Cohesiveness in microbial community coalescence

Juan Diaz-Colunga[1,*], Nanxi Lu[1,*], Alicia Sanchez-Gorostiaga[1,2,*], Joshua E. Goldford[3], Mikhail Tikhonov[4], and Álvaro Sánchez[1,†]

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

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

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

[4]*Department of Physics, Center for Science & Engineering of Living Systems, Washington University in St. Louis, St. Louis, MO, USA*

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

Our results indicate that co-selection varies in direction and strength depending on the supplied limiting resource. This primary resource, in turn, has been shown to shape the structure and interactions of the communities [30]. We observe that, when top-down co-selection is weak, bottom-up co-selection can be very strong, with positive co-selection being far more common than negative co-selection. We then turn to a Microbial Consumer-Resource Model (MicroCRM) [29, 31, 32] that is able to capture the dynamics of microbial communities dominated by metabolic interactions, as is the case for the ones assembled in our experimental conditions [29, 30]. We show that the empirically observed trends in ecological co-selection are reproduced with minimal model assumptions, and that tuning the complexity of the metabolic interactions in our *in silico* communities can modulate the recurrence of top-down or bottom-up co-selection. Our findings indicate that collective interactions play an important role at dictating community structure during coalescence.

Results & Discussion

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

Top-down ecological co-selection

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

We found that, for communities assembled in the glutamine environment, the relative frequency of a dominant against another in head-to-head pairwise competition is barely predictive of its relative frequency against that same other dominant when the cohorts are present too, i.e. during community coalescence (Figure 2a red dots, $R^2 = 0.04$, $p > 0.05$). This correlation is significantly higher for the citrate communities (Figure 2a blue dots, $R^2 = 0.83$, $p < 10^{-8}$). This suggests that, in the glutamine environments, head-to-head competition of dominants is heavily influenced by higher order effects introduced by the rare taxa of the communities. On the other hand, the cohorts seem to play a more passive role in the citrate environments. 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 again noticed differences between glutamine and citrate communities: 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 2b left panel, $R^2 = 0.15$, $p < 0.05$). For the latter, the structure of the coalesced communities tends to be more strongly dictated by the result of the dominant-dominant competition (Figure 2b middle panel, $R^2 = 0.57$, $p < 10^{-4}$). 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).

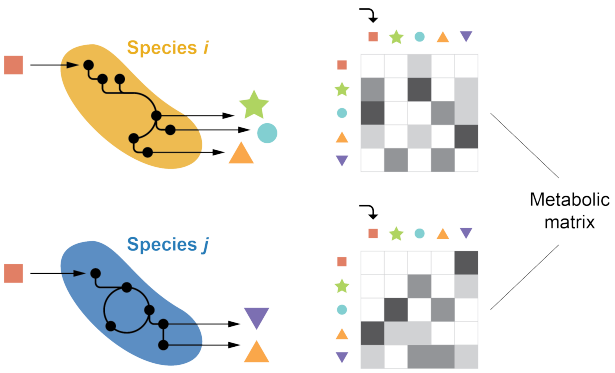
Together, these observations suggest that the strength of top-down co-selection depends on the environment where communities are assembled and coalescence takes place. Communities stabilized with citrate as the primary supplied resource display a strong degree of top-down cohesiveness, with the fates 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, 31] as implemented in the Community Simulator package for Python [32]. We chose this modeling framework because communities assembled under our experimental conditions (natural microbiomes re-assembled into multispecies 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, 30] for which the MicroCRM provides a good description. We carried out our simulations enabling a large variation across species in terms of their metabolic architectures; this choice is discussed in Box 1. 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 resident communities respectively, that were allowed 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 was able to capture a correlation between the head-to-head pairwise competition of dominants and the outcome of community coalescence (Figure 2b, right panel), further supporting the idea that top-down ecological co-selection can consistently emerge from metabolic interactions across species.

Box 1: Community cohesiveness in a Microbial Consumer-Resource Model

The Microbial Consumer-Resource Model (MicroCRM) [29, 31, 32] is a modeling framework based on the classic MacArthur's consumer resource model [37]. It encodes the dynamics of a system with S species and M resources in terms of a consumer preference matrix $\mathbf{c}$ and a metabolic matrix $\mathbf{D}$, with an additional set of parameters controlling the species maintenance costs (m_i for species i), the resource energy densities (w_α for resource α), the energy to growth rate conversion factor (g_i for species i) or the leakage fraction, i.e. the amount of energy lost as byproducts when a resource is consumed (l_α for resource α). The element $c_{i\alpha}$ of the consumer preference matrix represents the uptake rate of resource α by species i (although the relationship between $c_{i\alpha}$ and the uptake rate can be more complex in modeling scenarios that are not considered here, see [29, 31, 32]). The element $D_{\beta\alpha}$ of the metabolic matrix represents the amount of energy secreted in the form of resource β as a result of the metabolism of resource α.

This formulation assumes that all species metabolize resources through similar pathways, with any two species secreting the same byproducts when consuming a same resource. However, experimental evidence suggests that individual species can secrete different sets of metabolites to the environment when growing on a same primary resource [30, 38, 39]. This observation motivated us to introduce a new feature in the model: in short, we now define $\mathbf{D}$ as a three-dimensional matrix where the element $D_{i\beta\alpha}$ represents the energy flux in the form of resource β that is secreted *by species* i when it metabolizes resource α. Note that now $D_{i\beta\alpha}$ need not be equal to $D_{j\beta\alpha}$ if $i \neq j$ (see illustration below).



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

Bottom-up co-selection during community coalescence

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

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

We then asked whether this trend was also observed *in vitro*. We went back to our synthetic communities and carried out a new round of experiments where we invaded the resident communities with the invasive dominants alone (Methods: Coalescence, competition and invasion experiments). After stabilization, we quantified species abundance through 16S Illumina sequencing (Methods: Determination of community composition by 16S sequencing). Again, we observed that bottom-up co-selection is far more common in its positive than in its negative form (Figure 3d). Interestingly, bottom-up recruitment appears to be more frequent in the glutamine environments than in the citrate ones, consistent with our hypothesis that metabolic interactions among species are key in determining the strength and direction of ecological co-selection. We then repeated our analysis in Figure 3c, this time splitting our data according to the observed strength of bottom-up co-selection instead of the primary carbon source as we had done in Figure 2b. Our findings were in line with the model prediction: pairwise competition between dominants is only predictive of coalescence outcomes if bottom-up co-selection is weak.

Conclusions

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

Our data suggests that the strength and direction of ecological co-selection is modulated by the underlying metabolic networks that shape the structure of communities assembled in synthetic minimal conditions [29, 30]. This network is in turn regulated by the supplied primary carbon source. This idea is supported by the observation that a Microbial Consumer-Resource Model captures the trends observed experimentally when we enable a large variation in the metabolic fluxes across species. The model also predicts a trade-off between the strength of bottom-up co-selection and the ability of dominant-dominant pairwise competition

to dictate coalescence outcomes, which we have confirmed experimentally. These observations, together with previous results in different systems [24] as well as theoretical predictions [9, 19–23], suggest that collective interactions between microbes and the environment should be generically expected to produce ecological co-selection during community coalescence.

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

Methods

Stabilization of environmental communities in simple synthetic environments

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

Isolation of dominant species

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

Coalescence, competition and invasion experiments

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

Determination of community composition by 16S sequencing

The sequencing protocol was identical to that described in [29]. Community samples were collected by spinning down at 3500rpm for 25min in a bench-top centrifuge at room temperature; cell pellets were stored at −80°C before processing. To maximize Gram-positive bacteria cell wall lysis, the cell pellets were re-suspended and incubated at 37°C for 30min in enzymatic lysis buffer (20mM Tris-HCl, 2mM sodium EDTA, 1.2% Triton X-100) and 20mg/mL of lysozyme from chicken egg white (Sigma-Aldrich). After cell lysis, the DNA extraction and purification was performed using the DNeasy 96 protocol for animal tissues (Qiagen). The clean DNA in 100µL elution buffer of 10mM Tris-HCl, 0.5mM EDTA at pH 9.0 was quantified using Quan-iT PicoGreen dsDNA Assay Kit (Molecular Probes, Inc.) and normalized to 5ng/µL in nuclease-free water (Qiagen) for subsequent 16S rRNA Illumina sequencing. 16S rRNA amplicon library preparation was performed following a dual-index paired-end approach [40]. Briefly, PCR amplicon libraries of V4 regions of the 16S rRNA were prepared sing dual-index primers (F515/R805), then pooled and sequenced using the Illumina MiSeq chemistry and platform. Each sample went through a 30-cycle PCR in duplicate of 20µL reaction volumes using 5ng of DNA each, dual index primers, and AccuPrime Pfx SuperMix (Invitrogen). The thermocycling procedure includes a 2min initial denaturation step at 95°C, and 30 cycles of the following PCR scheme: (a) 20-second denaturation at 95°C, (b) 15-second annealing at 55°C, and (c) 5-minute extension at 72°C. The duplicate PCR products of each sample were pooled, purified, and normalized using SequalPrep PCR cleanup and normalization kit (Invitrogen). Barcoded amplicon libraries were then pooled and sequenced using Illumina Miseq v2 reagent kit, which

generated 2×250bp paired-end reads at the Yale Center for Genome Analysis (YCGA). The sequencing reads were demultiplexed on QIIME 1.9.0 [41]. The barcodes, indexes, and primers were removed from raw reads, producing FASTQ files with both the forward and reverse reads for each sample, ready for DADA2 analysis [36]. DADA2 version 1.1.6 was used to infer unique biological exact sequence variants (ESVs) for each sample and naïve Bayes was used to assign taxonomy using the SILVA version 123 database [42, 43].

Metrics of community distance

Beta-diversity indexes between the invasive and coalesced communities or the resident and coalesced communities were computed using various similarity metrics. For two arbitrary communities with ESV abundances represented by the vectors $\mathbf{x} = (x_1, x_2, \cdots, x_N)$ and $\mathbf{y} = (y_1, y_2, \cdots, y_N)$ (where x_i and y_i represent the relative abundance of the ith ESV in each community respectively and N is the total number of ESVs), the Bray-Curtis similarity $BC(\mathbf{x}, \mathbf{y})$ is calculated as [44]

$$BC(\mathbf{x}, \mathbf{y}) = \sum_i \min(x_i, y_i) \tag{1}$$

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

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

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

$$KL(\mathbf{x}, \mathbf{y}) = \sum_i x_i \log_2\left(\frac{x_i}{y_i}\right) \tag{3}$$

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

$$J(\mathbf{x}, \mathbf{y}) = \frac{|\,\mathbf{x} \cap \mathbf{y}\,|}{|\,\mathbf{x} \cup \mathbf{y}\,|} \tag{4}$$

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

For all the metrics above, we quantified the relative similarity between the invasive and the coalesced communities using relative metrics (denoted as Q):

$$Q(\mathbf{x}_\mathrm{I}, \mathbf{x}_\mathrm{R}, \mathbf{x}_\mathrm{C}) = \frac{F(\mathbf{x}_\mathrm{I}, \mathbf{x}_\mathrm{C})}{F(\mathbf{x}_\mathrm{I}, \mathbf{x}_\mathrm{C}) + F(\mathbf{x}_\mathrm{R}, \mathbf{x}_\mathrm{C})} \tag{5}$$

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

Simulations

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

$$J^{\mathrm{out}}_{i\beta} = \sum_\alpha D_{i\beta\alpha} l_\alpha J^{\mathrm{in}}_{i\alpha} \tag{6}$$

The documentation for the Community Simulator contains detailed descriptions of the model formulation, parameters and package use. For the updated package with the new functionality, see Data & code availability.

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

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

Data & code availability

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

Acknowledgements

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

References

1. Mansour I, Heppell CM, Ryo M and Rillig MC (2018). Application of the microbial community coalescence concept to riverine networks. *Biological Reviews* **93(4)**:1832–1845

2. Luo X, Xiang X, Yang Y, Huang G, Fu K, Che R and Chen L (2020). Seasonal effects of river flow on microbial community coalescence and diversity in a riverine network. *FEMS Microbiology Ecology* **96(8)**:fiaa132

3. Vass M, Székely AJ, Lindström ES, Osman OA and Langenheder S (2021). Warming mediates the resistance of aquatic bacteria to invasion during community coalescence. *Molecular Ecology* **30(5)**:1345–1356

4. Rillig MC, Lehmann A, Aguilar-Trigueros CA, Antonovics J, Caruso T, Hempel S, Lehmann J, Valyi K, Verbruggen E et al. (2016). Soil microbes and community coalescence. *Pedobiologia* **59(1-2)**:37–40

5. Evans SE, Bell-Dereske LP, Dougherty KM and Kittredge HA (2019). Dispersal alters soil microbial community response to drought. *Environmental Microbiology* **22(3)**:905–916

6. Dutton CL, Subalusky AL, Sanchez A, Estrela S, Lu N, Hamilton SK, Njoroge L, Rosi EJ and Post DM (2021). The meta-gut: Hippo inputs lead to community coalescence of animal and environmental microbiomes. *biorXiv*

7. Vandegrift R, Fahimipour AK, Muscarella M, Bateman AC, Wymelenberg KVD and Bohannan BJ (2019). Moving microbes: the dynamics of transient microbial residence on human skin. *biorXiv*

8. Rillig MC, Antonovics J, Caruso T, Lehmann A, Powell JR, Veresoglou SD and Verbruggen E (2015). Interchange of entire communities: microbial community coalescence. *Trends in Ecology & Evolution* **30(8)**:470–476

9. Gilpin M (1994). Community-level competition: asymmetrical dominance. *Proceedings of the National Academy of Sciences* **91(8)**:3252–3254

10. Simberloff D and Holle BV (1999). Positive Interactions of Nonindigenous Species: Invasional Meltdown? *Biological Invasions* **1(1)**:21–32

11. Grosholz ED (2005). Recent biological invasion may hasten invasional meltdown by accelerating historical introductions. *Proceedings of the National Academy of Sciences* **102(4)**:1088–1091

12. Simberloff D (2006). Invasional meltdown 6 years later: important phenomenon, unfortunate metaphor, or both? *Ecology Letters* **9(8)**:912–919

13. Gurevitch J (2006). Commentary on Simberloff (2006): Meltdowns, snowballs and positive feedbacks. *Ecology Letters* **9(8)**:919–921

14. Green PT, O'Dowd DJ, Abbott KL, Jeffery M, Retallick K and Nally RM (2011). Invasional meltdown: Invader–invader mutualism facilitates a secondary invasion. *Ecology* **92(9)**:1758–1768

15. Livingston G, Jiang Y, Fox JW and Leibold MA (2013). The dynamics of community assembly under sudden mixing in experimental microcosms. *Ecology* **94(12)**:2898–2906

16. Prior KM, Robinson JM, Dunphy SAM and Frederickson ME (2015). Mutualism between co-introduced species facilitates invasion and alters plant community structure. *Proceedings of the Royal Society B: Biological Sciences* **282(1800)**:20142846

17. O'Loughlin LS and Green PT (2017). Secondary invasion: When invasion success is contingent on other invaders altering the properties of recipient ecosystems. *Ecology and Evolution* **7(19)**:7628–7637

18. Castledine M, Sierocinski P, Padfield D and Buckling A (2020). Community coalescence: an eco-evolutionary perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375(1798)**:20190252

19. Toquenaga Y (1997). Historicity of a Simple Competition Model. *Journal of Theoretical Biology* **187(2)**:175–181

20. Tikhonov M (2016). Community-level cohesion without cooperation. *eLife* **5**:e15747

21. Tikhonov M and Monasson R (2017). Collective Phase in Resource Competition in a Highly Diverse Ecosystem. *Physical Review Letters* **118(4)**:048103

22. Vila JCC, Jones ML, Patel M, Bell T and Rosindell J (2019). Uncovering the rules of microbial community invasions. *Nature Ecology & Evolution* **3(8)**:1162–1171

23. Lechón P, Clegg T, Cook J, Smith TP and Pawar S (2021). The role of competition versus cooperation in microbial community coalescence. *biorXiv*

24. Sierocinski P, Milferstedt K, Bayer F, Großkopf T, Alston M, Bastkowski S, Swarbreck D, Hobbs PJ, Soyer OS et al. (2017). A Single Community Dominates Structure and Function of a Mixture of Multiple Methanogenic Communities. *Current Biology* **27(21)**:3390–3395.e4

25. Rillig MC and Mansour I (2017). Microbial Ecology: Community Coalescence Stirs Things Up. *Current Biology* **27(23)**:R1280–R1282

26. Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DS, Parfrey LW, Farjalla VF and Doebeli M (2016). High taxonomic variability despite stable functional structure across microbial communities. *Nature Ecology & Evolution* **1(1)**:0015

27. Winfree R, Fox JW, Williams NM, Reilly JR and Cariveau DP (2015). Abundance of common species, not species richness, drives delivery of a real-world ecosystem service. *Ecology Letters* **18(7)**:626–635

28. Rosenzweig RF, Sharp RR, Treves DS and Adams J (1994). Microbial evolution in a simple unstructured environment: genetic differentiation in Escherichia coli. *Genetics* **137(4)**:903–917

29. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, Segrè D, Mehta P and Sanchez A (2018). Emergent simplicity in microbial community assembly. *Science* **361(6401)**:469–474

30. Estrela S, Vila JCC, Lu N, Bajic D, Rebolleda-Gomez M, Chang CY and Sanchez A (2020). Metabolic rules of microbial community assembly. *biorXiv*

31. Marsland III R, Cui W, Goldford J, Sanchez A, Korolev K and Mehta P (2019). Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities. *PLoS Computational Biology* **15(2)**:e1006793

32. Marsland R, Cui W, Goldford J and Mehta P (2020). The Community Simulator: A Python package for microbial ecology. *PLoS ONE* **15(3)**:e0230430

33. Dimroth P (2004). Molecular Basis for Bacterial Growth on Citrate or Malonate. *EcoSal Plus* **1(1)**

34. Forchhammer K (2007). Glutamine signalling in bacteria. *Frontiers in Bioscience* **12(1)**:358

35. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA and Holmes SP (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13(7)**:581–583

36. Callahan BJ, McMurdie PJ and Holmes SP (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal* **11**:2639–2643

37. MacArthur R (1970). Species packing and competitive equilibrium for many species. *Theoretical Population Biology* **1(1)**:1–11

38. Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, Bonilla G, Kar A, Leiby N et al. (2014). Metabolic Resource Allocation in Individual Microbes Determines Ecosystem Interactions and Spatial Dynamics. *Cell Reports* **7(4)**:1104–1115

39. Pinu FR, Granucci N, Daniell J, Han TL, Carneiro S, Rocha I, Nielsen J and Villas-Boas SG (2018). Metabolite secretion in microorganisms: the theory of metabolic overflow put to the test. *Metabolomics* **14(4)**

40. Kozich JJ, Westcott SL, Baxter NT, Highlander SK and Schloss PD (2013). Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology* **79(17)**:5112–5120

41. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**:335–336

42. Wang Q, Garrity GM, Tiedje JM and Cole JR (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology* **73(16)**:5261–5267

43. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J and Glöckner FO (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41(D1)**:D590–D596

44. Curtis JT and Bray JR (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs* **27(4)**:325–349

45. Lin J (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory* **37(1)**:145–151

46. Kullback S and Leibler RA (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics* **22(1)**:79–86

47. Jaccard P (1912). The distribution of the flora in the alpine zone. *New Phytologist* **11(2)**:37–50

Figures

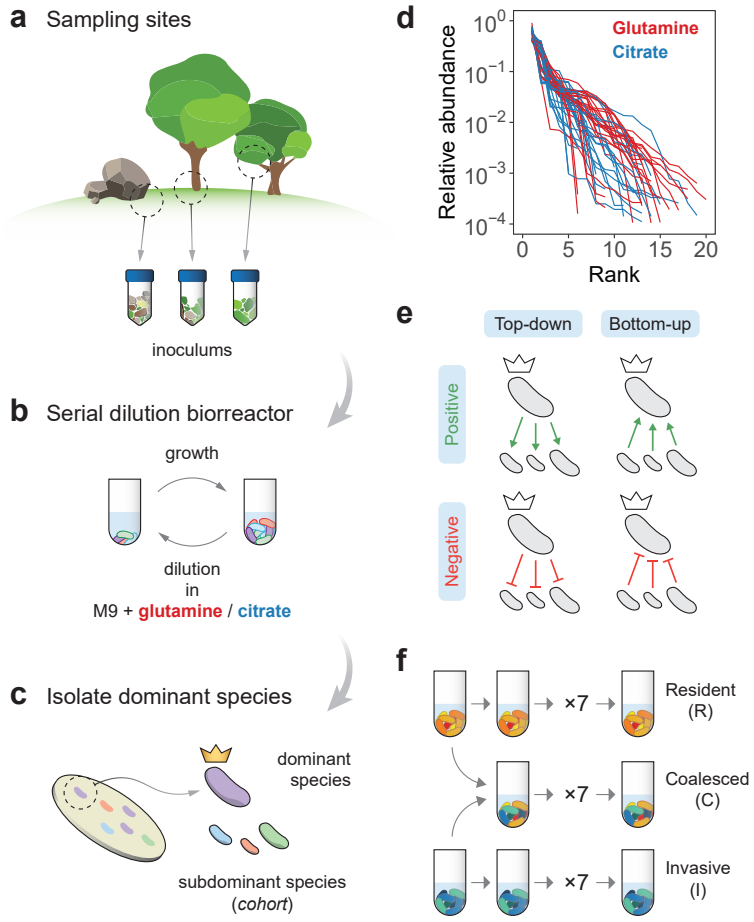


Figure 1. Overview of the experimental protocol. a. Environmental samples collected from eight different locations were used to inoculate our communities. **b.** Communities were stabilized in serial batch culture bioreactors [29] in minimal synthetic media with glutamine or citrate as the only supplied 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 all 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.

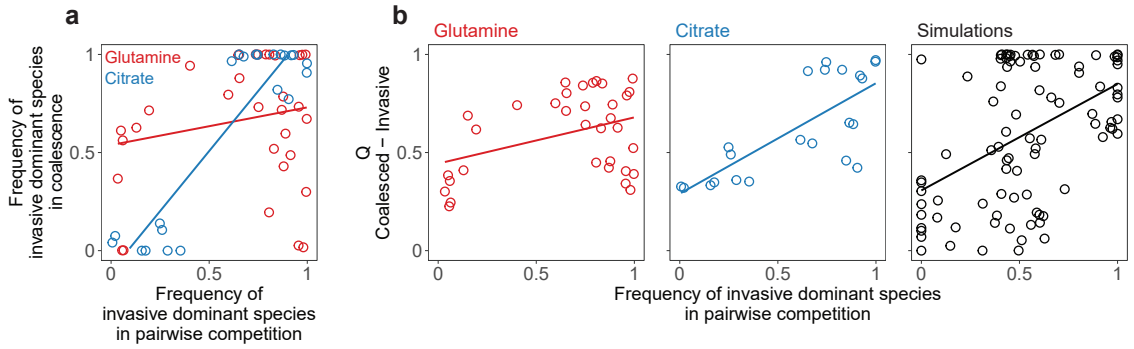


Figure 2. Top-down co-selection in microbial community coalescence. a. Pairwise competition of dominants with or without their cohorts. In the horizontal axis, we plot the frequency of the invasive dominant species in head-to-head pairwise competition with the resident dominant. In the vertical axis,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

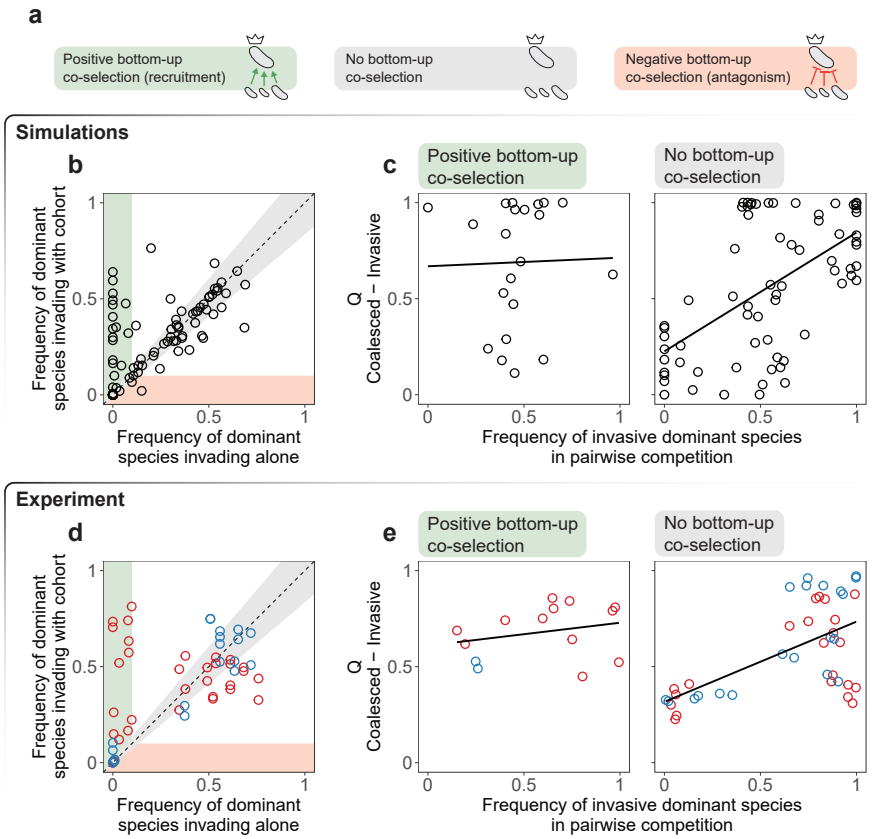


Figure 3. Trade offs between bottom-up and top-down ecological co-selection. **a.** We hypothesize that three scenarios are possible regarding bottom-up co-selection: sub-dominant species could co-select for (green) or against (red) their dominant in coalescence, or they could have no effect in the invasion success of the dominant taxa (gray). **b.** Simulations with a Microbial Consumer-Resource Model suggest that positive bottom-up co-selection is common and can be very strong, whereas negative bottom-up co-selection is rare. We plot the frequency reached by the invasive dominants when invading the resident communities in isolation versus the same frequency when invading together with their cohorts, i.e. in community coalescence. Points in the green/red area represent instances where the invasive dominant is able to invade with higher/lower success when accompanied by its cohort, evidencing positive/negative bottom-up co-selection. Points around the diagonal (gray area) correspond to cases where the success of the invasive dominant is only weakly affected by the presence or absence of its cohort. **c.** We divided the data from our simulations into two sets according to whether positive or no bottom-up co-selection was observed (that is, whether points fell into the green or gray areas of panel b). Here we reproduce the plots in Figure 2b for each set, representing the result of the dominant head-to-head pairwise competition versus the outcome of community coalescence. Left panel: when positive bottom-up co-selection is strong, head-to-head pairwise competition of dominants is poorly predictive of coalescence outcomes ($R^2 = 0.00$, $p > 0.05$). Right panel: on the other hand, when bottom-up co-selection is weak coalescence outcomes are more strongly dictated by the result of the dominant-dominant competition ($R^2 = 0.34$, $p < 10^{-6}$). **d.** Experiments show that in our conditions, positive bottom-up co-selection is indeed more frequent and strong than negative bottom-up co-selection. **e.** We reproduce the plots in panel c for our experimental data. This is equivalent to recreating Figure 2b, this time splitting our data by the strength of bottom-up co-selection instead of by the carbon source provided to the communities. Like simulations predicted, instances of community coalescence displaying strong positive bottom-up co-selection yield a poor correspondence between dominant pairwise competition and coalescence outcomes (left panel, $R^2 = 0.07$, $p > 0.05$) and vice-versa (right panel, $R^2 = 0.37$, $p < 10^{-4}$).

Supplementary Figures

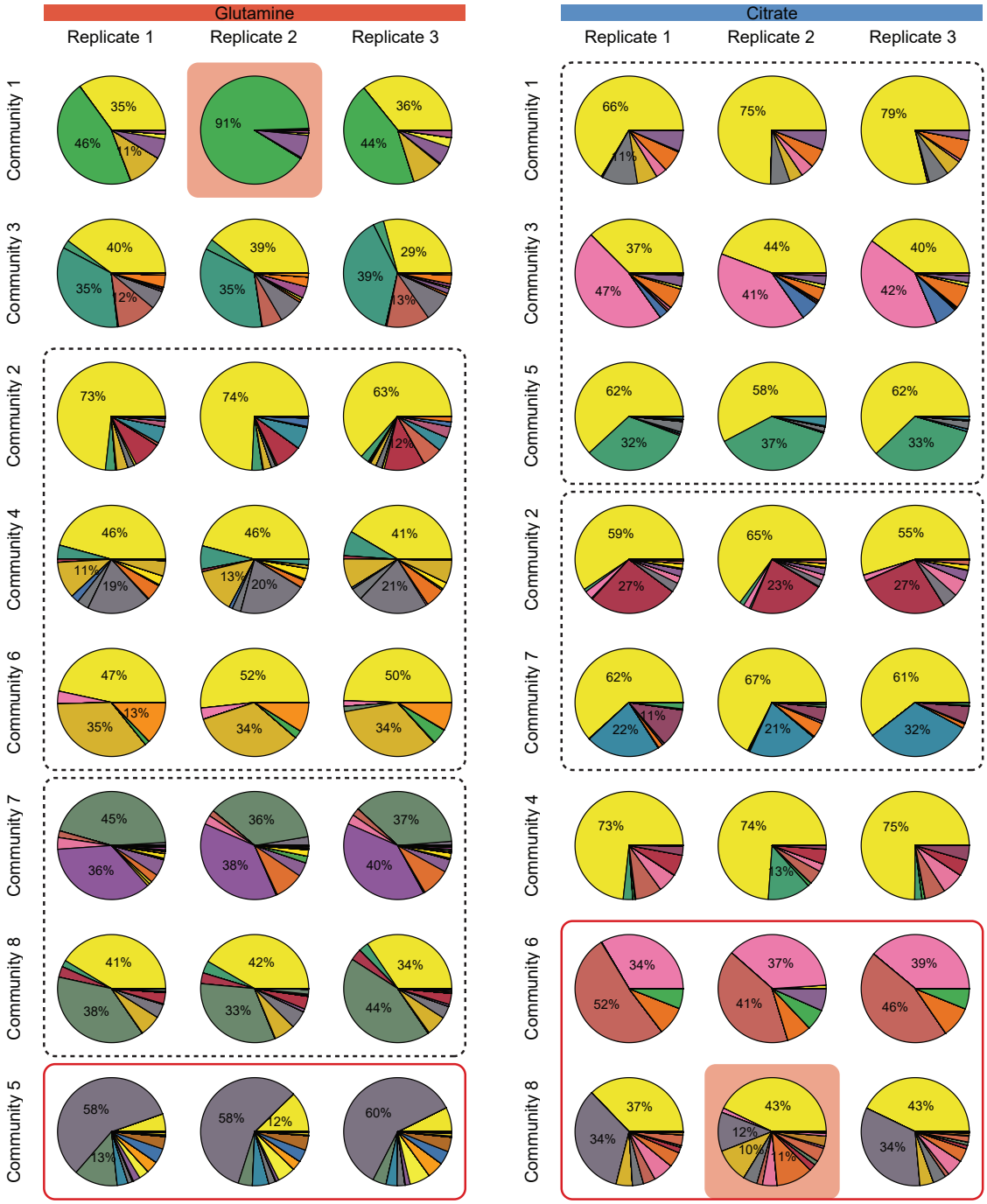


Figure S1. Community compositions after seven additional transfers without coalescence. Each color of the pie plots corresponds to a different exact sequence variant (Methods: Determination of community composition by 16S sequencing). Replicate 2 of community 1 from glutamine, as well as replicate 2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

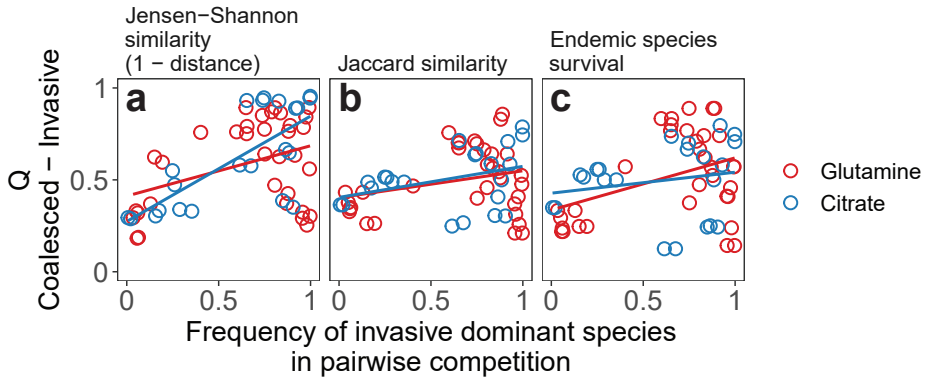


Figure S2. Alternative metrics of community distance. Quantifying coalescence outcomes using different metrics of community similarity (Methods: Metrics of community distance) gives similar results to those shown in Figure 2a. Metrics that account for the relative species abundances (Bray-Curtis or Jensen-Shannon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

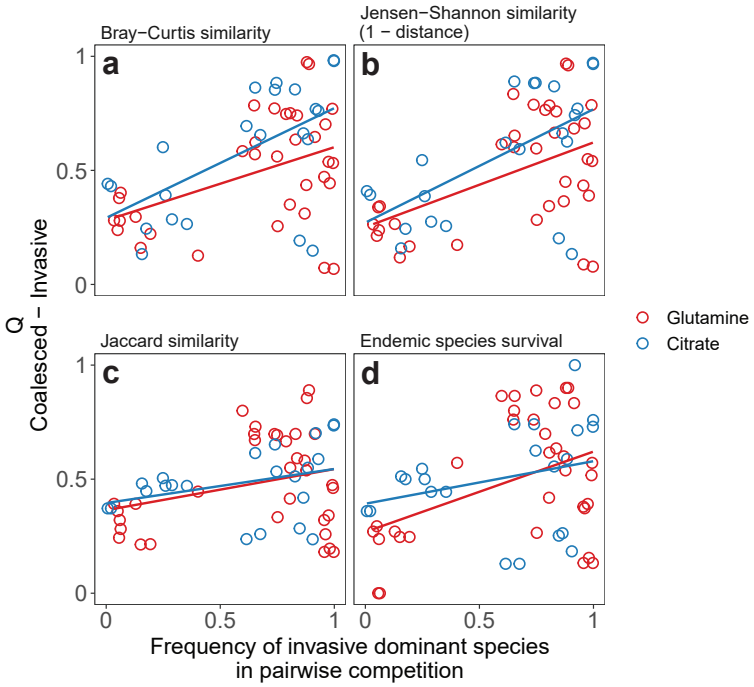


Figure S3. Dominant species have limited effects on coalescence outcomes quantification. We repeated the analyses shown in Figure 2a and Figure S2, but this time we removed the dominants from the compositional data prior to quantifying community distances. The trends observed before are maintained. **a.**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