

# Top-down versus 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

The abstract goes here.

## 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] or by aerially dispersed bacteria and fungi [5]. Gut microbiomes can invade external communities through the host animal secretions [6], and the skin microbiota is also subject to invasions when making contact with environmental sources of microbes [7].

The phenomenon by which entire microbiomes invade one another has been termed *community coalescence* [8]. Ecologists have long contemplated the idea that interactions between multiple co-invading species can produce correlated invasional outcomes [8–18]. 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. Early mathematical models of community-community invasions [9, 19] as well as more recent work [20–23] 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 [15, 24]. The situation where ecological partners in the invading community recruit each other into the final coalesced community has been called *ecological co-selection* [24, 25].

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 [26] and the role played by the less abundant species has long been subject to debate [27]. Laboratory cultures have also been found to contain uneven distributions of multiple strains that feed off the metabolic secretions of the dominant species [28, 29]. The fate of these sub-dominant taxa may be dependent on the invasional success of their dominant species, or, alternatively, the dominant itself may owe its dominance (at least in part) to cross-feeding or other forms of facilitation from the rarer members of the population. These scenarios would give rise to “top-down” or “bottom-up” community cohesiveness, respectively. Either of these forms of co-selection could, in principle, be positive (recruitment) or negative (antagonism), as illustrated in Figure 1e. Which of these situations are typically found in nature? Previous theoretical and computational studies suggest that the answer is determined by the type and strength of the interactions of the community members with one another and with the environment [20, 22, 23], but addressing this question has been experimentally challenging in the past [24, 25].

In previous work, we have shown that a large amount of soil and plant microbiomes can be cultured *ex situ* in synthetic minimal environments with a single supplied limiting resource under serial growth-dilution cycles [29] (Figure 1a-b). Under these conditions, environmental microbiomes spontaneously re-assemble into complex multi-species communities sustained by dense cross-feeding facilitation networks [29]. In addition, and just like

48 in natural consortia, species abundance distributions in these communities are generally long-tailed and uneven  
49 (Figure 1d and Figure S1), with the dominant (most abundant) species typically comprising most of the biomass  
50 (median = 46%, Figure S1). Because these communities are easy to manipulate and grow in high throughput, and  
51 are largely made up by culturable members, they represent good test cases to investigate ecological co-selection  
52 during community coalescence. Here we focus on the dominants and ask whether they can co-select or be co-  
53 selected by the sub-dominant species in their communities (henceforth referred to as their *cohorts*, Figure 1c).

54 Our results indicate that co-selection varies in direction and strength depending on the supplied limiting re-  
55 source. This primary resource, in turn, has been shown to shape the structure and interactions of the communities  
56 [30]. We observe that, when top-down co-selection is weak, bottom-up co-selection can be very strong, with  
57 positive co-selection being far more common than negative co-selection. We then turn to a Microbial Consumer-  
58 Resource Model (MicroCRM) [29, 31, 32] that is able to capture the dynamics of microbial communities domi-  
59 nated by metabolic interactions, as is the case for the ones assembled in our experimental conditions [29, 30]. We  
60 show that the empirically observed trends in ecological co-selection are reproduced with minimal model assump-  
61 tions, and that the recurrence of top-down and bottom-up co-selection is determined by the configuration of the  
62 cross-feeding networks in the MicroCRM. Our findings indicate that collective interactions play an important role  
63 at dictating community structure during coalescence.

## 64 Results & Discussion

65 We collected eight natural microbiomes from different soil and plant environmental samples (Figure 1a) and used  
66 them to inoculate our synthetic communities, which were stabilized in serial batch-culture biorreactors for 84  
67 generations in synthetic minimal media containing either glutamine or citrate as the only supplied carbon source  
68 (Figure 1b, Methods: Stabilization of environmental communities in simple synthetic environments). We chose  
69 these two carbon sources because they are metabolized through different pathways in bacteria [33, 34], and we  
70 hypothesize that communities assembled in either resource will be supported by cross-feeding networks of distinct  
71 sets of metabolites [29, 30] thus leading to potentially variable degrees of community cohesiveness and coales-  
72 cence outcomes [18, 20, 21, 23]. We isolated the dominant species of every community (Methods: Isolation  
73 of dominant species) and identified them by Sanger-sequencing their 16S rRNA gene (Methods: Determination  
74 of community composition by 16S sequencing), which correctly matched the dominant Exact Sequence Variant  
75 (ESV) [35, 36] found through community-level 16S Illumina sequencing (Figure S1). These dominants remained  
76 at high frequency after seven additional transfers with the exception of two of the citrate communities and one  
77 of the glutamine communities (where the dominants were presumably a transiently dominating species) that were  
78 excluded from further analysis (Figure S1). Similarly, pairs of communities where the dominants shared a same  
79 16S sequence and had similar colony morphology were excluded (Figure S1).

### 80 Top-down ecological co-selection

81 If communities being coalesced were highly cohesive from the top-down, the dominant species would co-select the  
82 rarer members of its community during coalescence (Figure 1e, left panels). In this scenario, we would expect the  
83 outcome of community coalescence to be predicted by which of the two dominants is most competitive in pairwise  
84 competition. Analogously, competition between dominants should be affected only weakly by the presence or ab-  
85 sence of the cohorts, that would play a passive role under these conditions. To test this hypothesis, we performed  
86 all pairwise competitions between dominant species in glutamine and citrate environments by mixing them 1:1 on  
87 their native media and propagating the cultures for seven serial transfers, roughly 42 generations (Methods: Coa-  
88 lescence, competition and invasion experiments). We then carried out all possible pairwise community coalescence  
89 experiments by mixing equal volumes of the communities and propagating the resulting cultures for seven extra  
90 transfers (Figure 1f). The frequencies of all species in both community-community and dominant-dominant com-  
91 petitions were determined by 16S Illumina sequencing (Methods: Determination of community composition by  
92 16S sequencing).

93 We found that, for communities assembled in the glutamine environment, the relative frequency of a dominant  
94 against another in head-to-head pairwise competition is barely predictive of its relative frequency against that  
95 same other dominant when the cohorts are present too, i.e. during community coalescence (Figure 2a red dots,  
96  $R^2 = 0.04, p > 0.05$ ). This correlation is significantly higher for the citrate communities (Figure 2a blue dots,  
97  $R^2 = 0.83, p < 10^{-8}$ ). This suggests that, in the glutamine environments, head-to-head competition of dominants  
98 is heavily influenced by higher order effects introduced by the rare taxa of the communities. On the other hand, the  
99 cohorts seem to play a more passive role in the citrate environments. To test the effects of top-down co-selection  
100 at the community level, we quantified the distances between the invasive and coalesced communities using the  
101 relative Bray-Curtis similarity (Methods: Metrics of community distance) and compared them to the outcomes of

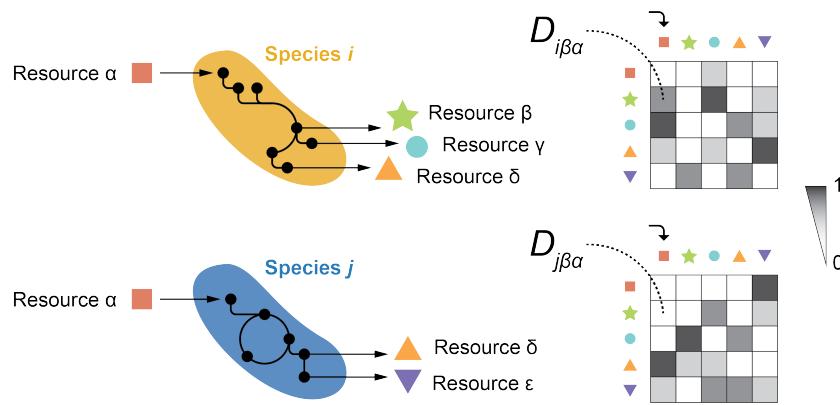
102 the pairwise competitions between dominants alone. We again noticed differences between glutamine and citrate  
103 communities: for the former, the pairwise competitive ability of an invasive dominant is only weakly predictive of  
104 the performance of the invasive community in coalescence ([Figure 2b](#) left panel,  $R^2 = 0.15$ ,  $p < 0.05$ ). For the  
105 latter, the structure of the coalesced communities tends to be more strongly dictated by the result of the dominant-  
106 dominant competition ([Figure 2b](#) middle panel,  $R^2 = 0.57$ ,  $p < 10^{-4}$ ). Alternative quantifications of community  
107 distance yield similar results, with weaker effects when the metric used accounts only for the presence/absence  
108 of specific species and not for their relative abundance in the communities ([Figure S2](#)). All these metrics include  
109 the presence of the dominant species themselves. To better disentangle the effect that these dominants have on the  
110 other members of their communities, we repeated the analysis this time excluding the dominant species from the  
111 compositional data, finding that our results still hold ([Figure S3](#)).

112 Together, these observations suggest that the strength of top-down co-selection depends on the environment  
113 where communities are assembled and coalescence takes place. Communities stabilized with citrate as the primary  
114 supplied resource display a strong degree of top-down cohesiveness, with the fates of the sub-dominant species  
115 determined to a large extent by dominant-dominant pairwise competition. This competition is, in turn, only weakly  
116 affected by the presence of the cohorts. For glutamine communities, although some level of top-down co-selection  
117 is consistent with our data, the cohorts do not appear to just be passively responding to their dominants but rather  
118 playing an active role in community coalescence.

119 To investigate the determinants of top-down co-selection and the factors modulating its strength, we ran a  
120 set of simulations of community coalescence. We used a Microbial Consumer-Resource Model (MicroCRM)  
121 [[29](#), [31](#)] as implemented in the Community Simulator package for Python [[32](#)] ([Box 1](#)). We chose this modeling  
122 framework because communities assembled under our experimental conditions (natural microbiomes re-assembled  
123 into multispecies communities through serial growth-dilution cycles in synthetic minimal media with a single  
124 carbon source) have been shown to be sustained by dense metabolic cross-feeding networks [[29](#), [30](#)] for which the  
125 MicroCRM provides a good description. Based on previous experimental work [[30](#), [38](#), [39](#)], we allowed different  
126 species in our model to secrete different sets of byproducts when metabolizing a same resource ([Box 1](#)). To  
127 reproduce our experimental protocol *in silico*, we first generated a library of resources and two non-overlapping  
128 pools of species. Each pool was used to seed a collection of 100 invasive and 100 resident communities respectively  
129 by randomly choosing 50 species and allowing them to stabilize through 20 growth-dilution cycles. We then mixed  
130 these stable communities in pairs to simulate our coalescence and dominant-dominant competition experiments  
131 ([Methods: Simulations](#)). We found that the MicroCRM was able to capture a correlation between the head-to-head  
132 pairwise competition of dominants and the outcome of community coalescence ([Figure 2b](#), right panel), further  
133 supporting the idea that top-down ecological co-selection can consistently emerge from metabolic interactions  
134 across species.

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

136 The Microbial Consumer-Resource Model (MicroCRM) [29, 31, 32] is a modeling framework based on the  
 137 classic MacArthur's consumer resource model [37]. It encodes the dynamics of a system with  $S$  species and  
 138  $M$  resources in terms of a consumer preference matrix  $\mathbf{c}$  and a metabolic matrix  $\mathbf{D}$ , with an additional set of  
 139 parameters controlling the species maintenance costs ( $m_i$  for species  $i$ ), the resource energy densities ( $w_\alpha$  for  
 140 resource  $\alpha$ ), the energy to growth rate conversion factor ( $g_i$  for species  $i$ ) and the leakage fraction, i.e. the  
 141 amount of energy lost as byproducts when a resource is consumed ( $l_\alpha$  for resource  $\alpha$ ). The element  $c_{ia}$  of the  
 142 consumer preference matrix represents the uptake rate of resource  $\alpha$  by species  $i$  (although the relationship  
 143 between  $c_{ia}$  and the uptake rate can be more complex in modeling scenarios that are not considered here, see  
 144 [29, 31, 32]). Experimental evidence suggests that individual species can secrete different sets of metabolites  
 145 to the environment when growing on a same primary resource [30, 38, 39]. Thus, we define  $\mathbf{D}$  as a three-  
 146 dimensional matrix where the element  $D_{i\beta\alpha}$  represents the energy flux in the form of resource  $\beta$  that is secreted  
 147 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  
 148 below).



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

$$152 \frac{dN_i}{dt} = g_i N_i \left[ \sum_{\alpha} (1 - l_{\alpha}) w_{\alpha} c_{i\alpha} R_{\alpha} - m_i \right] \quad (1)$$

$$153 \frac{dR_{\alpha}}{dt} = - \sum_j N_j c_{j\alpha} R_{\alpha} + \sum_j \sum_{\beta} N_j c_{j\beta} R_{\beta} \left[ l_{\beta} D_{j\beta\alpha} \frac{w_{\beta}}{w_{\alpha}} \right] \quad (2)$$

154 These equations can take slightly different forms in certain cases, e.g. if the primary resource is supplied  
 155 continuously instead of at the beginning of each growth cycle [31, 32], but they are a good approximation for the  
 156 community dynamics between consecutive serial dilutions in our setup. Here, we assembled *in silico* com-  
 157 munities by randomly sampling a set of species from a pool, then integrating equations 1 and 2, diluting the  
 158 final abundances, replenishing the primary resource, and repeating the process until generational equilib-  
 159 rium was achieved (Methods: Simulations). Coalescence simulations were carried out following the same logic,  
 160 this time seeding the coalesced communities by mixing the invasive and resident ones instead of sampling  
 161 from a species pool.

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

169 Our data indicates that the primary resource supplied to the communities can modulate the effect that the cohorts  
170 have in the dominants pairwise competition ([Figure 2a](#)) and the strength of top-down co-selection ([Figure 2b](#),  
171 left and middle panels). The fact that our model captures these trends suggests that this might be a result of the  
172 metabolic interactions between community members, including the rarer taxa. To investigate the potential role  
173 of the cohorts in coalescence, i.e. whether the dominants may be co-selected for or against by them ([Figure 1e](#),  
174 right panels), we ran a new set of simulations this time invading resident communities with the dominants alone  
175 ([Methods: Simulations](#)). We compared the invasion success of the dominants in isolation with respect to our  
176 previous simulations where they invaded accompanied by their cohorts. The invasion success of the dominants  
177 was quantified by their relative abundance in the final stabilized communities. Whenever positive bottom-up  
178 ecological co-selection is strong, we expect to see dominants reaching higher invasion success with their cohorts  
179 than by themselves ([Figure 3b](#), green shaded region). On the other hand, a high degree of bottom-up antagonism  
180 would result in dominants invading more effectively alone than in the presence of their cohorts ([Figure 3b](#), red  
181 shaded region). Alternatively, if both forms of bottom-up co-selection are weak, we would see a similar invasion  
182 success regardless of the presence or absence of the cohort ([Figure 3b](#), gray shaded region).

183 [Figure 3b](#) shows that, in the simulations, many dominants could not invade on their own (or could only do  
184 so at very low final relative abundances, below 0.1) but were able to reach high frequencies when they were  
185 accompanied by their cohorts. This indicates that positive bottom-up co-selection is frequent and potentially very  
186 strong, while negative bottom-up co-selection is far more uncommon. We then asked whether the ability of the  
187 pairwise competition of dominants to predict coalescence outcomes was dependent on the strength of bottom-up  
188 co-selection. We divided our simulations into two subsets: the first one was comprised of the instances where  
189 positive bottom-up co-selection was strong (i.e. dots in the green shaded region of [Figure 3b](#)), the second set  
190 included all other cases (dots near the diagonal of [Figure 3b](#)). We reexamined our original simulations and plotted  
191 the frequency of the invasive dominant in pairwise competition versus the relative similarity between the invasive  
192 and coalesced communities, i.e. the same plot as in [Figure 2b](#), for each subset. We found that when bottom-up  
193 positive co-selection is strong, the pairwise competition of dominants is not predictive of coalescence outcomes  
194 ([Figure 3c](#), left panel) and vice-versa ([Figure 3c](#), right panel).

195 We then asked whether this trend was also observed *in vitro*. We went back to our laboratory cultures and car-  
196 ried out a new round of experiments where we invaded the resident communities with the invasive dominants alone  
197 ([Methods: Coalescence, competition and invasion experiments](#)). After stabilization, we quantified species abun-  
198 dance through 16S Illumina sequencing ([Methods: Determination of community composition by 16S sequencing](#)).  
199 Again, we observed that bottom-up co-selection is far more common in its positive than in its negative form ([Fig-  
200 ure 3d](#)). Interestingly, bottom-up recruitment appears to be more frequent in the glutamine environments than in  
201 the citrate ones, consistent with our hypothesis that metabolic interactions among species are key in determining  
202 the strength and direction of ecological co-selection. We then repeated our analysis in [Figure 3c](#), this time splitting  
203 our data according to the observed strength of bottom-up co-selection instead of the primary carbon source as we  
204 had done in [Figure 2b](#). Our findings were in line with the model prediction: pairwise competition between domi-  
205 nants is only predictive of coalescence outcomes if bottom-up co-selection is weak ([Figure 3e](#),  $R^2 = 0.07$ ,  $p > 0.05$   
206 when bottom-up co-selection is strong;  $R^2 = 0.37$ ,  $p < 10^{-4}$  when bottom-up co-selection is weak).

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

208 To better understand the underlying mechanisms that govern the emergence of ecological co-selection, we devel-  
209 oped a minimal model of community coalescence. This model is comprised of two communities with two species  
210 each as illustrated in [Figure 4a](#). Within each community, the dominant species ( $s_1$  and  $s_3$  in the resident and inva-  
211 sive communities respectively) is able to utilize the primary resource that is supplied externally ( $R_1$ ). As a result  
212 of  $R_1$  metabolism, the dominants secrete a single byproduct ( $R_2$  and  $R_4$  respectively) off which the sub-dominants  
213 ( $s_2$  and  $s_4$  respectively) can feed. Finally, these sub-dominants secrete an additional resource ( $R_3$  and  $R_5$  respec-  
214 tively) that can again be metabolized by the corresponding dominants. This is the simplest structure where both  
215 top-down and bottom-up cohesiveness can emerge, with their relative strengths being determined by how effec-  
216 tively the species can utilize each other's metabolic byproducts. Several parameters can modulate the intensity and  
217 direction of community cohesiveness, and even in this simple model there can be complex interactions between  
218 them. For instance, a sub-dominant that is very efficient at metabolizing the dominant's secretions will therefore  
219 be a strong producer of secondary byproducts, in turn increasing the metabolic flow towards the dominant itself.  
220 To keep both communities symmetrical and maintain tractability of the model's behavior, we chose to make the two  
221 sub-dominant species equally able consumers of the byproduct of their respective dominants ([Methods: Minimal  
222 model](#)). We also made it so species secrete only one type of byproduct each, i.e. the metabolic matrix is binary in  
223 this case (see [Box 1](#)). The secretions of the resident and invasive communities are non-overlapping: this ensures

224 that coalescence always results in one of the two communities taking over and completely excluding the other one,  
225 which facilitates the interpretation of the model's outcomes. This means that the relative similarity between the  
226 invasive and the coalesced communities ( $Q$ , see [Methods: Metrics of community distance](#)) can only be 0 (if the  
227 resident community dominates in coalescence) or 1 (if the opposite happens).

228 With these constraints, the model is specified by four rates: the resident dominant's uptake rate of the primary  
229 resource ( $c_{11}$ ) and of the secondary byproduct secreted by  $s_2$  ( $c_{13}$ ), as well as the invasive dominant's uptake rate  
230 of the primary resource ( $c_{31}$ ) and of the secondary byproduct secreted by  $s_4$  ( $c_{35}$ ). Analogous to what we did for  
231 our initial experiments and simulations of community coalescence, we first stabilized each minimal community  
232 separately through serial growth-dilution cycles, and then mixed them 1:1 and stabilized again until generational  
233 equilibrium was achieved. Equations 1 and 2 in [Box 1](#) describe the dynamics of this system between consecutive  
234 dilutions with the corresponding choices for the parameter values ([Methods: Minimal model](#)). We ran simulations  
235 for both community-community competitions and invasions from the dominant alone. We considered four  
236 different scenarios illustrated in [Figure 4b-e](#): 1) both communities are cohesive from the top-down but not from  
237 the bottom-up, 2) the resident community is cohesive both from the top-down and from the bottom-up while the  
238 resident community is cohesive from the top-down only, 3) the resident community is cohesive from the top-down  
239 only while the invasive community is cohesive both from the top-down and from the bottom-up, and 4) both  
240 communities are cohesive from the top-down and from the bottom-up. The outcomes in each case are discussed  
241 below.

242 **Top-down resident versus top-down invasive.** This is the simplest scenario. When both communities are  
243 cohesive strictly from the top-down ( $c_{13} = 0$  and  $c_{35} = 0$ ), coalescence outcomes are determined by which domi-  
244 nant is a better consumer of the primary resource. The most competitive dominant excludes the least competitive  
245 one ( $s_1$  excludes  $s_3$  if  $c_{11} > c_{31}$  and vice-versa, [Figure 4b](#)) and, because the cross-feeding networks of the invasive  
246 and resident communities are non-overlapping, it co-selects its sub-dominant exclusively. The invasion success  
247 of the invasive dominant ( $s_3$ ) is unaffected by the presence or absence of its sub-dominant ( $s_4$ ) since it is not be-  
248 ing cross-fed by it. Thus, again  $s_3$  is only able to invade alone when it outcompetes  $s_1$  for the primary resource  
249 ( $c_{11} > c_{31}$ ).

250 **Bidirectional resident versus top-down invasive.** When the resident community is cohesive both top-down  
251 and bottom-up ( $c_{13} > 0$ ) and the invasive community is only cohesive from the top-down ( $c_{35} = 0$ ), coalescence  
252 outcomes are determined by the strength of bottom-up cohesiveness in the resident community, i.e. the magnitude  
253 of  $c_{13}$  in the minimal model ([Figure 4c](#)). If  $c_{13}$  is small (weak bottom-up cohesiveness), the system behaves as in  
254 the previous scenario, virtually as if both communities were only cohesive from the top-down. However, as  $c_{13}$   
255 increases (strong bottom-up cohesiveness) the resident community becomes more resistant to invasion. For large  
256 values of  $c_{13}$ , the invasive dominant being more competitive than the resident dominant for the primary resource  
257 does not ensure invasion success. Since the latter can also be cross-fed by its sub-dominant (through  $R_3$ ), it does  
258 not rely on the primary resource only ( $R_1$ ) to outcompete the invader. The same is true when the invasive dominant  
259 invades alone.

260 **Top-down resident versus bidirectional invasive.** More outcomes become possible if the resident community  
261 is cohesive from the top-down only ( $c_{13} = 0$ ) and the invasive one is cohesive in both directions ( $c_{35} > 0$ ). Like in  
262 the first scenario we considered, the invasive dominant will only be successful by itself if it is a better competitor  
263 for  $R_1$  than the resident dominant ( $c_{31} > c_{11}$ ). However, if the invasive sub-dominant is also present, the invasion  
264 can be successful even if  $c_{31} < c_{11}$  as long as the cross-feeding from  $s_4$  (invasive sub-dominant) to  $s_3$  (invasive  
265 dominant) is strong enough, i.e. if  $c_{35}$  sufficiently large ([Figure 4d](#)). In this scenario, we can see cases where  
266 the invasive dominant cannot invade in isolation but the invasive community still dominates during coalescence,  
267 similar to what we observed in our experiments and initial simulations.

268 **Bidirectional resident versus bidirectional invasive.** If both communities are cohesive from the top-down  
269 and from the bottom-up ( $c_{13} > 0$  and  $c_{35} > 0$ ), coalescence outcomes are generically determined by the relations  
270 of strength between all forms of cohesiveness. In limit cases we can recover scenarios considered previously, e.g.  
271 if  $c_{13} \gg c_{35}$  the system behaves as if the resident community was cohesive from the bottom-up but the invasive  
272 community was not. Situations where the invasive dominant can invade when accompanied by its sub-dominant  
273 despite being unable to do so by itself are again possible if the invasive community is more strongly cohesive from  
274 the bottom-up than the resident community ([Figure 4e](#)).

275 This minimal coalescence model shows that bottom-up co-selection of a dominant that is unable to invade by  
276 itself is only possible if said dominant is strongly cross-fed by its cohort. The model also demonstrates that this  
277 behavior can only happen if the byproducts of the cohorts are to some extent non-overlapping and preferentially  
278 utilized by the respective dominants. Otherwise, if both species can utilize all secretions, bottom-up co-selection  
279 becomes unspecific with respect to the dominant identity.

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

281 How do the ideas above scale to more complex and diverse communities?

282 We reason that the choice of species-specific metabolic architectures is necessary to potentially generate cohe-  
283 siveness at the community level during coalescence. If the secretions of all species were identical (or only slightly  
284 different), higher order cross-feeding effects would be very unspecific: the establishment of new invasive species  
285 –given that they could outcompete resident taxa within their metabolic niches, i.e. more effectively feed off the  
286 same resource or set of resources as them– would not alter (or only do so moderately) the metabolic flows through  
287 the rest of the community’s cross-feeding network. On the other hand, said network could undergo a profound  
288 and further-reaching restructuring if the invasive species secreted very different sets of metabolites with respect  
289 to the resident ones, potentially disabling existing niches and/or enabling new ones where more invaders could  
290 be co-selected. For a similar reason, we argue that the sparsity of the metabolic matrix could also modulate the  
291 emergence of cohesiveness in the face of coalescence. A dense metabolic matrix corresponds to a situation where  
292 all species secrete a wide variety of byproducts. New-coming invasive species that secrete similar byproducts as  
293 resident ones –even if they do so in different relative amounts– might only induce moderate quantitative changes in  
294 the metabolic fluxes. But if the sets of secretions are qualitatively different, co-selection of species adapted to each  
295 of those sets becomes possible. These ideas are supported by experimental observations suggesting that species  
296 with a history of coexistence make up cohesive communities with highly specific cross-feeding configurations  
297 [28–30].

298 **Conclusions**

299 Understanding the mechanisms underlying the responses of microbial communities to invasions is an essential  
300 but poorly understood question in microbial ecology [8]. Theory has suggested that communities may exhibit an  
301 emergent cohesiveness [9, 15, 20, 21], leading to members of the same community recruiting one another during  
302 community-community invasions. Our results provide direct experimental evidence of ecological co-selection in  
303 a large number of community coalescence experiments, and highlight the critical role played by the rarer, sub-  
304 dominant species in the generation of community cohesiveness.

305 Our data suggests that the strength and direction of ecological co-selection is modulated by the underlying  
306 metabolic networks that shape the structure of communities assembled in synthetic minimal conditions [29, 30].  
307 This network is in turn regulated by the supplied primary carbon source in our minimal laboratory conditions.  
308 This idea is supported by the observation that a Microbial Consumer-Resource Model captures the trends observed  
309 experimentally when we enable a large variation in the metabolic fluxes across species. The model also predicts a  
310 trade-off between the strength of bottom-up co-selection and the ability of dominant-dominant pairwise competi-  
311 tion to dictate coalescence outcomes, which we have confirmed experimentally. These observations, together with  
312 previous results in different systems [24] as well as theoretical predictions [9, 19–23], suggest that collective inter-  
313 actions between microbes and the environment should be generically expected to produce ecological co-selection  
314 during community coalescence.

315 Additional work will be necessary to further clarify the relationship between metabolic feedbacks, community  
316 cohesiveness and ecological co-selection. The experimental system that we introduced in this work can be eas-  
317 ily expanded so that large numbers of community coalescence experiments can be carried out in parallel. It thus  
318 represents a promising tool to explore the properties of microbial community coalescence in high throughput and  
319 test quantitative theories about its role in microbiome assembly. On the other hand, coalescence of communities  
320 under different settings (e.g. in spatially structured environments) might be governed by additional factors. Un-  
321 derstanding them and quantifying their relative contributions in natural communities remains an open question in  
322 ecology.

323 **Methods**

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

325 Communities were stabilized *ex situ* as described in [29]. In short, environmental samples (soil, leaves...) within  
326 one meter radius in eight different geographical locations were collected with sterile tweezers or spatulas into 50mL  
327 sterile tubes ([Figure 1a](#)). One gram of each sample was allowed to sit at room temperature in 10mL of phosphate  
328 buffered saline (1×PBS) containing 200µg/mL cycloheximide to suppress eukaryotic growth. After 48h, samples  
329 were mixed 1:1 with 80% glycerol and kept frozen at -80°C. Starting microbial communities were prepared by  
330 scrapping the frozen stocks into 200µL of 1×PBS and adding a volume of 4µL to 500µL of synthetic minimal  
331 media (1×M9) supplemented with 200µg/mL cycloheximide and 0.07 C-mol/L glutamine or sodium citrate as  
332 the carbon source in 96 deep-well plates (1.2mL; VWR). Cultures were then incubated still at 30°C to allow  
333 for re-growth. After 48h, samples were fully homogenized and biomass increase was followed by measuring the  
334 optical density (620nm) of 100µL of the cultures in a Multiskan FC plate reader (Thermo Scientific). Communities  
335 were stabilized [29] by passaging 4µL of the cultures into 500µL of fresh media (1×M9 with the carbon source)  
336 every 48h for a total of 12 transfers at a dilution factor of 1:100, roughly equivalent to 80 generations per culture  
337 ([Figure 1b](#)). Cycloheximide was not added to the media after the first two transfers.

338 **Isolation of dominant species**

339 For each community, the most abundant colony morphotype at the end of the ninth transfer was selected ([Figure 1c](#)),  
340 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  
341 corresponding synthetic minimal media and allowed to regrow at 30°C for 48h. Dominants were then identified,  
342 inoculated into 500µL of fresh media and incubated still at 30°C for 48h. After this period, the communities  
343 stabilized for eleven transfers and the isolated dominants were ready for the competition experiments at the onset  
344 of the twelfth transfer.

345 **Coalescence, competition and invasion experiments**

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

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

353 The sequencing protocol was identical to that described in [29]. Community samples were collected by spinning  
354 down at 3500rpm for 25min in a bench-top centrifuge at room temperature; cell pellets were stored at -80°C  
355 before processing. To maximize Gram-positive bacteria cell wall lysis, the cell pellets were re-suspended and  
356 incubated at 37°C for 30min in enzymatic lysis buffer (20mM Tris-HCl, 2mM sodium EDTA, 1.2% Triton X-100)  
357 and 20mg/mL of lysozyme from chicken egg white (Sigma-Aldrich). After cell lysis, the DNA extraction and  
358 purification was performed using the DNeasy 96 protocol for animal tissues (Qiagen). The clean DNA in 100µL  
359 elution buffer of 10mM Tris-HCl, 0.5mM EDTA at pH 9.0 was quantified using Quan-iT PicoGreen dsDNA Assay  
360 Kit (Molecular Probes, Inc.) and normalized to 5ng/µL in nuclease-free water (Qiagen) for subsequent 16S rRNA  
361 Illumina sequencing. 16S rRNA amplicon library preparation was performed following a dual-index paired-end  
362 approach [40]. Briefly, PCR amplicon libraries of V4 regions of the 16S rRNA were prepared sing dual-index  
363 primers (F515/R805), then pooled and sequenced using the Illumina MiSeq chemistry and platform. Each sample  
364 went through a 30-cycle PCR in duplicate of 20µL reaction volumes using 5ng of DNA each, dual index primers,  
365 and AccuPrime Pfx SuperMix (Invitrogen). The thermocycling procedure includes a 2min initial denaturation step  
366 at 95°C, and 30 cycles of the following PCR scheme: (a) 20-second denaturation at 95°C, (b) 15-second annealing  
367 at 55°C, and (c) 5-minute extension at 72°C. The duplicate PCR products of each sample were pooled, purified,  
368 and normalized using SequalPrep PCR cleanup and normalization kit (Invitrogen). Barcoded amplicon libraries  
369 were then pooled and sequenced using Illumina Miseq v2 reagent kit, which generated 2×250bp paired-end reads  
370 at the Yale Center for Genome Analysis (YCGA). The sequencing reads were demultiplexed on QIIME 1.9.0 [41].  
371 The barcodes, indexes, and primers were removed from raw reads, producing FASTQ files with both the forward  
372 and reverse reads for each sample, ready for DADA2 analysis [36]. DADA2 version 1.1.6 was used to infer unique  
373 biological exact sequence variants (ESVs) for each sample and naïve Bayes was used to assign taxonomy using  
374 the SILVA version 123 database [42, 43].

375 **Metrics of community distance**

376 Beta-diversity indexes between the invasive and coalesced communities or the resident and coalesced communities  
 377 were computed using various similarity metrics. For two arbitrary communities with ESV abundances represented  
 378 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  
 379  $i$ th ESV in each community respectively and  $S$  is the total number of ESVs), the Bray-Curtis similarity  $BC(\mathbf{x}, \mathbf{y})$   
 380 is calculated as [44]

$$BC(\mathbf{x}, \mathbf{y}) = \sum_i \min(x_i, y_i) \quad (3)$$

381 The Jensen-Shannon similarity  $JS(\mathbf{x}, \mathbf{y})$  is defined as one minus the Jensen-Shannon distance (which is, in turn,  
 382 the square root of the Jensen-Shannon divergence [45])

$$JS(\mathbf{x}, \mathbf{y}) = 1 - \sqrt{\frac{1}{2}KL(\mathbf{x}, \mathbf{m}) + \frac{1}{2}KL(\mathbf{y}, \mathbf{m})} \quad (4)$$

383 where  $\mathbf{m} = (\mathbf{x} + \mathbf{y}) / 2$  and  $KL$  denotes the Kullback-Leibler divergence [46]

$$KL(\mathbf{x}, \mathbf{y}) = \sum_i x_i \log_2 \left( \frac{x_i}{y_i} \right) \quad (5)$$

384 Using base-two logarithms ensures that the metric is bounded between 0 and 1. The Jaccard similarity is given by  
 385  $J(\mathbf{x}, \mathbf{y})$  [47]

$$J(\mathbf{x}, \mathbf{y}) = \frac{|\mathbf{x} \cap \mathbf{y}|}{|\mathbf{x} \cup \mathbf{y}|} \quad (6)$$

386 Additionally, we quantified coalescence outcomes by examining the fraction of the endemic cohort of the original  
 387 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  
 388 found in  $\mathbf{y}$ .

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

$$Q(\mathbf{x}_I, \mathbf{x}_R, \mathbf{x}_C) = \frac{F(\mathbf{x}_I, \mathbf{x}_C)}{F(\mathbf{x}_I, \mathbf{x}_C) + F(\mathbf{x}_R, \mathbf{x}_C)} \quad (7)$$

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

394 **Simulations**

395 We used the Community Simulator package [32] and included new features for our simulations. In the package,  
 396 species are characterized by their resource uptake rates ( $c_{i\alpha}$  for species  $i$  and resource  $\alpha$ ), and they all share a  
 397 common metabolic matrix  $\mathbf{D}$ . The element  $D_{\alpha\beta}$  of this matrix represents the fraction of energy in the form of  
 398 resource  $\alpha$  secreted when resource  $\beta$  is consumed. Here we implemented a new operation mode in which species  
 399 can secrete different metabolites (and/or in different abundances) when consuming a same resource. We call  $D_{i\alpha\beta}$  to  
 400 the fraction of energy in the form of resource  $\alpha$  secreted by species  $i$  when consuming resource  $\beta$ . In the Community  
 401 Simulator underlying Microbial Consumer-Resource Model, this means that the energy flux  $J_{i\beta}^{\text{out}}$  [29, 31] now takes  
 402 the form

$$J_{i\beta}^{\text{out}} = \sum_{\alpha} D_{i\beta\alpha} l_{\alpha} J_{i\alpha}^{\text{in}} \quad (8)$$

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

405 For our simulations, we first generated a library of 2400 species divided into three specialist families of 800  
 406 members each and a generalist family of 240 members. We split this library into two non-overlapping pools of  
 407 1320 species each. We randomly sampled 50 species from each pool in equal ratios to seed 100 resident and 100  
 408 invasive communities respectively. We then let grow and diluted the communities serially, replenishing the primary  
 409 resource after each dilution. We repeated the process 20 times to ensure generational equilibrium was achieved  
 410 [29]. We then performed the *in silico* experiments by using the generationally stable communities to seed 100  
 411 coalesced communities that were again stabilized as described previously. Similarly, we identified the dominant

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

414 Most other parameters were set to the defaults of the original Community Simulator package, with the only  
 415 exception of the maintenance costs ( $m$ ) which are set to zero for all species (equivalent to assuming cell death is  
 416 negligible through the duration of our growth cycles) and the sparsity of the metabolic matrices ( $s$ ) which is set to  
 417 0.9 to generate significant variability in the secretion fluxes across different species (see main text).

#### 418 Minimal model

419 Our minimal model is set within the same MicroCRM framework that we used for the previous simulations. As  
 420 described in the main text, the model contains two communities of two species each ( $s_1$  to  $s_4$ ), with five resources  
 421 in total, out of which the first one ( $R_1$ ) is replenished externally at the beginning of each growth cycle and the  
 422 rest correspond to the species' metabolic byproducts. Each species secretes a unique byproduct, meaning that  
 423 the metabolic matrix  $\mathbf{D}$  is binary in this case. The specific structure of  $\mathbf{D}$  is displayed below –because it is a 3-  
 424 dimensional matrix in our framework, we have “sliced” it into the five 2-dimensional matrices corresponding to  
 425 our five species.

$$\mathbf{D}_1 = \begin{array}{ccccc} R_1 & R_2 & R_3 & R_4 & R_5 \\ \hline R_1 & (0 & * & 0 & * & *) \\ R_2 & 1 & * & 0 & * & * \\ R_3 & 0 & * & 1 & * & * \\ R_4 & 0 & * & 0 & * & * \\ R_5 & 0 & * & 0 & * & * \end{array}$$

$$\mathbf{D}_2 = \begin{array}{ccccc} R_1 & R_2 & R_3 & R_4 & R_5 \\ \hline R_1 & (* & 0 & * & * & *) \\ R_2 & * & 0 & * & * & * \\ R_3 & * & 1 & * & * & * \\ R_4 & * & 0 & * & * & * \\ R_5 & (*) & 0 & * & * & * \end{array}$$

$$\mathbf{D}_3 = \begin{array}{ccccc} R_1 & R_2 & R_3 & R_4 & R_5 \\ \hline R_1 & (0 & * & * & * & 0) \\ R_2 & 0 & * & * & * & 0 \\ R_3 & 0 & * & * & * & 0 \\ R_4 & 1 & * & * & * & 0 \\ R_5 & 0 & * & * & * & 1 \end{array}$$

$$\mathbf{D}_4 = \begin{array}{ccccc} R_1 & R_2 & R_3 & R_4 & R_5 \\ \hline R_1 & (* & * & * & 0 & *) \\ R_2 & * & * & * & 0 & * \\ R_3 & * & * & * & 0 & * \\ R_4 & * & * & * & 0 & * \\ R_5 & * & * & * & 1 & * \end{array}$$

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

$$\mathbf{c} = \begin{array}{ccccc} R_1 & R_2 & R_3 & R_4 & R_5 \\ \hline s_1 & (c_{11} & 0 & c_{13} & 0 & 0) \\ s_2 & 0 & 100 & 0 & 0 & 0 \\ s_3 & c_{31} & 0 & 0 & 0 & c_{35} \\ s_4 & 0 & 0 & 0 & 100 & 0 \end{array}$$

429 Where we made the sub-dominants equally strong consumers of their dominants' secretions ( $c_{22} = c_{44} = 100$ ),  
 430 and we varied all other uptake rates depending on the scenario we were considering (see main text). Whenever  
 431 we were interested in the ratio between two rates (e.g.  $c_{35}/c_{13}$  in Figure 4e) we gave the one in the denominator a  
 432 fixed value of 1 and let one in the numerator range within the specified limits.

433 **Data & code availability**

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

436 **Acknowledgements**

437 The authors wish to thank Pankaj Mehta, Wenping Cui, Robert Marsland and all members of the Sanchez labora-  
438 tory for many helpful discussions. We also wish to express our gratitude to the Goodman laboratory at Yale for  
439 technical help during the early stages of this project. The funding for this work partly results from a Scialog Pro-  
440 gram sponsored jointly by the Research Corporation for Science Advancement and the Gordon and Betty Moore  
441 Foundation through grants to Yale University by the Research Corporation and the Simons Foundation.

## 442 References

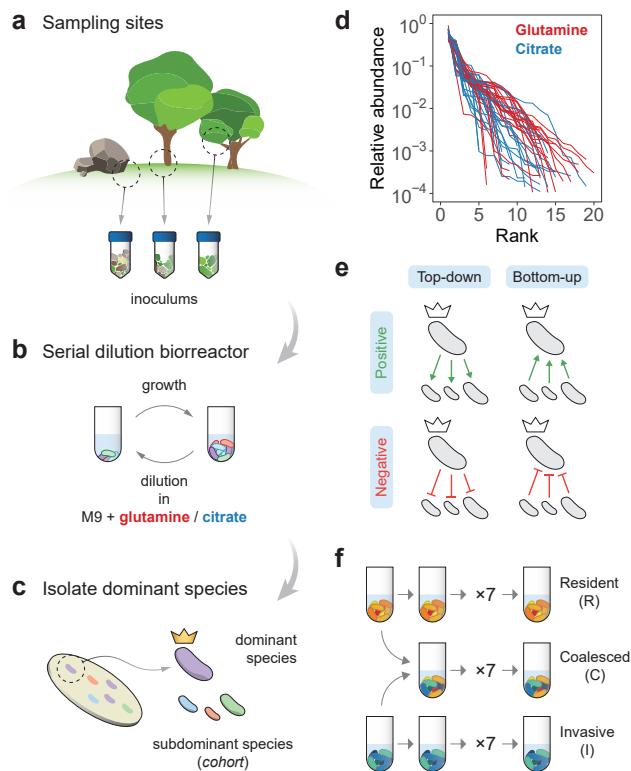
- 443 1. Mansour I, Heppell CM, Ryo M and Rillig MC (2018). Application of the microbial community coalescence  
444 concept to riverine networks. *Biological Reviews* **93**(4):1832–1845
- 445 2. Luo X, Xiang X, Yang Y, Huang G, Fu K, Che R and Chen L (2020). Seasonal effects of river flow on microbial  
446 community coalescence and diversity in a riverine network. *FEMS Microbiology Ecology* **96**(8):fiaa132
- 447 3. Vass M, Székely AJ, Lindström ES, Osman OA and Langenheder S (2021). Warming mediates the resistance  
448 of aquatic bacteria to invasion during community coalescence. *Molecular Ecology* **30**(5):1345–1356
- 449 4. Rillig MC, Lehmann A, Aguilar-Trigueros CA, Antonovics J, Caruso T, Hempel S, Lehmann J, Valyi K,  
450 Verbruggen E et al. (2016). Soil microbes and community coalescence. *Pedobiologia* **59**(1-2):37–40
- 451 5. Evans SE, Bell-Dereske LP, Dougherty KM and Kittredge HA (2019). Dispersal alters soil microbial commu-  
452 nity response to drought. *Environmental Microbiology* **22**(3):905–916
- 453 6. Dutton CL, Subalusky AL, Sanchez A, Estrela S, Lu N, Hamilton SK, Njoroge L, Rosi EJ and Post DM  
454 (2021). The meta-gut: Hippo inputs lead to community coalescence of animal and environmental micro-  
455 biomes. *biorXiv*
- 456 7. Vandegrift R, Fahimipour AK, Muscarella M, Bateman AC, Wymelenberg KVD and Bohannan BJ (2019).  
457 Moving microbes: the dynamics of transient microbial residence on human skin. *biorXiv*
- 458 8. Rillig MC, Antonovics J, Caruso T, Lehmann A, Powell JR, Veresoglou SD and Verbruggen E (2015). Inter-  
459 change of entire communities: microbial community coalescence. *Trends in Ecology & Evolution* **30**(8):470–  
460 476
- 461 9. Gilpin M (1994). Community-level competition: asymmetrical dominance. *Proceedings of the National  
462 Academy of Sciences* **91**(8):3252–3254
- 463 10. Simberloff D and Holle BV (1999). Positive Interactions of Nonindigenous Species: Invasional Meltdown?  
464 *Biological Invasions* **1**(1):21–32
- 465 11. Grosholz ED (2005). Recent biological invasion may hasten invasional meltdown by accelerating historical  
466 introductions. *Proceedings of the National Academy of Sciences* **102**(4):1088–1091
- 467 12. Simberloff D (2006). Invasional meltdown 6 years later: important phenomenon, unfortunate metaphor, or  
468 both? *Ecology Letters* **9**(8):912–919
- 469 13. Gurevitch J (2006). Commentary on Simberloff (2006): Meltdowns, snowballs and positive feedbacks. *Ecol-  
470 ogy Letters* **9**(8):919–921
- 471 14. Green PT, O'Dowd DJ, Abbott KL, Jeffery M, Retallick K and Nally RM (2011). Invasional meltdown:  
472 Invader-invader mutualism facilitates a secondary invasion. *Ecology* **92**(9):1758–1768
- 473 15. Livingston G, Jiang Y, Fox JW and Leibold MA (2013). The dynamics of community assembly under sudden  
474 mixing in experimental microcosms. *Ecology* **94**(12):2898–2906
- 475 16. Prior KM, Robinson JM, Dunphy SAM and Frederickson ME (2015). Mutualism between co-introduced  
476 species facilitates invasion and alters plant community structure. *Proceedings of the Royal Society B: Biolog-  
477 ical Sciences* **282**(1800):20142846
- 478 17. O'Loughlin LS and Green PT (2017). Secondary invasion: When invasion success is contingent on other  
479 invaders altering the properties of recipient ecosystems. *Ecology and Evolution* **7**(19):7628–7637
- 480 18. Castledine M, Sierociński P, Padfield D and Buckling A (2020). Community coalescence: an eco-evolutionary  
481 perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**(1798):20190252
- 482 19. Toquenaga Y (1997). Historicity of a Simple Competition Model. *Journal of Theoretical Biology* **187**(2):175–  
483 181
- 484 20. Tikhonov M (2016). Community-level cohesion without cooperation. *eLife* **5**:e15747
- 485 21. Tikhonov M and Monasson R (2017). Collective Phase in Resource Competition in a Highly Diverse Ecosys-  
486 tem. *Physical Review Letters* **118**(4):048103

- 487 22. Vila JCC, Jones ML, Patel M, Bell T and Rosindell J (2019). Uncovering the rules of microbial community  
488 invasions. *Nature Ecology & Evolution* **3**(8):1162–1171
- 489 23. Lechón P, Clegg T, Cook J, Smith TP and Pawar S (2021). The role of competition versus cooperation in  
490 microbial community coalescence. *biorXiv*
- 491 24. Sierociński P, Milferstedt K, Bayer F, Großkopf T, Alston M, Bastkowski S, Swarbreck D, Hobbs PJ, Soyer OS  
492 et al. (2017). A Single Community Dominates Structure and Function of a Mixture of Multiple Methanogenic  
493 Communities. *Current Biology* **27**(21):3390–3395.e4
- 494 25. Rillig MC and Mansour I (2017). Microbial Ecology: Community Coalescence Stirs Things Up. *Current  
495 Biology* **27**(23):R1280–R1282
- 496 26. Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DS, Parfrey LW, Farjalla VF and Doebeli M (2016).  
497 High taxonomic variability despite stable functional structure across microbial communities. *Nature Ecology  
498 & Evolution* **1**(1):0015
- 499 27. Winfree R, Fox JW, Williams NM, Reilly JR and Cariveau DP (2015). Abundance of common species, not  
500 species richness, drives delivery of a real-world ecosystem service. *Ecology Letters* **18**(7):626–635
- 501 28. Rosenzweig RF, Sharp RR, Treves DS and Adams J (1994). Microbial evolution in a simple unstructured  
502 environment: genetic differentiation in *Escherichia coli*. *Genetics* **137**(4):903–917
- 503 29. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, Segré D, Mehta P and Sanchez A  
504 (2018). Emergent simplicity in microbial community assembly. *Science* **361**(6401):469–474
- 505 30. Estrela S, Vila JCC, Lu N, Bajic D, Rebollo-Gómez M, Chang CY and Sanchez A (2020). Metabolic rules  
506 of microbial community assembly. *biorXiv*
- 507 31. Marsland III R, Cui W, Goldford J, Sanchez A, Korolev K and Mehta P (2019). Available energy fluxes drive  
508 a transition in the diversity, stability, and functional structure of microbial communities. *PLoS Computational  
509 Biology* **15**(2):e1006793
- 510 32. Marsland R, Cui W, Goldford J and Mehta P (2020). The Community Simulator: A Python package for  
511 microbial ecology. *PLoS ONE* **15**(3):e0230430
- 512 33. Dimroth P (2004). Molecular Basis for Bacterial Growth on Citrate or Malonate. *EcoSal Plus* **1**(1)
- 513 34. Forchhammer K (2007). Glutamine signalling in bacteria. *Frontiers in Bioscience* **12**(1):358
- 514 35. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA and Holmes SP (2016). DADA2: High-  
515 resolution sample inference from Illumina amplicon data. *Nature Methods* **13**(7):581–583
- 516 36. Callahan BJ, McMurdie PJ and Holmes SP (2017). Exact sequence variants should replace operational taxo-  
517 nomic units in marker-gene data analysis. *The ISME Journal* **11**:2639–2643
- 518 37. MacArthur R (1970). Species packing and competitive equilibrium for many species. *Theoretical Population  
519 Biology* **1**(1):1–11
- 520 38. Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, Bonilla G, Kar A, Leiby N et al.  
521 (2014). Metabolic Resource Allocation in Individual Microbes Determines Ecosystem Interactions and Spatial  
522 Dynamics. *Cell Reports* **7**(4):1104–1115
- 523 39. Pinu FR, Granucci N, Daniell J, Han TL, Carneiro S, Rocha I, Nielsen J and Villas-Boas SG (2018). Metabolite  
524 secretion in microorganisms: the theory of metabolic overflow put to the test. *Metabolomics* **14**(4)
- 525 40. Kozich JJ, Westcott SL, Baxter NT, Highlander SK and Schloss PD (2013). Development of a Dual-Index  
526 Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina  
527 Sequencing Platform. *Applied and Environmental Microbiology* **79**(17):5112–5120
- 528 41. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich  
529 JK et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*  
530 **7**:335–336

- 531 42. Wang Q, Garrity GM, Tiedje JM and Cole JR (2007). Naïve Bayesian Classifier for Rapid Assignment of  
532 rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology* **73**(16):5261–  
533 5267
- 534 43. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J and Glöckner FO (2013). The SILVA  
535 ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*  
536 **41**(D1):D590–D596
- 537 44. Curtis JT and Bray JR (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin.  
538 *Ecological Monographs* **27**(4):325–349
- 539 45. Lin J (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*  
540 **37**(1):145–151
- 541 46. Kullback S and Leibler RA (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*  
542 **22**(1):79–86
- 543 47. Jaccard P (1912). The distribution of the flora in the alpine zone. *New Phytologist* **11**(2):37–50

## Figures

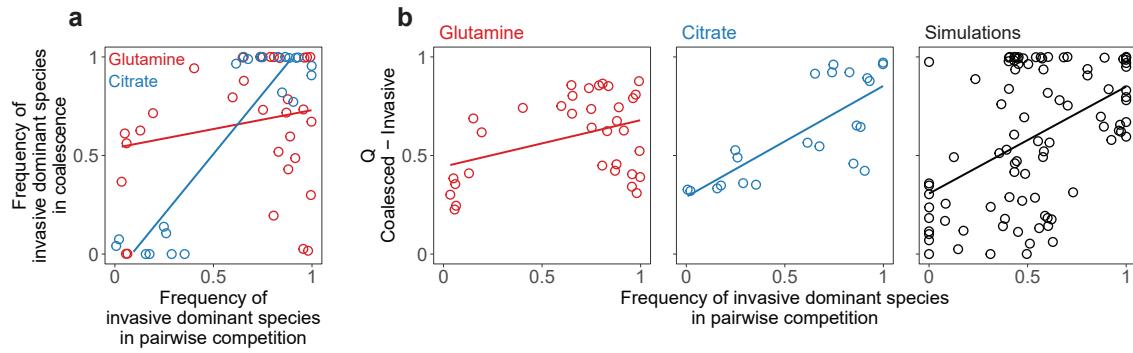
544



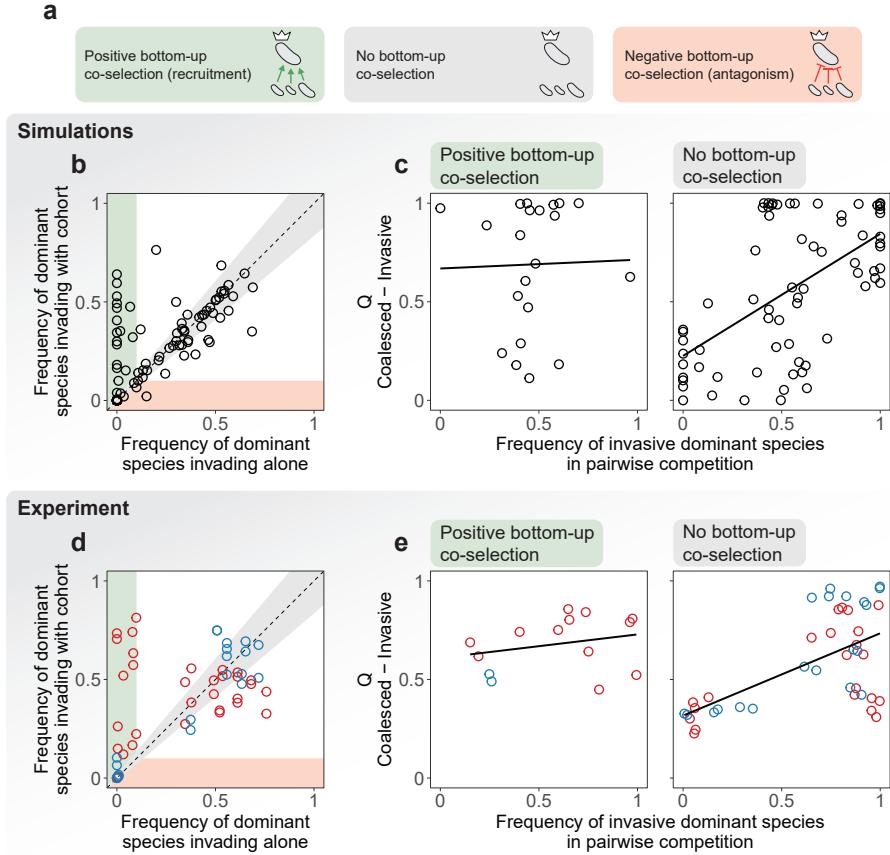
545

**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.

559

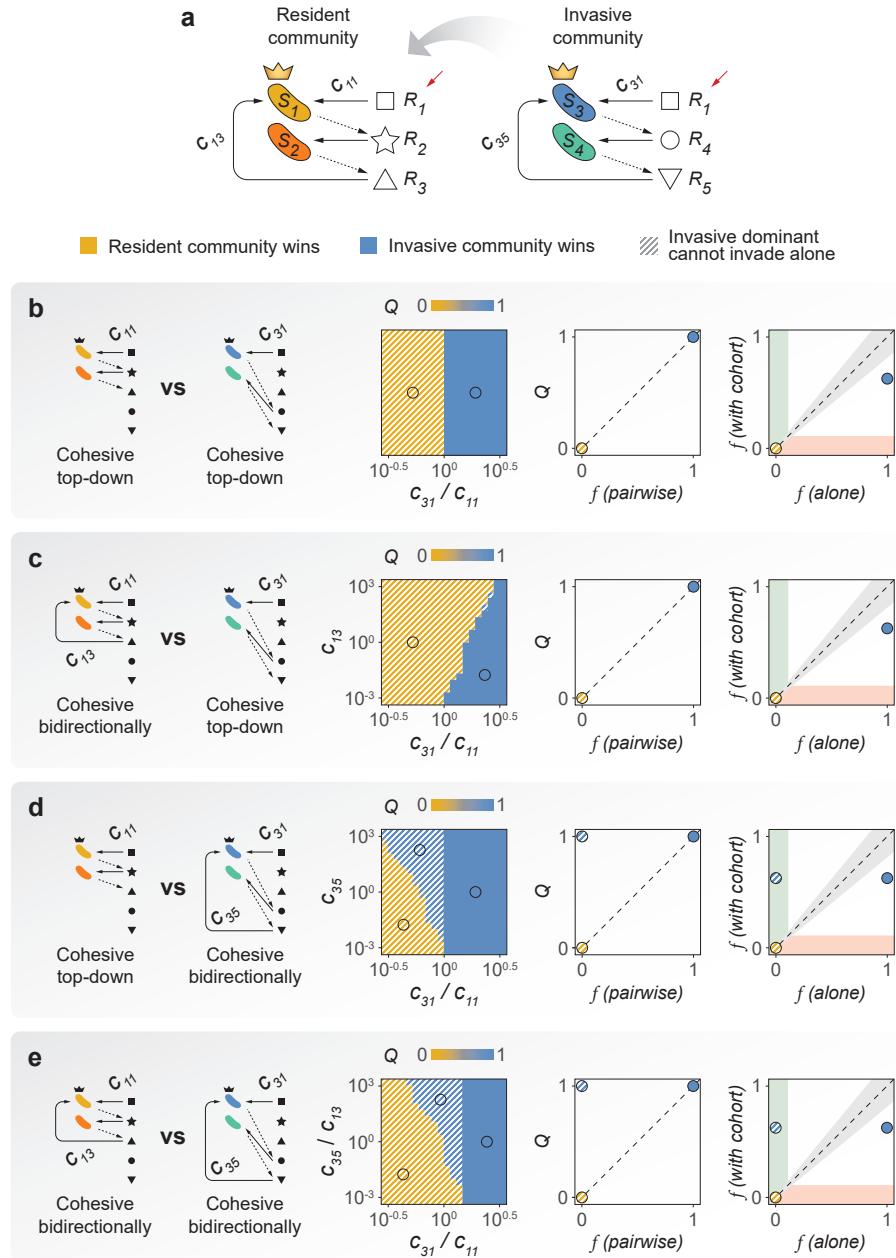


560 **Figure 2. Top-down co-selection in microbial community coalescence.** **a.** Pairwise competition of dominants with or  
 561 without their cohorts. In the horizontal axis, we plot the frequency of the invasive dominant species in head-to-head pairwise  
 562 competition with the resident dominant. In the vertical axis, we plot the same relative frequency when the two species compete  
 563 in the presence of their cohorts, i.e. during community coalescence.  $R^2 = 0.04$ ,  $p > 0.05$  for glutamine (red) and  $R^2 = 0.83$ ,  
 564  $p < 10^{-8}$  for citrate (blue). **b.** Coalescence outcomes are quantified by the relative Bray-Curtis similarity ( $Q$ ) between the  
 565 coalesced and invasive communities. These outcomes are predicted by the pairwise competition between the invasive and  
 566 resident dominant species. Left panel (red): glutamine communities,  $R^2 = 0.15$ ,  $p < 0.05$ . Middle panel (blue): citrate  
 567 communities,  $R^2 = 0.57$ ,  $p < 10^{-4}$ . A high correlation is consistent with a scenario of strong top-down positive co-selection  
 568 where dominants recruit their cohorts for the final coalesced community. Two biological replicates per experiment are plotted  
 569 individually. Right panel (black): simulations with a Microbial Consumer-Resource Model are able to capture these trends  
 570 ( $R^2 = 0.22$ ,  $p < 10^{-5}$ ).



572

573 **Figure 3. Trade offs between bottom-up and top-down ecological co-selection.** **a.** We hypothesize that three scenarios are  
 574 possible regarding bottom-up co-selection: sub-dominant species could co-select for (green) or against (red) their dominant in  
 575 coalescence, or they could have no effect in the invasion success of the dominant taxa (gray). **b.** Simulations with a Microbial  
 576 Consumer-Resource Model: we plot the frequency reached by the invasive dominants when invading the resident communities  
 577 in isolation versus the same frequency when invading together with their cohorts, i.e. in community coalescence. Points in the  
 578 green/red area represent instances where the invasive dominant is able to invade with higher/lower success when accompanied  
 579 by its cohort, evidencing positive/negative bottom-up co-selection. Points around the diagonal (gray area) correspond to cases  
 580 where the success of the invasive dominant is only weakly affected by the presence or absence of its cohort. **c.** We divided  
 581 the data from our simulations into two sets according to whether positive or no bottom-up co-selection was observed (that is,  
 582 whether points fell into the green or gray areas of panel b). Here we reproduce the plots in Figure 2b for each set, representing  
 583 the result of the dominant head-to-head pairwise competition versus the outcome of community coalescence. Left panel: strong  
 584 positive 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 show that in our conditions, positive bottom-up co-selection is indeed more frequent and strong than negative  
 585 bottom-up co-selection. **e.** We reproduce the plots in panel c for our experimental data, i.e. we recreate Figure 2b but this  
 586 time splitting our data by the strength of bottom-up co-selection instead of by the carbon source supplied to the communities.  
 587 Left panel: strong positive bottom-up co-selection ( $R^2 = 0.07, p > 0.05$ ). Right panel: no bottom-up co-selection ( $R^2 = 0.37,$   
 588  $p < 10^{-4}$ ).

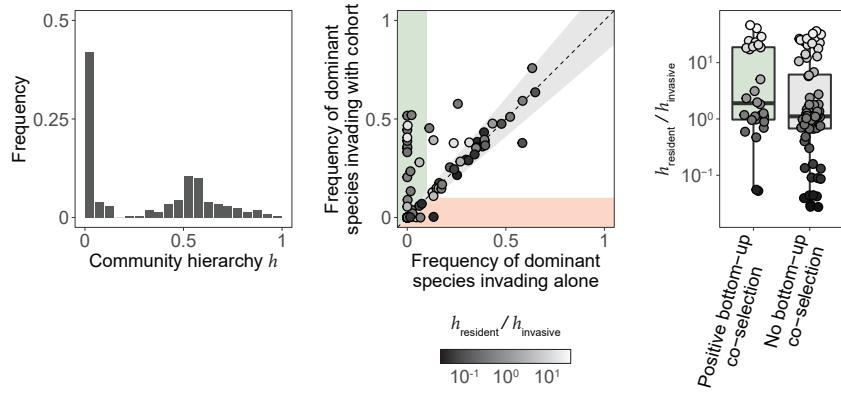


591

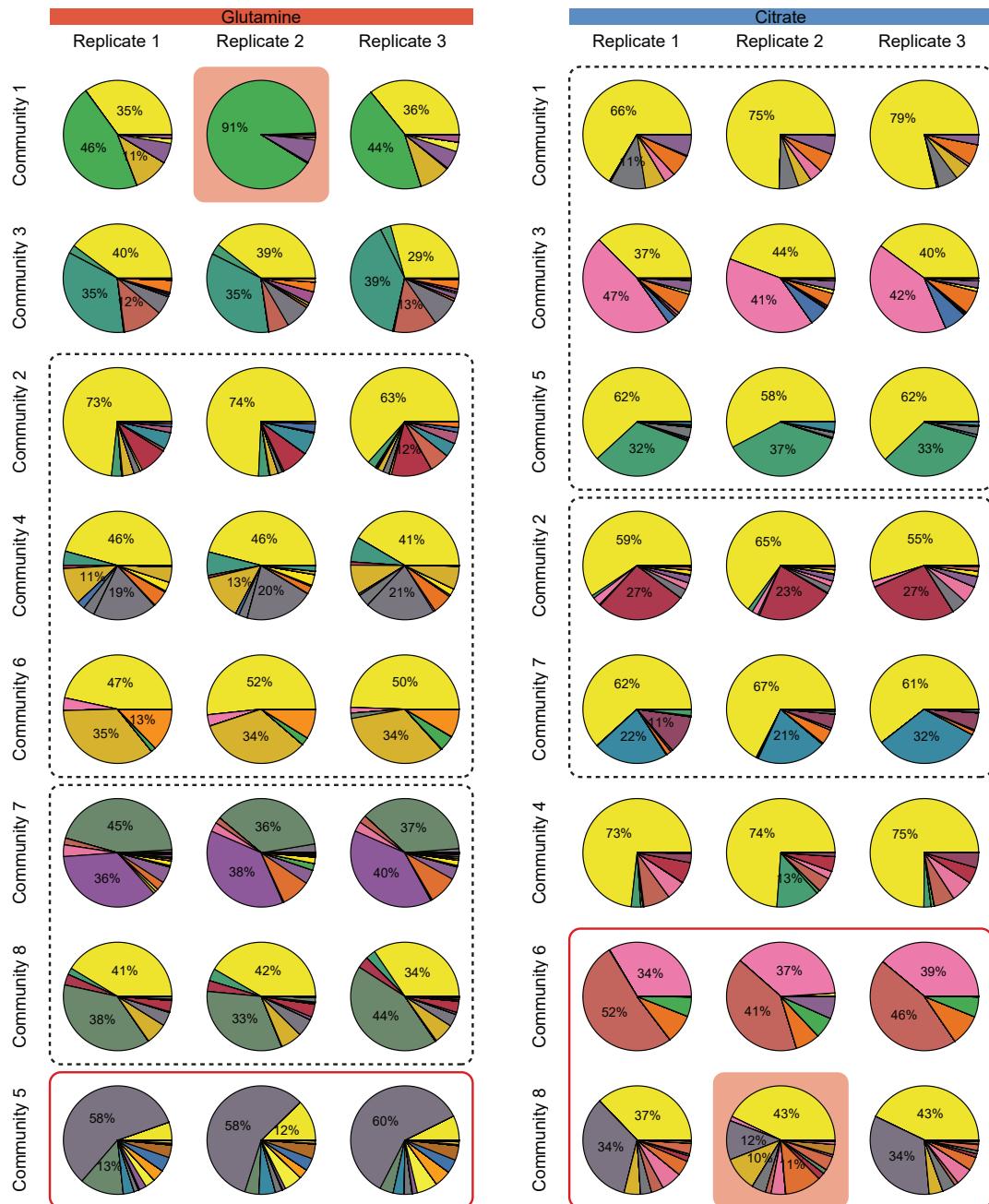
592 **Figure 4. A minimal model of community coalescence.** **a.** Illustration of the model structure and parameters. The primary  
 593 resource ( $R_1$ ) is replenished after each growth-dilution cycle (red arrows). Solid arrows indicate resource consumption, dashed  
 594 arrows represent resource secretion. **b-e.** Coalescence outcomes in the minimal model under different relations of cohesiveness  
 595 between the resident and the invasive communities.

597

598 **Figure 5. Community hierarchies determine coalescence outcomes.** **a.** Blablabla... **b.** Blablabla... **c.** Blablabla... ( $p =$   
599 0.083).

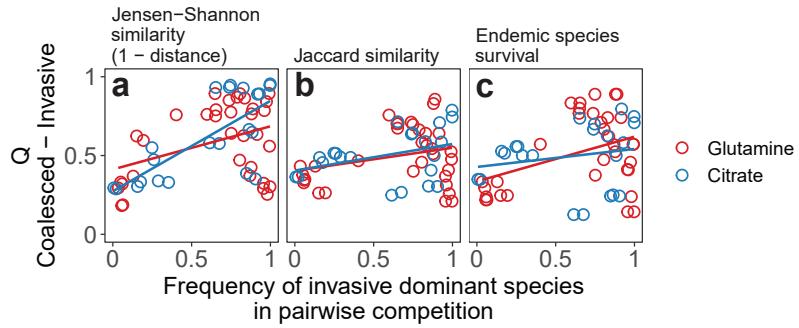


601 **Supplementary Figures**



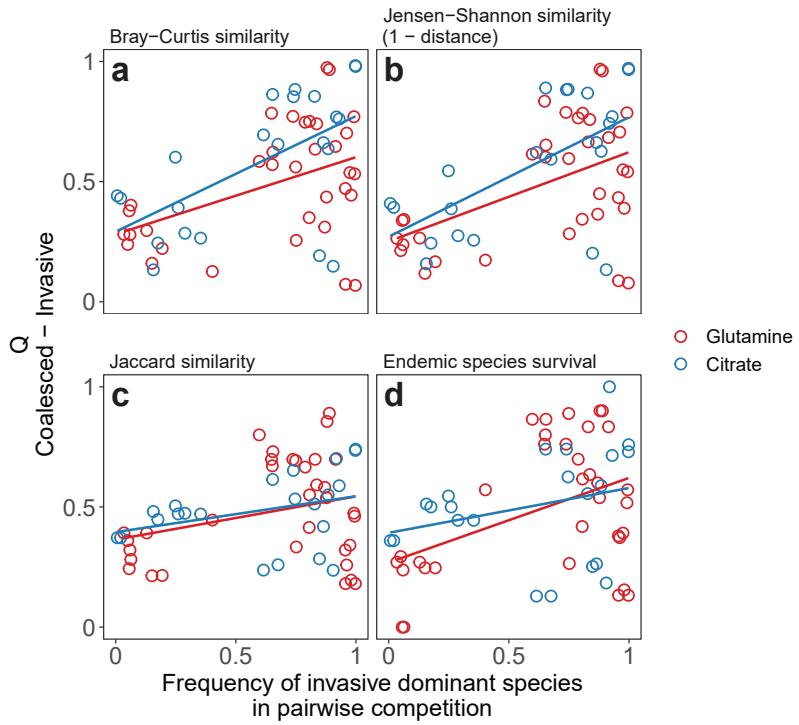
602

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



611

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



621

622 **Figure S3. Dominant species have limited effects on coalescence outcomes quantification.** We repeated the analyses shown  
 623 in [Figure 2a](#) and [Figure S2](#), but this time we removed the dominants from the compositional data prior to quantifying community  
 624 distances. The trends observed before are maintained. **a.** Relative Bray-Curtis similarity ( $R^2 = 0.20, p < 0.01$  for glutamine  
 625 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  $R^2 = 0.36,$   
 626  $p < 0.005$  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.**  
 627 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  
 628 citrate).