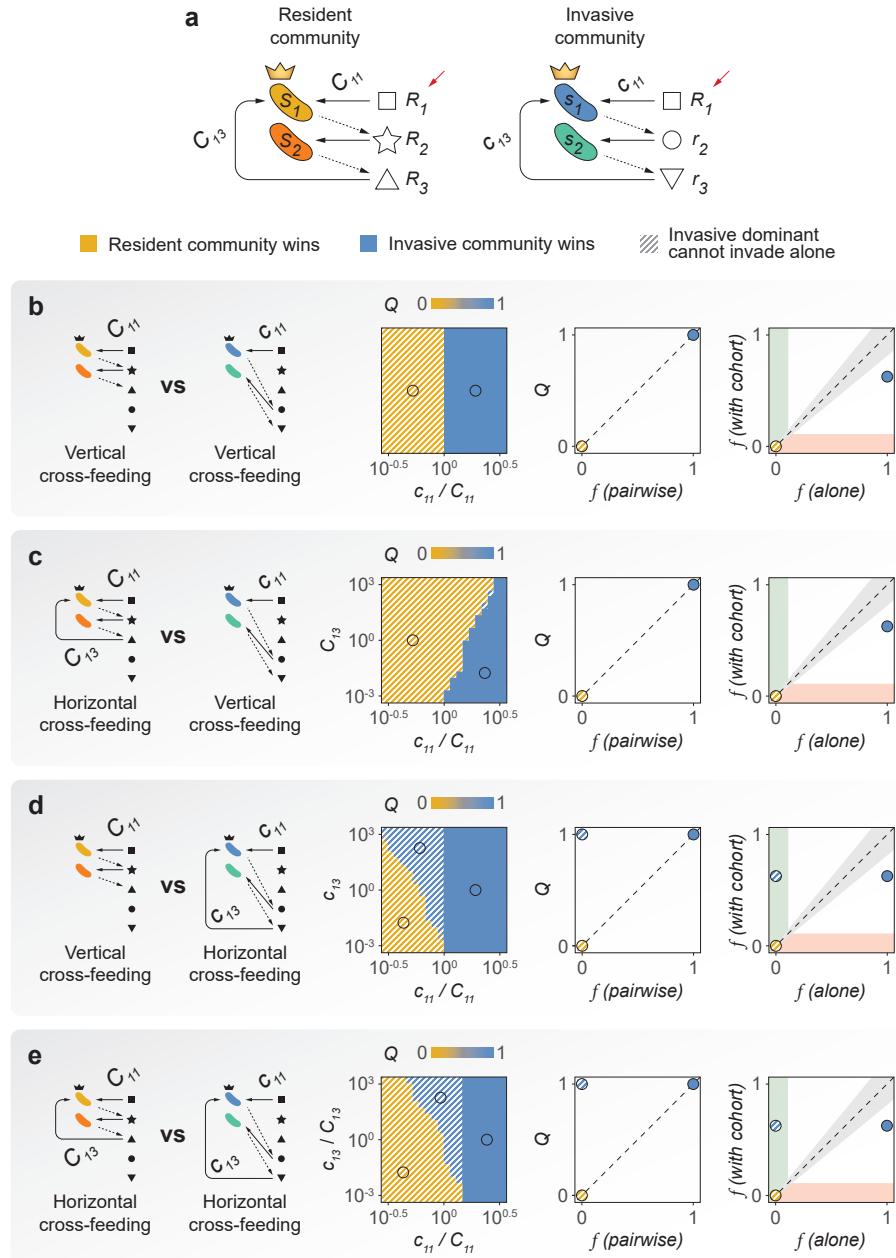
596

597 **Figure 2. Top-down co-selection in microbial community coalescence.** **a.** Experimental setup and hypothesis. We carried  
 598 out coalescence experiments by mixing pairs of generationally stable resident and invasive communities in 1:1 ratios, and then  
 599 allowing the coalesced community to stabilize for seven additional transfers. The relative similarity between the coalesced and  
 600 the invasive communities ( $Q$ ) is quantified using the Bray-Curtis similarity ( $BC$ ). We also carried out head-to-head competition  
 601 experiments between all pairs of dominants from each community. We hypothesize that, if top-down positive co-selection was  
 602 strong, the dominant that is most competitive would co-select its cohort and therefore a positive correlation would be observed  
 603 between  $Q$  and the fraction of the invasive dominant in pairwise competition. Alternatively, top-down negative co-selection  
 604 would result in a negative correlation as the most competitive dominant would antagonize its own cohort. We would see no  
 605 correlation if none of these forms of top-down co-selection were substantial. **b.** Coalescence outcomes are predicted by the  
 606 pairwise competition between the invasive and resident dominant species in our experiments. Left panel (red): glutamine  
 607 communities,  $R^2 = 0.15$ ,  $p < 0.05$ . Right panel (blue): citrate communities,  $R^2 = 0.57$ ,  $p < 10^{-4}$ . A high correlation is  
 608 consistent with a scenario of strong top-down positive co-selection where dominants recruit their cohorts for the final coalesced  
 609 community. Two biological replicates per experiment are plotted individually. **c.** Pairwise competition of dominants with or  
 610 without their cohorts. In the horizontal axis, we plot the frequency of the invasive dominant species in head-to-head pairwise  
 611 competition with the resident dominant. In the vertical axis, we plot the same relative frequency when the two species compete  
 612 in the presence of their cohorts, i.e. during community coalescence.  $R^2 = 0.04$ ,  $p > 0.05$  for glutamine (red) and  $R^2 = 0.83$ ,  
 613  $p < 10^{-8}$  for citrate (blue). **d.** Simulations with a Microbial Consumer-Resource Model are able to capture these trends  
 614 ( $R^2 = 0.22$ ,  $p < 10^{-5}$ ).



616

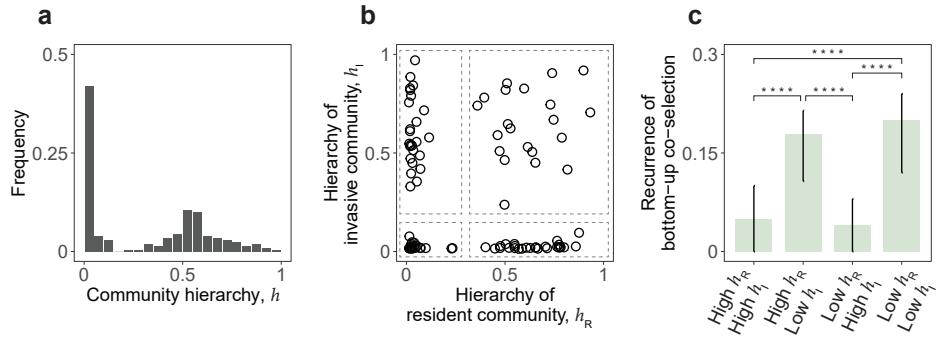
617 **Figure 3. Trade offs between bottom-up and top-down ecological co-selection.** **a.** Experimental setup and hypothesis.  
618 In addition to our coalescence experiments, we carried out a series of assays where we invaded the resident communities  
619 with the invasive dominants in isolation. We hypothesize that three scenarios are possible regarding bottom-up co-selection:  
620 sub-dominant species could co-select for (green) or against (red) their dominant in coalescence, which would result in the  
621 invasive dominant reaching higher (positive bottom-up co-selection) or lower (negative bottom-up co-selection) abundances  
622 when accompanied by its cohort with respect to when invading alone. Alternatively, the cohorts could have no effect in the  
623 invasion success of the dominant taxa (no bottom-up co-selection, gray). **b.** Simulations with a Microbial Consumer-Resource  
624 Model: we plot the frequency reached by the invasive dominants when invading the resident communities in isolation versus  
625 the same frequency when invading together with their cohorts, i.e. in community coalescence. Points in the green/red area  
626 represent instances where the invasive dominant is able to invade with higher/lower success when accompanied by its cohort,  
627 evidencing positive/negative bottom-up co-selection. Points around the diagonal (gray area) correspond to cases where the  
628 success of the invasive dominant is only weakly affected by the presence or absence of its cohort. **c.** We divided the data from  
629 our simulations into two sets according to whether positive or no bottom-up co-selection was observed (that is, whether points  
630 fell into the green or gray areas of panel b). Here we reproduce the plots in Figure 2b for each set, representing the result of  
631 the dominant head-to-head pairwise competition versus the outcome of community coalescence. Left panel: strong positive  
632 bottom-up co-selection ( $R^2 = 0.00, p > 0.05$ ). Right panel: no bottom-up co-selection ( $R^2 = 0.34, p < 10^{-6}$ ). **d.** Experiments  
633 show that in our conditions, positive bottom-up co-selection is indeed more frequent and strong than negative bottom-up co-  
634 selection. **e.** We reproduce the plots in panel c for our experimental data, i.e. we recreate Figure 2b but this time splitting our  
635 data by the strength of bottom-up co-selection instead of by the carbon source supplied to the communities. Left panel: strong  
636 positive bottom-up co-selection ( $R^2 = 0.07, p > 0.05$ ). Right panel: no bottom-up co-selection ( $R^2 = 0.37, p < 10^{-4}$ ).



638

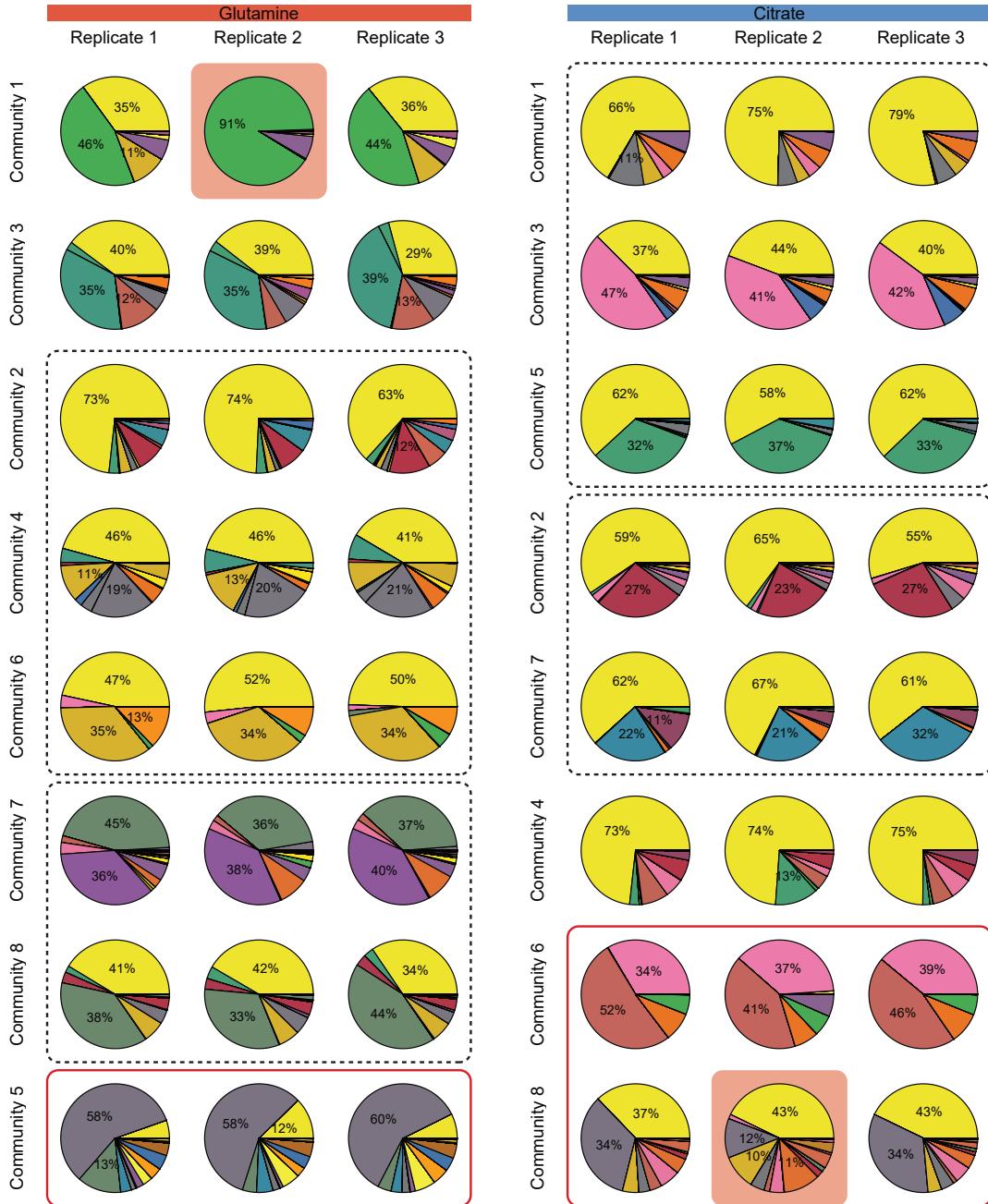
639 **Figure 4. A minimal model of community coalescence.** **a.** Illustration of the model structure and parameters. The primary  
 640 resource ( $R_1$ ) is replenished after each growth-dilution cycle (red arrows). Solid arrows indicate resource consumption, dashed  
 641 arrows represent resource secretion. **b-e.** Coalescence outcomes in the minimal model under different relations of cohesiveness  
 642 between the resident and the invasive communities. We represent the relative Bray-Curtis similarity between the invasive and  
 643 the coalesced communities ( $Q$ ) as a function of the relevant model parameters. For the specific representative cases indicated  
 644 by the hollow circles, we also show  $Q$  as a function of the frequency of the invasive dominant in pairwise competition with the  
 645 resident dominant, as well as the frequency of the invasive dominant invading alone versus invading accompanied by its cohort.

647



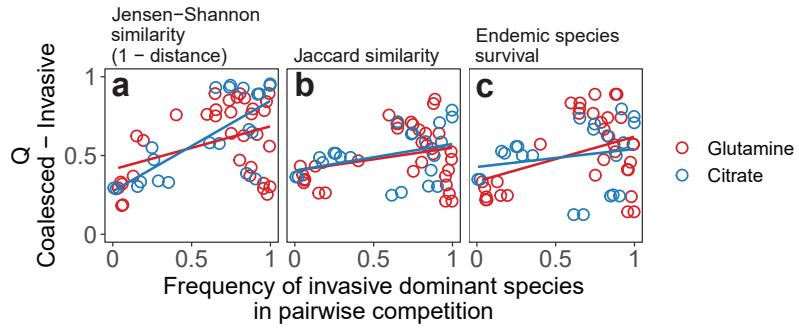
**Figure 5. Community hierarchy modulates the recurrence of bottom-up co-selection.** **a.** Distribution of community hierarchies for our *in silico* communities. **b.** We divided our coalescence simulations into four groups according to the hierarchies of the resident ( $h_R$ ) and invasive ( $h_I$ ) communities as indicated by the dashed boxes. For every group, we calculated the fraction of cases where bottom-up co-selection was observed, i.e. the invasive dominant was unsuccessful when invading in isolation but successful when invading with its cohort. **c.** Bottom-up co-selection of the invasive dominant during coalescence is significantly more frequent when the invasive community is non-hierarchical. Error bars representing 95% confidence intervals and p-values were computed by bootstrapping ( $p < 10^{-4}$  where indicated).

656 **Supplementary Figures**



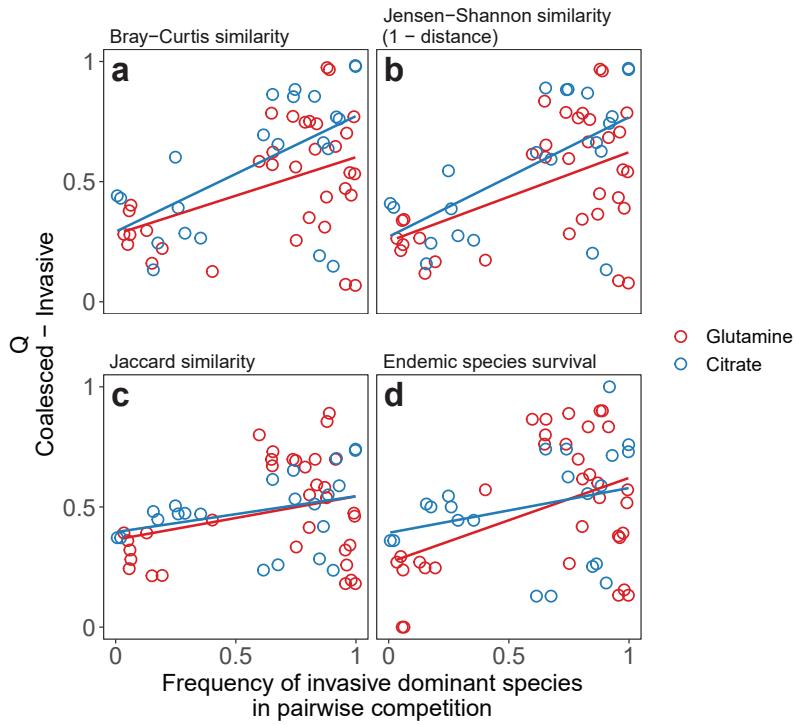
657

658 **Figure S1. Community compositions after seven additional transfers without coalescence.** Each color of the pie plots  
 659 corresponds to a different exact sequence variant ([Methods: Determination of community composition by 16S sequencing](#)).  
 660 Replicate 2 of community 1 from glutamine, as well as replicate 2 of community 8 from citrate (highlighted)  
 661 were removed based on their dissimilarity to the other two replicates (details in code for data analysis, see [Data & code availability](#)). Communities  
 662 clustered in dashed boxes shared the same dominant species as revealed by sequencing data. For communities enclosed  
 663 in red boxes, sequencing data showed that the species isolated by plating was not detectable in the community after seven  
 664 additional transfers (i.e. the dominant was incorrectly identified) and were therefore excluded from downstream analyses.



666

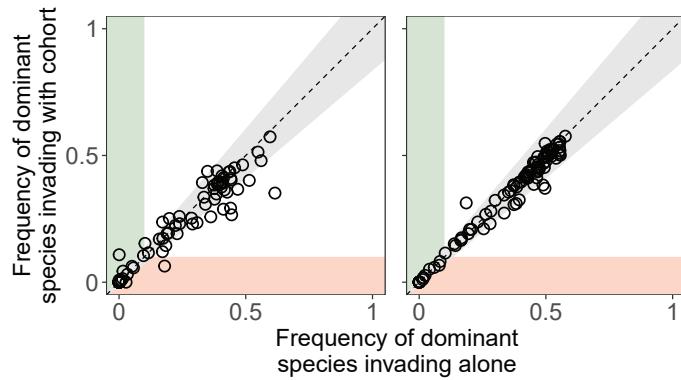
667 **Figure S2. Alternative metrics of community distance.** Quantifying coalescence outcomes using different metrics of commu-  
 668 nity similarity (Methods: Metrics of community distance) gives similar results to those shown in Figure 2b. Metrics that account  
 669 for the relative species abundances (Bray-Curtis or Jensen-Shannon similarities) yield higher correlations than less quantitative  
 670 metrics that only account for species presence/absence (Jaccard similarity or the fraction of endemic invasive species persisting  
 671 in the coalesced community). **a.** Relative Jensen-Shannon similarity ( $R^2 = 0.15$ ,  $p < 0.05$  for glutamine and  $R^2 = 0.53$ ,  
 672  $p < 5 \times 10^{-4}$  for citrate) **b.** Relative Jaccard similarity ( $R^2 = 0.08$ ,  $p > 0.05$  for glutamine and  $R^2 = 0.13$ ,  $p > 0.05$  for citrate)  
 673 **c.** Relative survival of invasive endemic species after coalescence ( $R^2 = 0.16$ ,  $p < 0.05$  for glutamine and  $R^2 = 0.04$ ,  $p > 0.05$   
 674 for citrate).



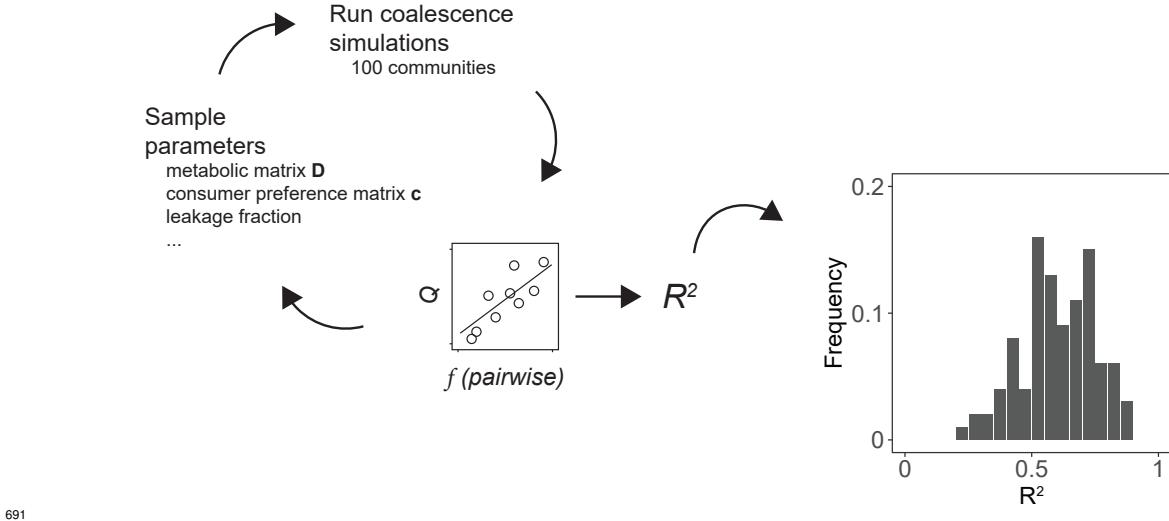
676

677 **Figure S3. Dominant species have limited effects on coalescence outcomes quantification.** We repeated the analyses shown  
 678 in [Figure 2b](#) and [Figure S2](#), but this time we removed the dominants from the compositional data prior to quantifying community  
 679 distances. The trends observed before are maintained. **a.** Relative Bray-Curtis similarity ( $R^2 = 0.20$ ,  $p < 0.01$  for glutamine  
 680 and  $R^2 = 0.34$ ,  $p < 0.005$  for citrate) **b.** Relative Jensen-Shannon similarity ( $R^2 = 0.24$ ,  $p < 0.005$  for glutamine and  $R^2 = 0.36$ ,  
 681  $p < 0.005$  for citrate) **c.** Relative Jaccard similarity ( $R^2 = 0.09$ ,  $p > 0.05$  for glutamine and  $R^2 = 0.11$ ,  $p > 0.05$  for citrate) **d.**  
 682 Relative survival of invasive endemic species after coalescence ( $R^2 = 0.18$ ,  $p < 0.05$  for glutamine and  $R^2 = 0.08$ ,  $p > 0.05$  for  
 683 citrate).

685



686 **Figure S4. Bottom-up ecological co-selection is not observed when species have similar metabolic architectures.** We  
 687 ran simulations of community coalescence following the same procedure described in the main text, but this time we used a  
 688 dense metabolic matrix (left panel, sparsity = 0.05 in the Community Simulator package [35]) or a species-unspecific metabolic  
 689 matrix (right panel,  $D_{i\beta\alpha} = D_{j\beta\alpha}$  for all  $i, j, \alpha$  and  $\beta$ , see Box 1). Virtually no bottom-up co-selection is observed in either case.



691

692 **Figure S5. Top-down co-selection is robust in the MicroCRM.** We ran simulations of community coalescence following the  
 693 same procedure described in the main text, but randomly sampling the following parameters of the MicroCRM uniformly within  
 694 the indicated ranges: number of species per community sampled at initialization ( $S$ ) between 10 and 90; total number of species  
 695 families between 1 and 5; total number of species per family ( $S_A$ ) between 400 and 1200; total number of generalist species  
 696 ( $S_{gen}$ ) between 100 and 380; specialist species preference strength ( $q$ ) between 0.5 and 1; number of resource classes between 3  
 697 and 8; number of resources per class ( $M_A$ ) between 3 and 17; leakage fraction ( $l_a$ ) between 0.45 and 0.95; maintenance cost ( $m_i$ )  
 698 between 0 and 0.2; standard deviation of the sum of consumption rates ( $\sigma_c$ ) between 2 and 4; sparsity of the metabolic matrix  
 699 between 0.05 and 0.95; fraction of secretion flux to resources of the same type ( $f_s$ ) or to waste resources ( $f_w$ ) between 0.05 and  
 700 0.45. The metabolic matrix was randomly chosen to be such that  $D_{i\beta\alpha} = D_{j\beta\alpha}$  or such that  $D_{i\beta\alpha}$  need not be equal to  $D_{j\beta\alpha}$  for all  
 701 species ( $i, j$ ) and resources ( $\alpha, \beta$ ). Both options were given equal probabilities. For each randomly sampled set of parameters,  
 702 we ran 100 simulations and quantified the ability of the pairwise competition of dominants to predict coalescence outcomes  
 703 (see Figure 2), i.e. the strength of top-down co-selection in each regime. The histogram shows that top-down co-selection is  
 704 robust throughout the parameter space in the MicroCRM.