

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

Juan Diaz-Colunga^{1*}, Nanxi Lu^{1*}, Alicia Sanchez-Gorostiaga^{1,2*}, Chang-Yu Chang¹, Helen S. Cai¹, Joshua E. Goldford³, Mikhail Tikhonov⁴, and Álvaro Sánchez¹✉

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

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

³Physics of Living Systems, Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA

⁴Department 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’s secretions [6], and the skin microbiota is also subject to invasions when it makes 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 invasion 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 invasion 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 community members 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 invasion 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. We refer to these two opposite scenarios as “top-down” or “bottom-up” forms of community cohesiveness, respectively. Top-down cohesiveness emerges when the dominant invader co-selects other sub-dominant taxa into the final community during coalescence. Alternatively, bottom-up cohesiveness refers to the case when the dominant is co-selected by the more rare members of its community. 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].

47 In previous work, we have shown that a large amount of soil and plant microbiomes can be cultured *ex situ*
48 in synthetic minimal environments with a single supplied limiting resource under serial growth-dilution cycles
49 [29] ([Figure 1a-b](#)). Under these conditions, environmental microbiomes spontaneously re-assemble into complex
50 multi-species communities sustained by dense cross-feeding facilitation networks [29]. In addition, and just like
51 in natural consortia, species abundance distributions in these communities are generally long-tailed and uneven
52 ([Figure 1d](#) and [Figure S1](#)), with the dominant (most abundant) species typically comprising most of the biomass
53 (median = 46%, [Figure S1](#)). Because these communities are easy to manipulate and grow in high throughput,
54 they represent good test cases to investigate ecological co-selection during community coalescence. Here we
55 focus on the dominants and ask whether they can co-select or be co-selected by the sub-dominant species in their
56 communities (henceforth referred to as their *cohorts*, [Figure 1c](#)).

57 Our results indicate that when top-down co-selection is weak, bottom-up co-selection can be very strong, with
58 positive co-selection being far more common than negative co-selection. We then turn to a Microbial Consumer-
59 Resource Model (MicroCRM) [29–31] that is able to capture the dynamics of microbial communities dominated
60 by metabolic interactions, as is the case for the ones assembled in our experimental conditions [29, 32]. We show
61 that the empirically observed trends in ecological co-selection are reproduced with minimal model assumptions,
62 and that the recurrence of top-down and bottom-up co-selection is determined by the configuration of the cross-
63 feeding networks in the MicroCRM. Our findings indicate that collective interactions play an important role at
64 dictating community structure during coalescence.

65 Results & Discussion

66 We collected eight natural microbiomes from different soil and plant environmental samples ([Figure 1a](#)) and used
67 them to inoculate eight identical habitats containing minimal media with either glutamine or citrate as the only sup-
68 plied carbon source. We chose these two carbon sources because they are metabolized through different pathways
69 in bacteria [33, 34], and we hypothesize that communities assembled in either resource will be supported by cross-
70 feeding networks of distinct sets of metabolites [29, 32], thus leading to potentially variable degrees of community
71 cohesiveness and coalescence outcomes [18, 20, 21, 23]. After inoculation, all communities were serially passaged
72 for 12 transfers (84 generations), with an incubation time of 48 hours and a dilution factor of 1:100. ([Figure 1b](#),
73 [Methods: Stabilization of environmental communities in simple synthetic environments](#)). In previous work we
74 have shown that under these conditions, 12 transfers allow communities to approach a state of “generational equi-
75 librium”, where the community composition at the end of one batch incubation will be the same as in consecutive
76 incubations. We isolated the dominant species of every community ([Methods: Isolation of dominant species](#)) and
77 identified them by Sanger-sequencing their 16S rRNA gene ([Methods: Determination of community composition
78 by 16S sequencing](#)), which correctly matched the dominant Exact Sequence Variant (ESV) [35, 36] found through
79 community-level 16S Illumina sequencing ([Figure S1](#)). These dominants remained at high frequency after seven
80 additional transfers with the exception of two of the citrate communities and one of the glutamine communities
81 (where the dominants were presumably a transiently dominating species) that were excluded from further analysis
82 ([Figure S1](#)). Similarly, pairs of communities where the dominants shared a same 16S sequence and had similar
83 colony morphology were excluded ([Figure S1](#)).

84 Top-down ecological co-selection

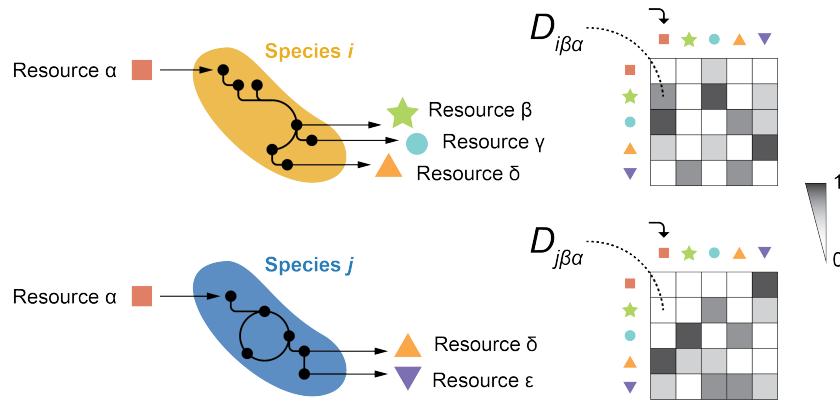
85 One form of cohesiveness may arise when the sub-dominant members of the community depend on the dominant
86 species. This can occur, for instance, when the dominant provides resources (or stressors) that select for the sub-
87 dominant taxa ([Figure 1e](#), left panels). If communities being coalesced are highly cohesive from the top-down,
88 the fate of the sub-dominant community members will be tied to their dominant: if it gets excluded, they will
89 be likely to fall with it, and if it is able to resist coalescence, they will be likely to follow suit. In this scenario,
90 we would expect the outcome of community coalescence to be predicted by which of the two dominants is most
91 competitive in pairwise competition. Likewise, competition between dominants should be affected only weakly by
92 the presence or absence of sub-dominant species, which would play a passive role under these conditions. To test
93 this hypothesis, we performed all pairwise competitions between dominant species in glutamine and citrate envi-
94 ronments by mixing them 1:1 on their native media and propagating the cultures for seven serial transfers, roughly
95 42 generations ([Methods: Coalescence, competition and invasion experiments](#)). We then carried out all possible
96 pairwise community coalescence experiments by mixing equal volumes of the communities and propagating the re-
97 sulting cultures for seven extra transfers ([Figure 1f](#)). The frequencies of all species in both community-community
98 and dominant-dominant competitions were determined by 16S Illumina sequencing ([Methods: Determination of
99 community composition by 16S sequencing](#)).

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

To investigate the determinants of top-down co-selection and the factors modulating its strength, we ran a set of simulations of community coalescence. We used a Microbial Consumer-Resource Model (MicroCRM) [29, 30] as implemented in the Community Simulator package for Python [31] ([Box 1](#)). We chose this modeling framework because communities assembled under our experimental conditions (natural microbiomes re-assembled into multi-species communities through serial growth-dilution cycles in synthetic minimal media with a single carbon source) have been shown to be sustained by dense metabolic cross-feeding networks [29, 32] for which the MicroCRM provides a good description. We and others have previously found a strong concordance between the behavior of laboratory and natural microbial communities and the generic behavior of the MicroCRM [29–31, 37, 38]. To reproduce our experimental protocol *in silico*, we first generated a library of resources and two non-overlapping pools of species. Each pool was used to seed a collection of 100 invasive and 100 resident communities respectively by randomly choosing 50 species and allowing them to stabilize through 20 growth-dilution cycles. We then mixed these stable communities in pairs to simulate our coalescence and dominant-dominant competition experiments ([Methods: Simulations](#)). We found that the MicroCRM simulations naturally exhibit the observed correlation between the head-to-head pairwise competition of dominants and the outcome of community coalescence ([Figure 2a](#), right panel), further supporting the idea that top-down ecological co-selection consistently emerges from metabolic interactions across species. Moreover, we found that top-down co-selection is observed under a wide range of different simulation conditions and cross-feeding networks ([Figure ??](#)), indicating that it is a robust phenomenon.

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

145 The Microbial Consumer-Resource Model (MicroCRM) [29–31] is a modeling framework based on the clas-
 146 sic MacArthur's consumer resource model [39]. It encodes the dynamics of a system with S species and M
 147 resources in terms of a consumer preference matrix \mathbf{c} and a metabolic matrix \mathbf{D} , with an additional set of
 148 parameters controlling the species maintenance costs (m_i for species i), the resource energy densities (w_α for
 149 resource α), the energy to growth rate conversion factor (g_i for species i) and the leakage fraction, i.e. the
 150 amount of energy lost as byproducts when a resource is consumed (l_α for resource α). The element c_{ia} of the
 151 consumer preference matrix represents the uptake rate of resource α by species i (although the relationship
 152 between c_{ia} and the uptake rate can be more complex in modeling scenarios that are not considered here,
 153 see [29–31]). Experimental evidence suggests that individual species can secrete different sets of metabo-
 154 lites to the environment when growing on a same primary resource [32, 40, 41]. Thus, we define \mathbf{D} as a
 155 three-dimensional matrix where the element $D_{i\beta\alpha}$ represents the energy flux in the form of resource β that is
 156 secreted by species i when it metabolizes resource α . Note that $D_{i\beta\alpha}$ need not be equal to $D_{j\beta\alpha}$ if $i \neq j$ (see
 157 illustration below).



158 The following equations describe the kinetics of the abundances of the i -th species (denoted as N_i) and
 159 the α -th resource (denoted as R_α):

$$160 \frac{dN_i}{dt} = g_i N_i \left[\sum_\alpha (1 - l_\alpha) w_\alpha c_{i\alpha} R_\alpha - m_i \right] \quad (1)$$

$$161 \frac{dR_\alpha}{dt} = - \sum_j N_j c_{j\alpha} R_\alpha + \sum_j \sum_\beta N_j c_{j\beta} R_\beta \left[l_\beta D_{j\beta\alpha} \frac{w_\beta}{w_\alpha} \right] \quad (2)$$

162 These equations can take slightly different forms in certain cases, e.g. if the primary resource is supplied
 163 continuously instead of at the beginning of each growth cycle [30, 31]. They represent a good approxima-
 164 tion for the community dynamics between consecutive serial dilutions in our setup. Here, we assembled *in*
 165 *silico* communities by randomly sampling a set of species from a pool, then integrating equations 1 and 2,
 166 diluting the final abundances, replenishing the primary resource, and repeating the process until generational
 167 equilibrium was achieved (Methods: Simulations). Coalescence simulations were carried out following the
 168 same logic, this time seeding the coalesced communities by mixing the invasive and resident ones instead of
 169 sampling from a species pool.

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

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

193 In simulations of the MicroCRM, we find no instances of bottom-up antagonism but multiple such instances
194 of positive bottom-up co-selection ([Figure 3b](#)). Many dominant members of our *in silico* communities could not
195 invade another community on their own (or could only do so at very low final relative abundances, below 0.1)
196 but were able to reach high frequencies when they were accompanied by their cohorts in community coalescence.
197 Thus, theory indicates that positive bottom-up co-selection is frequent and potentially very strong, while negative
198 bottom-up co-selection is far more uncommon. Interestingly, our simulations suggest that strong bottom-up co-
199 selection should only be observed in communities where top-down co-selection is weak, while top down co-selection
200 is only seen when bottom-up co-selection is weak. To better illustrate this prediction, we divided our simulations
201 into two subsets: the first one was comprised of the instances where positive bottom-up co-selection was strong
202 (i.e. dots in the green shaded region of [Figure 3b](#)), the second set included all other cases (dots near the diagonal of
203 [Figure 3b](#)). We reexamined our original simulations and found that when bottom-up positive co-selection is
204 strong, the pairwise competition of dominants is not predictive of coalescence outcomes ([Figure 3c](#), left panel)
205 indicating that top-down co-selection is weak. At the same time, when considering only those coalesced com-
206 munities in the diagonal of [Figure 3b](#) (where bottom-up co-selection is weak), our model predicts that the fates
207 of the sub-dominant community members after coalescence are much more strongly determined by the head-to-
208 head competition between dominants in isolation ($R^2 = 0.34$ for instances where bottom-up co-selection is weak,
209 [Figure 3c](#) right panel; $R^2 = 0.22$ when all instances are considered, [Figure 2a](#) right panel).

210 We then asked whether this trend was also observed in our experimental communities. To address this question,
211 we carried out a new round of experiments where we invaded the resident communities with the invasive dominants
212 alone ([Methods: Coalescence, competition and invasion experiments](#)). After stabilization ([Methods: Stabilization
213 of environmental communities in simple synthetic environments](#)), we quantified species abundance through 16S
214 Illumina sequencing ([Methods: Determination of community composition by 16S sequencing](#)). Consistent with
215 the behavior of our model, we observed that bottom-up co-selection is far more common in its positive than in
216 its negative form ([Figure 3d](#)). Interestingly, bottom-up recruitment appears to be more frequent in the glutamine
217 environments, where top-down co-selection was weak, than in the citrate ones, where top-down co-selection was
218 strong ([Figure 2](#)). We then repeated our analysis in [Figure 3c](#), this time splitting our data according to the observed
219 strength of bottom-up co-selection instead of the primary carbon source as we had done in [Figure 2a](#). Our findings
220 were in line with the model prediction: pairwise competition between dominants is only predictive of coalescence
221 outcomes if bottom-up co-selection is weak ([Figure 3e](#), $R^2 = 0.07$, $p > 0.05$ when bottom-up co-selection is strong;
222 $R^2 = 0.37$, $p < 10^{-4}$ when bottom-up co-selection is weak). Once the bottom-up communities are removed, the
223 glutamine and citrate communities behave very similarly and both have similar degrees of top-down cohesiveness
224 ([Figure 3e](#) right panel, $R^2 = 0.27$ for the glutamine environments, in red; $R^2 = 0.54$ for the citrate environments,
225 in blue; and $R^2 = 0.37$ when both environments are considered, black line). This suggests that the main difference
226 between citrate and glutamine is that the latter is richer in communities exhibiting bottom-up cohesiveness than
227 the former.

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

229 To better understand the underlying mechanisms that govern the emergence of ecological co-selection in our ex-
230 periments, we found it useful to study a minimal model of community coalescence ([Methods: Minimal model](#)).
231 This model is comprised of two communities with two species each as illustrated in [Figure 4a](#). Within each com-
232 munity, the dominant species (s_1 and s_3 in the resident and invasive communities respectively) are able to utilize

233 the single externally supplied resource (R_1). They secrete a single byproduct (R_2 and R_4 respectively) off which
 234 the sub-dominants (s_2 and s_4 respectively) can feed. Finally, these sub-dominants secrete an additional resource
 235 (R_3 and R_5 respectively) that can in turn be metabolized by the corresponding dominants. With this structure, com-
 236 munity members interact via simple cross-feeding networks that can be very vertical if the dominant cross-feeds
 237 the sub-dominant but the sub-dominant does not cross-feed the dominant (i.e. if the dominant cannot utilize the
 238 sub-dominant's secretions), or more horizontal (bidirectional) if the cohort also cross-feeds the dominant. We hy-
 239 pothesize that when two such communities coalesce, bottom-up ecological co-selection can emerge if the invasive
 240 community is sustained by a bidirectional cross-feeding network.

241 In the limit case when the cross-feeding networks of both the invasive and resident communities are strictly
 242 vertical (that is, the sub-dominants are passively sustained by the dominants but do not cross-feed them), it is
 243 straightforward that the outcome of community coalescence will depend on the competitive ability of the dominants
 244 to grow on the single externally supplied resource. The most competitive dominant will co-select its sub-dominant
 245 (from the top-down) through the secretion of specific metabolic byproducts (Figure 4b). If the resident community
 246 is maintained by a more horizontal cross-feeding network, it can display further resistance to invasion. In this
 247 scenario, even if the resident dominant is less competitive for the externally supplied resource than the invasive
 248 dominant, cross-feeding from the resident cohort can favor its success in coalescence. The stronger the metabolic
 249 flux from the resident cohort towards the dominant, the more prominent this effect can be (Figure 4c). On the
 250 other hand, if the cross-feeding network of the invasive community is horizontal (i.e. the sub-dominant is cross-fed
 251 by and also cross-feeds the dominant), more complex behaviors can emerge. The invasive dominant may not be
 252 able to invade the resident community by itself if it is less competitive for the externally supplied resource than
 253 the resident dominant (Figure 4d), or if despite being more competitive, cross-feeding from the resident cohort
 254 towards the resident dominant favors the success of the latter (Figure 4e). But even then, the invasive community
 255 could dominate in coalescence (i.e. when the invasive sub-dominant is also present). In order for this to happen,
 256 cross-feeding from the invasive cohort towards the invasive dominant should be strong enough to overcome the
 257 competitive disadvantage that said dominant may have in isolation.

258 In summary, coalescence outcomes are contingent on the direction of the cross-feeding networks sustaining the
 259 communities in this simple setting. We ran simulations of all scenarios described above with the minimal model of
 260 community coalescence implemented in the MicroCRM framework (Methods: Minimal model). In line with our
 261 hypothesis, simulations indicate that bottom-up co-selection of a dominant that is unable to invade by itself is only
 262 possible if said dominant is strongly cross-fed by its cohort (Figure 4).

263 Community hierarchy regulates the strength of bottom-up co-selection

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

$$269 h = \frac{\Delta N_{\text{dom}}^{\text{R}1}}{\Delta N_{\text{dom}}} \quad (3)$$

270 where ΔN_{dom} represents the overall increase in dominant biomass within a single batch incubation for a genera-
 271 tionally stable community, and $\Delta N_{\text{dom}}^{\text{R}1}$ represents the increase in said biomass resulting from the metabolism of
 272 the primary resource (R_1) only. If the dominant was just utilizing the primary resource, the cross-feeding network
 273 would be very hierarchical ($h \sim 1$), whereas if it was growing mostly on the secretions of other taxa, it would be
 274 more distributed ($h \ll 1$). We quantified the hierarchies of the resident and invasive communities in our Micro-
 275 CRM, finding that h follows a bimodal distribution (Figure 5a). We therefore divided our simulations into four
 276 groups according to whether the cross-feeding networks of both resident and invasive communities were vertical
 277 (high h) or horizontal (low h) as shown in Figure 5b. For each group, we evaluated the frequency of instances of
 278 bottom-up co-selection, i.e. the fraction of cases where a dominant that could not invade in isolation was success-
 279 ful when accompanied by its cohort (green area of Figure 3b). We found that bottom-up ecological co-selection
 280 is significantly more frequent when the invasive community is non-hierarchical (Figure 5c), in line with what the
 minimal model anticipated (Figure 4d-e).

281 Conclusions

282 Understanding the mechanisms underlying the responses of microbial communities to invasions is an essential
 283 but poorly understood question in microbial ecology [8]. Theory has suggested that communities may exhibit an
 284 emergent cohesiveness [9, 15, 20, 21], leading to members of the same community recruiting one another during

285 community-community invasions. Our results provide direct experimental evidence of ecological co-selection in
286 a large number of community coalescence experiments, and highlight the critical role that may be played by the
287 rarer, sub-dominant species in the generation of community cohesiveness.

288 Our simulations suggest that the strength and direction of ecological co-selection is modulated by the under-
289 lying cross-feeding networks that shape the structure of communities in synthetic minimal environments [29, 32].
290 This idea is supported by the observation that our Microbial Consumer-Resource Model captures the trends ob-
291 served experimentally when we enable a large variation in the metabolic fluxes across species. The model predicts
292 a trade-off between the strength of bottom-up co-selection and the ability of dominant-dominant pairwise com-
293 petition to dictate coalescence outcomes, which we have confirmed experimentally. It also suggests that rarer taxa
294 may play a more prominent role in co-selecting dominant species when the metabolic fluxes across the community
295 are distributed rather than hierarchical. These observations, together with previous results in different systems
296 [24] as well as theoretical predictions [9, 19–23], suggest that collective interactions of microbes with one another
297 and with the environment should be generically expected to produce ecological co-selection during community
298 coalescence. Testing this theoretical prediction would require to map the cross-feeding networks of all of our com-
299 munities. Keeping track of every molecule secreted by every species in co-culture and by which species they are
300 uptaken is still a low throughput process that is both labor intensive and expensive. Recent progress in metabolomic
301 tools promise to help us test this hypothesis in future work.

302 **Methods**

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

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

317 **Isolation of dominant species**

318 For each community, the most abundant colony morphotype at the end of the ninth transfer was selected ([Figure 1c](#)),
319 resuspended in 100µL 1×PBS and serially diluted (1:10). Next, 20µL of the cells diluted to 10⁻⁶ were plated in the
320 corresponding synthetic minimal media and allowed to regrow at 30°C for 48h. Dominants were then identified,
321 inoculated into 500µL of fresh media and incubated still at 30°C for 48h. After this period, the communities
322 stabilized for eleven transfers and the isolated dominants were ready for the competition experiments at the onset
323 of the twelfth transfer.

324 **Coalescence, competition and invasion experiments**

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

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

332 The sequencing protocol was identical to that described in [29]. Community samples were collected by spinning
333 down at 3500rpm for 25min in a bench-top centrifuge at room temperature; cell pellets were stored at -80°C
334 before processing. To maximize Gram-positive bacteria cell wall lysis, the cell pellets were re-suspended and
335 incubated at 37°C for 30min in enzymatic lysis buffer (20mM Tris-HCl, 2mM sodium EDTA, 1.2% Triton X-100)
336 and 20mg/mL of lysozyme from chicken egg white (Sigma-Aldrich). After cell lysis, the DNA extraction and
337 purification was performed using the DNeasy 96 protocol for animal tissues (Qiagen). The clean DNA in 100µL
338 elution buffer of 10mM Tris-HCl, 0.5mM EDTA at pH 9.0 was quantified using Quan-iT PicoGreen dsDNA Assay
339 Kit (Molecular Probes, Inc.) and normalized to 5ng/µL in nuclease-free water (Qiagen) for subsequent 16S rRNA
340 Illumina sequencing. 16S rRNA amplicon library preparation was performed following a dual-index paired-end
341 approach [42]. Briefly, PCR amplicon libraries of V4 regions of the 16S rRNA were prepared sing dual-index
342 primers (F515/R805), then pooled and sequenced using the Illumina MiSeq chemistry and platform. Each sample
343 went through a 30-cycle PCR in duplicate of 20µL reaction volumes using 5ng of DNA each, dual index primers,
344 and AccuPrime Pfx SuperMix (Invitrogen). The thermocycling procedure includes a 2min initial denaturation step
345 at 95°C, and 30 cycles of the following PCR scheme: (a) 20-second denaturation at 95°C, (b) 15-second annealing
346 at 55°C, and (c) 5-minute extension at 72°C. The duplicate PCR products of each sample were pooled, purified,
347 and normalized using SequalPrep PCR cleanup and normalization kit (Invitrogen). Barcoded amplicon libraries
348 were then pooled and sequenced using Illumina Miseq v2 reagent kit, which generated 2×250bp paired-end reads
349 at the Yale Center for Genome Analysis (YCGA). The sequencing reads were demultiplexed on QIIME 1.9.0 [43].
350 The barcodes, indexes, and primers were removed from raw reads, producing FASTQ files with both the forward
351 and reverse reads for each sample, ready for DADA2 analysis [36]. DADA2 version 1.1.6 was used to infer unique
352 biological exact sequence variants (ESVs) for each sample and naïve Bayes was used to assign taxonomy using
353 the SILVA version 123 database [44, 45].

354 **Metrics of community distance**

355 Beta-diversity indexes between the invasive and coalesced communities or the resident and coalesced communities
 356 were computed using various similarity metrics. For two arbitrary communities with ESV abundances represented
 357 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
 358 i th ESV in each community respectively and S is the total number of ESVs), the Bray-Curtis similarity $BC(\mathbf{x}, \mathbf{y})$
 359 is calculated as [46]

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

360 The Jensen-Shannon similarity $JS(\mathbf{x}, \mathbf{y})$ is defined as one minus the Jensen-Shannon distance (which is, in turn,
 361 the square root of the Jensen-Shannon divergence [47])

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

362 where $\mathbf{m} = (\mathbf{x} + \mathbf{y}) / 2$ and KL denotes the Kullback-Leibler divergence [48]

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

363 Using base-two logarithms ensures that the metric is bounded between 0 and 1. The Jaccard similarity is given by
 364 $J(\mathbf{x}, \mathbf{y})$ [49]

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

365 Additionally, we quantified coalescence outcomes by examining the fraction of the endemic cohort of the original
 366 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
 367 found in \mathbf{y} .

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

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

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

373 **Simulations**

374 We used the Community Simulator package [31] and included new features for our simulations. In the package,
 375 species are characterized by their resource uptake rates ($c_{i\alpha}$ for species i and resource α), and they all share a
 376 common metabolic matrix \mathbf{D} . The element $D_{\alpha\beta}$ of this matrix represents the fraction of energy in the form of
 377 resource α secreted when resource β is consumed. Here we implemented a new operation mode in which species
 378 can secrete different metabolites (and/or in different abundances) when consuming a same resource. We call $D_{i\alpha\beta}$ to
 379 the fraction of energy in the form of resource α secreted by species i when consuming resource β . In the Community
 380 Simulator underlying Microbial Consumer-Resource Model, this means that the energy flux $J_{i\beta}^{\text{out}}$ [29, 30] now takes
 381 the form

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

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

384 For our simulations, we first generated a library of 2400 species divided into three specialist families of 800
 385 members each and a generalist family of 240 members. We split this library into two non-overlapping pools of
 386 1320 species each. We randomly sampled 50 species from each pool in equal ratios to seed 100 resident and 100
 387 invasive communities respectively. We then let grow and diluted the communities serially, replenishing the primary
 388 resource after each dilution. We repeated the process 20 times to ensure generational equilibrium was achieved
 389 [29]. We then performed the *in silico* experiments by using the generationally stable communities to seed 100
 390 coalesced communities that were again stabilized as described previously. Similarly, we identified the dominant

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

393 Most other parameters were set to the defaults of the original Community Simulator package, with the only
 394 exception of the maintenance costs (m) which are set to zero for all species (equivalent to assuming cell death is
 395 negligible through the duration of our growth cycles) and the sparsity of the metabolic matrices (s) which is set to
 396 0.9 to generate significant variability in the secretion fluxes across different species (see main text).

397 Minimal model

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

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

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

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

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

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

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

408 Where we made the sub-dominants equally strong consumers of their dominants' secretions ($c_{22} = c_{44} = 100$),
 409 and we varied all other uptake rates depending on the scenario we were considering (see main text). Whenever
 410 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
 411 fixed value of 1 and let one in the numerator range within the specified limits.

412 **Data & code availability**

413 Experimental data and code for the analysis, as well as code for the simulations and the updated Community
414 Simulator package with instructions for enabling the new features are in github.com/jdiazc9/coalescence.

415 **Acknowledgements**

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

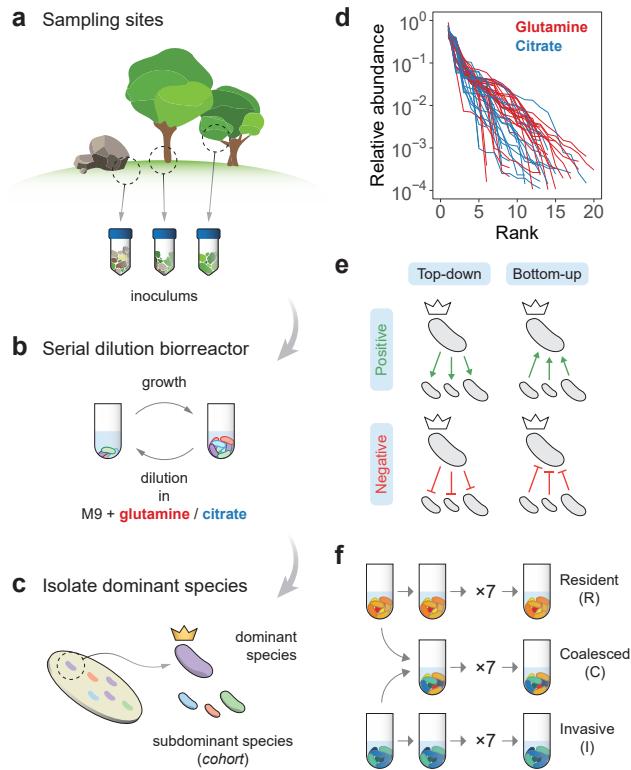
421 References

- 422 1. Mansour I, Heppell CM, Ryo M and Rillig MC (2018). Application of the microbial community coalescence
423 concept to riverine networks. *Biological Reviews* **93**(4):1832–1845
- 424 2. Luo X, Xiang X, Yang Y, Huang G, Fu K, Che R and Chen L (2020). Seasonal effects of river flow on microbial
425 community coalescence and diversity in a riverine network. *FEMS Microbiology Ecology* **96**(8):fiaa132
- 426 3. Vass M, Székely AJ, Lindström ES, Osman OA and Langenheder S (2021). Warming mediates the resistance
427 of aquatic bacteria to invasion during community coalescence. *Molecular Ecology* **30**(5):1345–1356
- 428 4. Rillig MC, Lehmann A, Aguilar-Trigueros CA, Antonovics J, Caruso T, Hempel S, Lehmann J, Valyi K,
429 Verbruggen E et al. (2016). Soil microbes and community coalescence. *Pedobiologia* **59**(1-2):37–40
- 430 5. Evans SE, Bell-Dereske LP, Dougherty KM and Kittredge HA (2019). Dispersal alters soil microbial commu-
431 nity response to drought. *Environmental Microbiology* **22**(3):905–916
- 432 6. Dutton CL, Subalusky AL, Sanchez A, Estrela S, Lu N, Hamilton SK, Njoroge L, Rosi EJ and Post DM
433 (2021). The meta-gut: Hippo inputs lead to community coalescence of animal and environmental micro-
434 biomes. *biorXiv*
- 435 7. Vandegrift R, Fahimipour AK, Muscarella M, Bateman AC, Wymelenberg KVD and Bohannan BJ (2019).
436 Moving microbes: the dynamics of transient microbial residence on human skin. *biorXiv*
- 437 8. Rillig MC, Antonovics J, Caruso T, Lehmann A, Powell JR, Veresoglou SD and Verbruggen E (2015). Inter-
438 change of entire communities: microbial community coalescence. *Trends in Ecology & Evolution* **30**(8):470–
439 476
- 440 9. Gilpin M (1994). Community-level competition: asymmetrical dominance. *Proceedings of the National
441 Academy of Sciences* **91**(8):3252–3254
- 442 10. Simberloff D and Holle BV (1999). Positive Interactions of Nonindigenous Species: Invasional Meltdown?
443 *Biological Invasions* **1**(1):21–32
- 444 11. Grosholz ED (2005). Recent biological invasion may hasten invasional meltdown by accelerating historical
445 introductions. *Proceedings of the National Academy of Sciences* **102**(4):1088–1091
- 446 12. Simberloff D (2006). Invasional meltdown 6 years later: important phenomenon, unfortunate metaphor, or
447 both? *Ecology Letters* **9**(8):912–919
- 448 13. Gurevitch J (2006). Commentary on Simberloff (2006): Meltdowns, snowballs and positive feedbacks. *Ecol-
449 ogy Letters* **9**(8):919–921
- 450 14. Green PT, O'Dowd DJ, Abbott KL, Jeffery M, Retallick K and Nally RM (2011). Invasional meltdown:
451 Invader-invader mutualism facilitates a secondary invasion. *Ecology* **92**(9):1758–1768
- 452 15. Livingston G, Jiang Y, Fox JW and Leibold MA (2013). The dynamics of community assembly under sudden
453 mixing in experimental microcosms. *Ecology* **94**(12):2898–2906
- 454 16. Prior KM, Robinson JM, Dunphy SAM and Frederickson ME (2015). Mutualism between co-introduced
455 species facilitates invasion and alters plant community structure. *Proceedings of the Royal Society B: Biolog-
456 ical Sciences* **282**(1800):20142846
- 457 17. O'Loughlin LS and Green PT (2017). Secondary invasion: When invasion success is contingent on other
458 invaders altering the properties of recipient ecosystems. *Ecology and Evolution* **7**(19):7628–7637
- 459 18. Castledine M, Sierociński P, Padfield D and Buckling A (2020). Community coalescence: an eco-evolutionary
460 perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**(1798):20190252
- 461 19. Toquenaga Y (1997). Historicity of a Simple Competition Model. *Journal of Theoretical Biology* **187**(2):175–
462 181
- 463 20. Tikhonov M (2016). Community-level cohesion without cooperation. *eLife* **5**:e15747
- 464 21. Tikhonov M and Monasson R (2017). Collective Phase in Resource Competition in a Highly Diverse Ecosys-
465 tem. *Physical Review Letters* **118**(4):048103

- 466 22. Vila JCC, Jones ML, Patel M, Bell T and Rosindell J (2019). Uncovering the rules of microbial community
467 invasions. *Nature Ecology & Evolution* **3**(8):1162–1171
- 468 23. Lechón P, Clegg T, Cook J, Smith TP and Pawar S (2021). The role of competition versus cooperation in
469 microbial community coalescence. *biorXiv*
- 470 24. Sierociński P, Milferstedt K, Bayer F, Großkopf T, Alston M, Bastkowski S, Swarbreck D, Hobbs PJ, Soyer OS
471 et al. (2017). A Single Community Dominates Structure and Function of a Mixture of Multiple Methanogenic
472 Communities. *Current Biology* **27**(21):3390–3395.e4
- 473 25. Rillig MC and Mansour I (2017). Microbial Ecology: Community Coalescence Stirs Things Up. *Current
474 Biology* **27**(23):R1280–R1282
- 475 26. Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DS, Parfrey LW, Farjalla VF and Doebeli M (2016).
476 High taxonomic variability despite stable functional structure across microbial communities. *Nature Ecology
477 & Evolution* **1**(1):0015
- 478 27. Winfree R, Fox JW, Williams NM, Reilly JR and Cariveau DP (2015). Abundance of common species, not
479 species richness, drives delivery of a real-world ecosystem service. *Ecology Letters* **18**(7):626–635
- 480 28. Rosenzweig RF, Sharp RR, Treves DS and Adams J (1994). Microbial evolution in a simple unstructured
481 environment: genetic differentiation in *Escherichia coli*. *Genetics* **137**(4):903–917
- 482 29. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, Segré D, Mehta P and Sanchez A
483 (2018). Emergent simplicity in microbial community assembly. *Science* **361**(6401):469–474
- 484 30. Marsland III R, Cui W, Goldford J, Sanchez A, Korolev K and Mehta P (2019). Available energy fluxes drive
485 a transition in the diversity, stability, and functional structure of microbial communities. *PLoS Computational
486 Biology* **15**(2):e1006793
- 487 31. Marsland R, Cui W, Goldford J and Mehta P (2020). The Community Simulator: A Python package for
488 microbial ecology. *PLoS ONE* **15**(3):e0230430
- 489 32. Estrela S, Vila JCC, Lu N, Bajic D, Rebollo-Gómez M, Chang CY and Sanchez A (2020). Metabolic rules
490 of microbial community assembly. *biorXiv*
- 491 33. Dimroth P (2004). Molecular Basis for Bacterial Growth on Citrate or Malonate. *EcoSal Plus* **1**(1)
- 492 34. Forchhammer K (2007). Glutamine signalling in bacteria. *Frontiers in Bioscience* **12**(1):358
- 493 35. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA and Holmes SP (2016). DADA2: High-
494 resolution sample inference from Illumina amplicon data. *Nature Methods* **13**(7):581–583
- 495 36. Callahan BJ, McMurdie PJ and Holmes SP (2017). Exact sequence variants should replace operational taxo-
496 nomic units in marker-gene data analysis. *The ISME Journal* **11**:2639–2643
- 497 37. Marsland R, Cui W and Mehta P (2020). A minimal model for microbial biodiversity can reproduce experi-
498 mentally observed ecological patterns. *Scientific Reports* **10**:3308
- 499 38. Estrela S, Sanchez-Gorostiaga A, Vila JC and Sanchez A (2021). Nutrient dominance governs the assembly
500 of microbial communities in mixed nutrient environments. *eLife* **10**
- 501 39. MacArthur R (1970). Species packing and competitive equilibrium for many species. *Theoretical Population
502 Biology* **1**(1):1–11
- 503 40. Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, Bonilla G, Kar A, Leiby N et al.
504 (2014). Metabolic Resource Allocation in Individual Microbes Determines Ecosystem Interactions and Spatial
505 Dynamics. *Cell Reports* **7**(4):1104–1115
- 506 41. Pinu FR, Granucci N, Daniell J, Han TL, Carneiro S, Rocha I, Nielsen J and Villas-Boas SG (2018). Metabolite
507 secretion in microorganisms: the theory of metabolic overflow put to the test. *Metabolomics* **14**(4)
- 508 42. Kozich JJ, Westcott SL, Baxter NT, Highlander SK and Schloss PD (2013). Development of a Dual-Index
509 Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina
510 Sequencing Platform. *Applied and Environmental Microbiology* **79**(17):5112–5120

- 511 43. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich
512 JK et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*
513 7:335–336
- 514 44. Wang Q, Garrity GM, Tiedje JM and Cole JR (2007). Naïve Bayesian Classifier for Rapid Assignment of
515 rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology* 73(16):5261–
516 5267
- 517 45. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J and Glöckner FO (2013). The SILVA
518 ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*
519 41(D1):D590–D596
- 520 46. Curtis JT and Bray JR (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin.
521 *Ecological Monographs* 27(4):325–349
- 522 47. Lin J (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*
523 37(1):145–151
- 524 48. Kullback S and Leibler RA (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*
525 22(1):79–86
- 526 49. Jaccard P (1912). The distribution of the flora in the alpine zone. *New Phytologist* 11(2):37–50

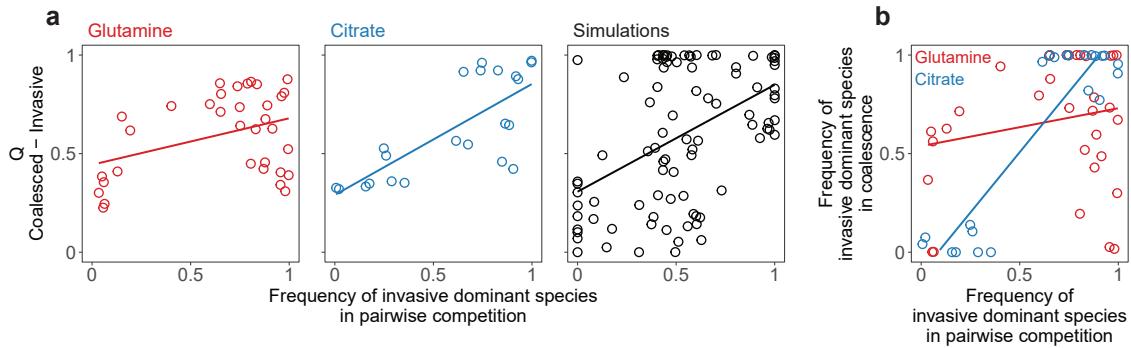
527 **Figures**



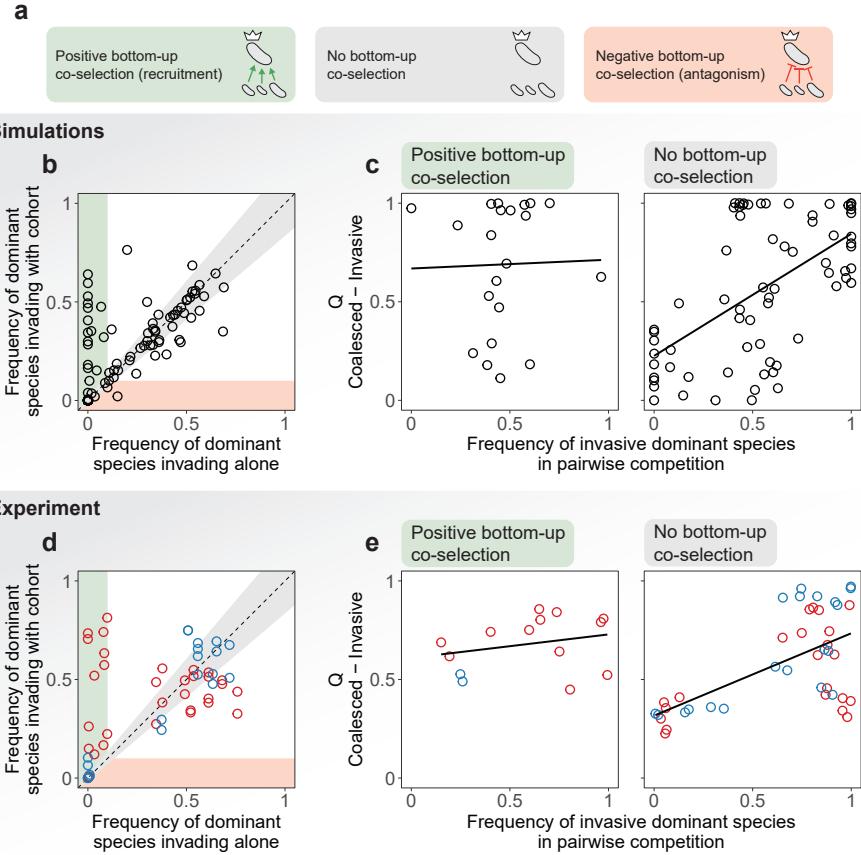
528

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

542

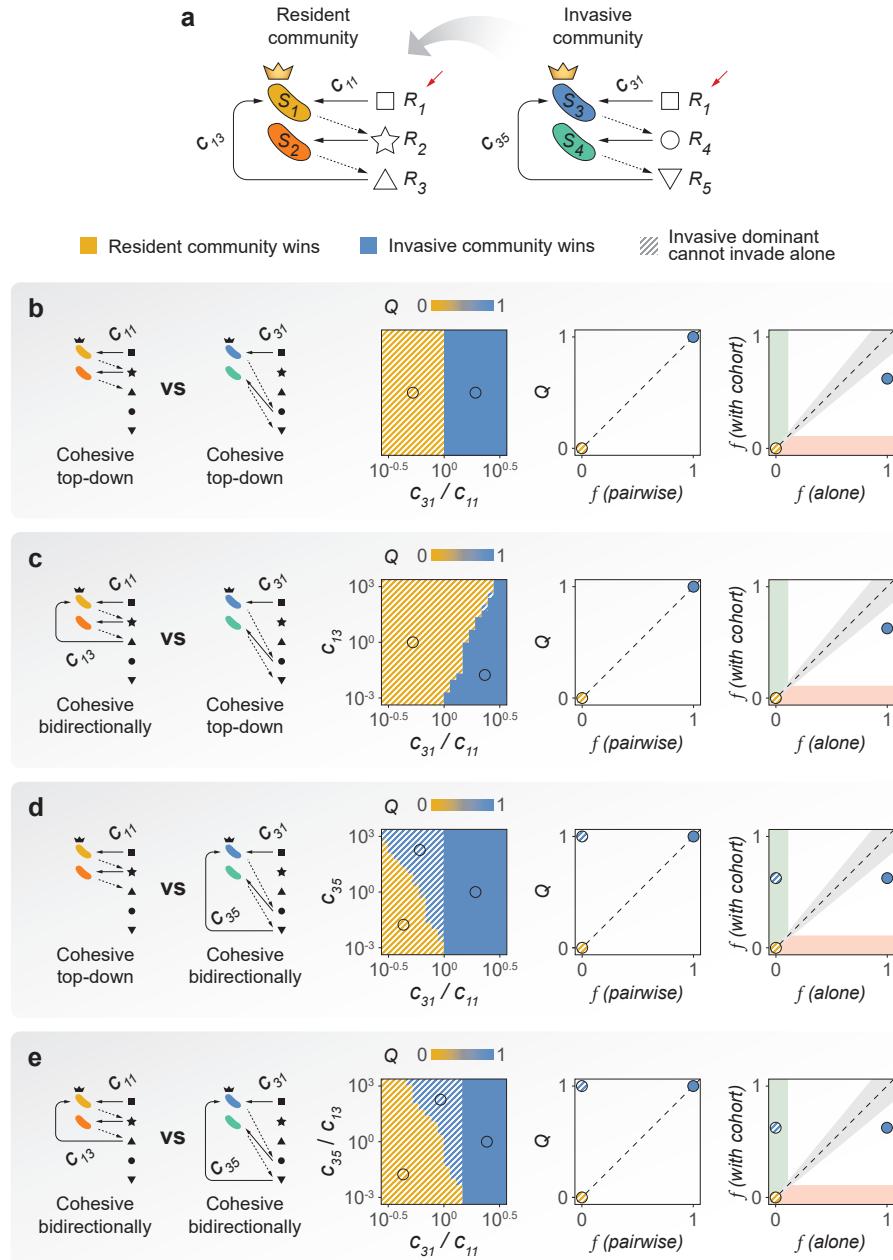


543 **Figure 2. Top-down co-selection in microbial community coalescence.** **a.** Coalescence outcomes are quantified by the
 544 relative Bray-Curtis similarity (Q) between the coalesced and invasive communities. These outcomes are predicted by the
 545 pairwise competition between the invasive and resident dominant species. Left panel (red): glutamine communities, $R^2 = 0.15$,
 546 $p < 0.05$. Middle panel (blue): citrate communities, $R^2 = 0.57$, $p < 10^{-4}$. A high correlation is consistent with a scenario of
 547 strong top-down positive co-selection where dominants recruit their cohorts for the final coalesced community. Two biological
 548 replicates per experiment are plotted individually. Right panel (black): simulations with a Microbial Consumer-Resource Model
 549 are able to capture these trends ($R^2 = 0.22$, $p < 10^{-5}$). **b.** Pairwise competition of dominants with or without their cohorts.
 550 In the horizontal axis, we plot the frequency of the invasive dominant species in head-to-head pairwise competition with the
 551 resident dominant. In the vertical axis, we plot the same relative frequency when the two species compete in the presence of
 552 their cohorts, i.e. during community coalescence. $R^2 = 0.04$, $p > 0.05$ for glutamine (red) and $R^2 = 0.83$, $p < 10^{-8}$ for citrate
 553 (blue).



555

556 **Figure 3. Trade offs between bottom-up and top-down ecological co-selection.** **a.** We hypothesize that three scenarios are
 557 possible regarding bottom-up co-selection: sub-dominant species could co-select for (green) or against (red) their dominant in
 558 coalescence, or they could have no effect in the invasion success of the dominant taxa (gray). **b.** Simulations with a Microbial
 559 Consumer-Resource Model: we plot the frequency reached by the invasive dominants when invading the resident communities
 560 in isolation versus the same frequency when invading together with their cohorts, i.e. in community coalescence. Points in the
 561 green/red area represent instances where the invasive dominant is able to invade with higher/lower success when accompanied
 562 by its cohort, evidencing positive/negative bottom-up co-selection. Points around the diagonal (gray area) correspond to cases
 563 where the success of the invasive dominant is only weakly affected by the presence or absence of its cohort. **c.** We divided
 564 the data from our simulations into two sets according to whether positive or no bottom-up co-selection was observed (that is,
 565 whether points fell into the green or gray areas of panel b). Here we reproduce the plots in Figure 2a for each set, representing
 566 the result of the dominant head-to-head pairwise competition versus the outcome of community coalescence. Left panel: strong
 567 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
 568 bottom-up co-selection. **e.** We reproduce the plots in panel c for our experimental data, i.e. we recreate Figure 2a but this
 569 time splitting our data by the strength of bottom-up co-selection instead of by the carbon source supplied to the communities.
 570 Left panel: strong positive bottom-up co-selection ($R^2 = 0.07, p > 0.05$). Right panel: no bottom-up co-selection ($R^2 = 0.37,$
 571 $p < 10^{-4}$).



574

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

583

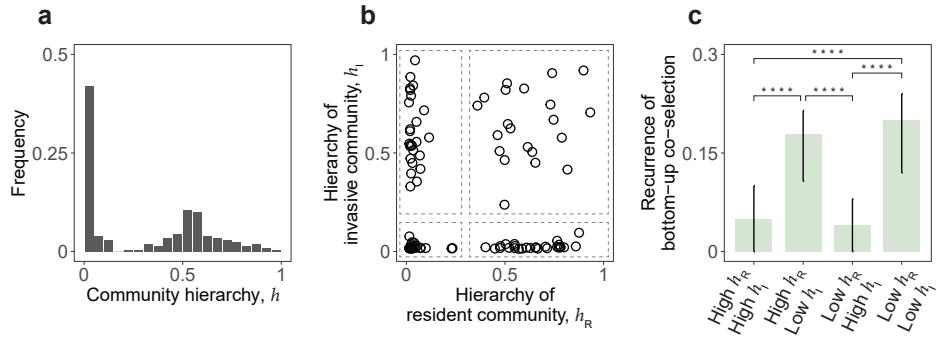
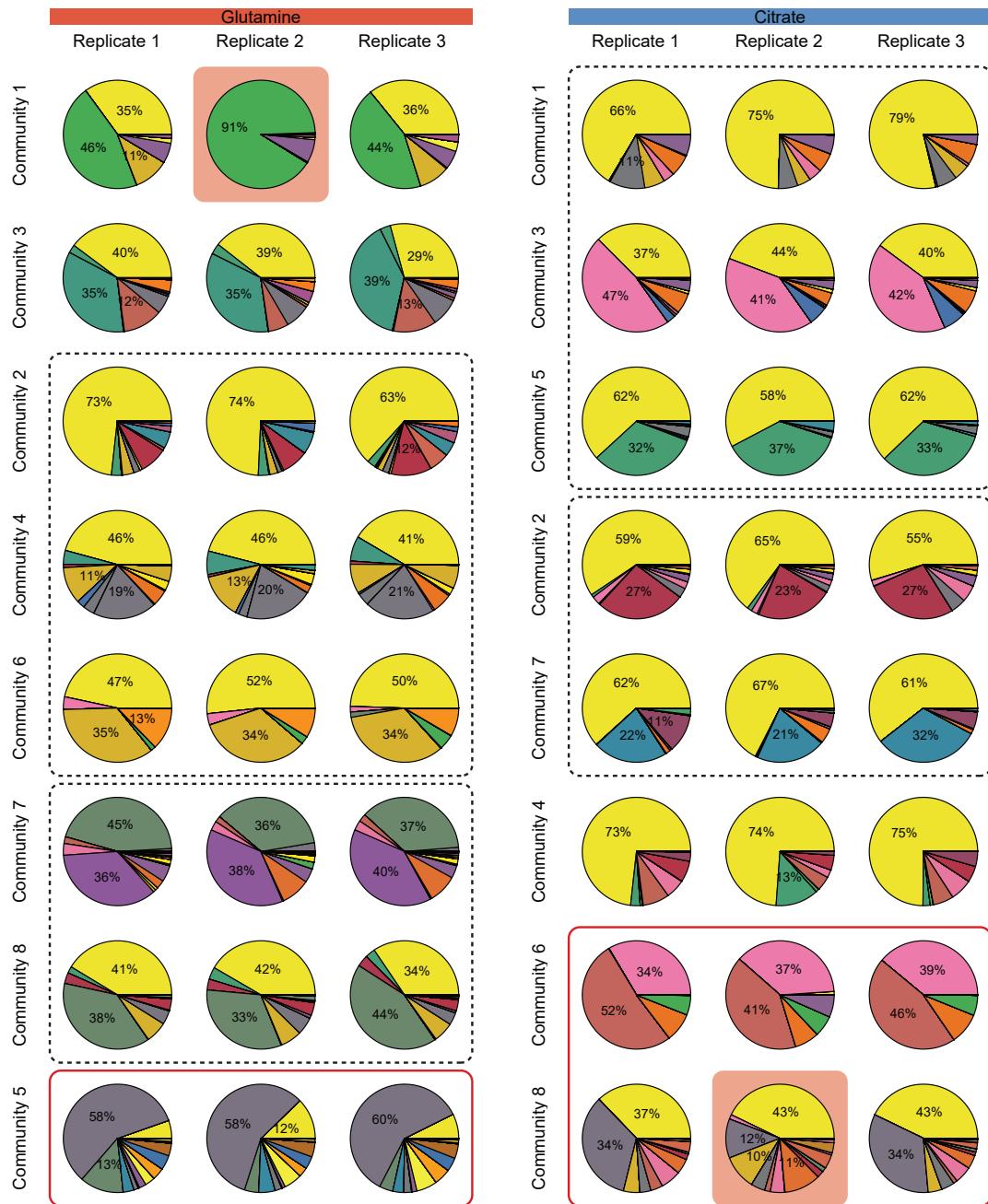


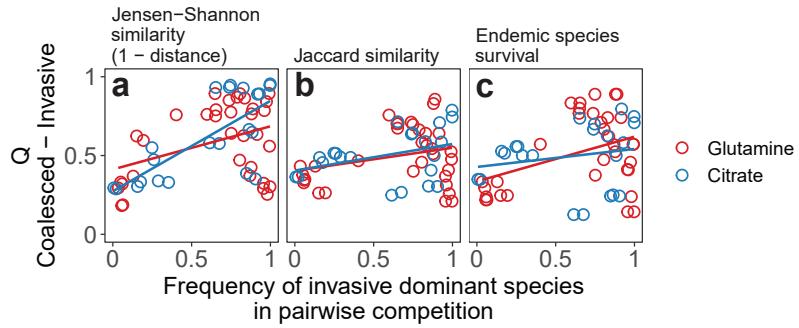
Figure 5. Community hierarchy modulates the recurrence of bottom-up co-selection. **a.** Distribution of community hierarchies for our *in silico* communities. **b.** We divided our coalescence simulations into four groups according to the hierarchies of the resident (h_R) and invasive (h_I) communities 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).

592 **Supplementary Figures**



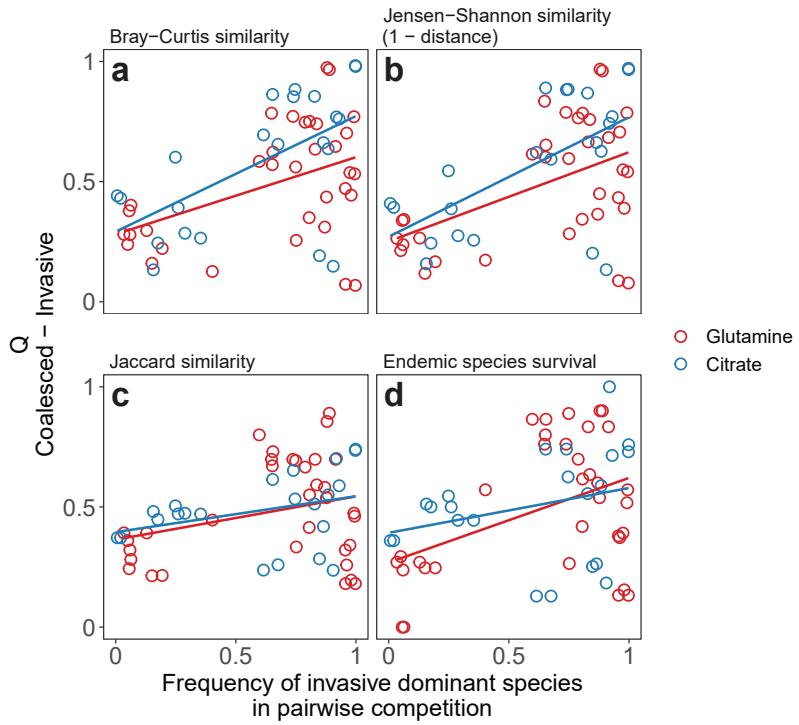
593

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



602

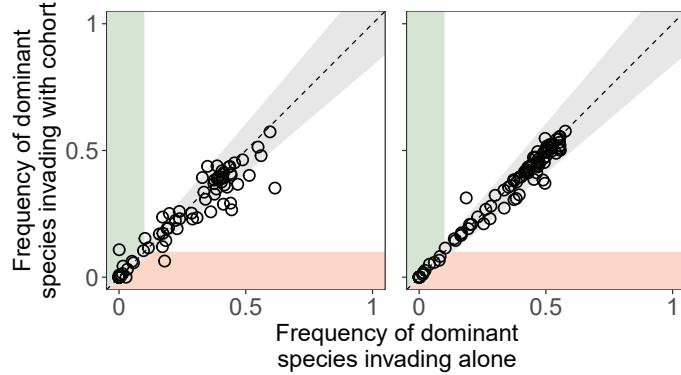
603 **Figure S2. Alternative metrics of community distance.** Quantifying coalescence outcomes using different metrics of commu-
 604 nity similarity (Methods: Metrics of community distance) gives similar results to those shown in Figure 2a. Metrics that account
 605 for the relative species abundances (Bray-Curtis or Jensen-Shannon similarities) yield higher correlations than less quantitative
 606 metrics that only account for species presence/absence (Jaccard similarity or the fraction of endemic invasive species persisting
 607 in the coalesced community). **a.** Relative Jensen-Shannon similarity ($R^2 = 0.15$, $p < 0.05$ for glutamine and $R^2 = 0.53$,
 608 $p < 5 \times 10^{-4}$ 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)
 609 **c.** Relative survival of invasive endemic species after coalescence ($R^2 = 0.16$, $p < 0.05$ for glutamine and $R^2 = 0.04$, $p > 0.05$
 610 for citrate).



612

613 **Figure S3. Dominant species have limited effects on coalescence outcomes quantification.** We repeated the analyses shown
 614 in [Figure 2a](#) and [Figure S2](#), but this time we removed the dominants from the compositional data prior to quantifying community
 615 distances. The trends observed before are maintained. **a.** Relative Bray-Curtis similarity ($R^2 = 0.20$, $p < 0.01$ for glutamine
 616 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 $R^2 = 0.36$,
 617 $p < 0.005$ 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.**
 618 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
 619 citrate).

621



622 **Figure S4. Bottom-up ecological co-selection is not observed when species have similar metabolic architectures.** We
 623 ran simulations of community coalescence following the same procedure described in the main text, but this time we used a
 624 dense metabolic matrix (left panel, sparsity = 0.05 in the Community Simulator package [31]) or a species-unspecific metabolic
 625 matrix (right panel, $D_{i\beta\alpha} = D_{j\beta\alpha}$ for all i, j, α and β , see Box 1). Virtually no bottom-up co-selection is observed in either case.