

Helen S. Cai
Candidate for B.S. in Ecology & Evolutionary Biology
Yale College '20

Advisors: Dr. Alvaro Sanchez & Chang-Yu Chang
Report for E&EB 476
April 2020

Trends in pairwise species competition are related to results of two-community microbial coalescence

Abstract

Microbial coalescence describes the encounter and population fluctuation of two or more microbial communities in the same physical environment. Outcomes of microbial coalescence depend on abiotic factors, nutrient resources available in the environment, and inter-species metabolic interactions. Here we use *in silico* methods to model communities and predict the final composition of coalesced communities. Individual species that are prominent in these preliminary communities are identified and competed against each other. We find that the abundance of the dominant invasive species after direct competition with another species predicts the success of the same species in the outcome of community coalescence.

1. Introduction

1.1 Community coalescence: an understanding of population coexistence

Microbes are ubiquitous in the world around us. Microscopic interactions between these microns-long organisms characterize trophic levels within both aerobic and anaerobic environments. The sheer diversity and biomass of these microbes is staggering in comparison to multicellular life. Understanding the guidelines for microbial existence may be key in deciphering key principles behind community engineering, coexistence between highly different individuals, and foundations of energy and nutrient exchange within an ecosystem.

Microbial assemblages, or *communities*, uniquely populate physical pieces of the environment. The small length scale of these organisms—typically on the order of microns—allows for high density of individuals within a population, while high nutrient availability allows for high diversity of species present.

Multiple forces present within natural environments bring pieces of the environment together, providing opportunities for encounters between communities. For instance, gravity works upon tree leaves, which fall upon each other; animals form burrows and casts, providing mixing of the soil, wind stirs dust particulates; natural and industrial liquids provide viscous media for turbulent mixing (Rillig 2015). These processes thus bring two previously established and stable communities in contact with each other. The interchange of microbiota through these physical events may be termed *community coalescence*. The term may also be applied to the process of species diversity extending for generations after the original interchange event.

Coalescence is an interaction between cross-sections of communities at the local scale. Such cross-sections do not necessarily interact with each other; even at local scales, communities do not necessarily exchange species or change community structures in an appreciable manner (Rillig 2015). Spatial restructuring is necessary for the coalescence of two communities together, and to facilitate feedback between species.

Nutrient resources present within the environment of a community provides key metabolic feedback for residing species. Thus, species interact with both the nutrient environment and with the metabolic products of their neighbors. Microbes consume different nutrient sources with varying levels of utility and efficiency. Common sources of carbon include glucose, citrate, and

leucine. Thus, the growth of certain families may be modulated by the limiting carbon source within an environment. A specialist species (for instance, a species that solely feeds upon glucose) may thrive upon an environment composed solely of glucose. However, the same specialist species will almost surely encounter extinction in an environment devoid of this nutrient.

Interactions with other species present in the environment may mitigate the limiting effects of scarce carbon environments. Upon consumption and metabolism of available nutrients, species secrete metabolic by-products into the environment. Importantly, these by-products often constitute a new nutrient source: a novel carbon source differing from that present in the environment and providing opportunities for exploitation by new consumer species. Thus, this *cross-feeding* within scarce environments allows for the persistence of much more diverse populations than previously allowable (Goldford 2018).

Equivalently, introduction of invasive species may lead to antagonistic relations. The presence of multiple mutualistic invasive species may instead facilitate the dominance of relatively few species. This *invasional meltdown* exhibits a negative check upon the positive facilitation of cross-feeding that is possible (Simberloff 1999), and may result in the coalesced community composition more closely resembling either individual assemblage before a coalescence event. Indeed, certain trends in microbial compositions have been repeatedly found after experimental coalescence events with metanogenic microbes (Sieroncinski 2017).

Several factors may affect the process of coalescence. Availability of nutrient sources and metabolic cross-feeding of species is but a small sampling. Other parameters may include: “environmental conditions (entry of communities into new environment versus adding one community to another), the mixing ratios (equal versus unequal community proportions), the interaction interface (communities coalesce via surface touching versus wholesale mixing), or the temporal dynamics of community coalescence events (intermittent pulses versus regular exchange) (Rillig 2015).” Exhaustive consideration of each of these parameters is difficult at this stage of maturity in the field. We instead consider *qualitative trends* relating community- and species-level interactions.

1.2 Building communities: bottom-up and top-down approaches

Consider the following inductive approach to building microbial assemblages from individual species.

Base case: Two species are brought together within a nutrient-limiting environment. The two species may consume the same nutrient resource, or a different one.

If the two species compete for the same resources: a stable population concentration favoring a dominant species is reached. Metabolites are produced and secreted into the environment, where they may be used by members of the same or different species.

If the two species do not compete for the same resources: their abundances will stabilize regardless. Metabolites are similarly produced and secreted.

Inductive step: A new species *invades* the existing community. The new species integrates into the environment as it consumes existing nutrients in the environment and secretes metabolic products for use by others. At no point is a species eliminated from the coalesced community.

Result: The inductive step is repeated for each species present within a coalesced community. Abundances of each species in the community is subject to change with the introduction of each new producer or consumer.

Incorporating cross-feeding into this pairwise inductive model is challenging. Feasibly, microbes significantly alter the chemical content of their existing habitats. The metabolic by-products of one species may influence the survival of more than one species. Each of these species, in turn, generates metabolic by-products that may be included in a *nutrient cycle* involving any combination (or sub-combination) of all species within the community. Small communities of four species have been shown to exist in resource-depleted environments on the principle of cross-feeding alone (Goldford 2018). Such is an example of *higher-order dynamics* at play.

These network-derived dynamics must consider interactions beyond those of simple pairwise competitions. Friedman (2017) explored a similar bottom-up approach by predicting the composition of a coalesced community based on the pairwise interactions between individual species. While this model was valid for small-scale communities (e.g. approximately four or less species present), prediction for a community of eight species was “barely higher than the 61% accuracy obtained when using only the average probability that a species will survive these competitions” (Friedman 2017). Thus, higher-order, holistic network interactions must be considered for any community analysis.

In a network-based inductive approach, the presence of relatively rare species (e.g. at a lesser abundance) nonetheless provides metabolic by-products that may facilitate the recruitment and presence of a new dominant species. Thus, the coalescence of communities requires consideration of the metabolic interactions across multiple species: including the sole *dominant* species, as well as its *cohort* of *sub-dominant* species. Indeed, incorporation of three-species interactions within the pairwise inductive model (Friedman 2017) resulted in higher positive predictive power. We expect this predictive power to increase with consideration of four-species interactions, five-species interactions, etc. The definition and characterization of these multi-species interactions proves to be a difficult question, and one that approximates the complex issue of community assembly itself.

In contrast, consider the following reductive approach to building a microbial assemblage.

Base case: The original coalesced community contains all species present in either the *invasive* or *resident* community.

Reductive step: Competition at each generation results in increase of abundance of some species and decrease in abundance of others. Attrition occurs as consumer species are out-competed for resources and are eliminated from the community, one-by-one (Goldford 2018).

Result: A stable microbial assemblage is predicted, along with the relative abundances of each species.

This top-down approach appears to most accurately capture paradigms at work in the natural environment, in which the number of species present in an environment can only monotonically decrease after a coalescence encounter. However, challenges are endemic to experimentally analyzing and verifying this model. Experimental verification by Goldford (2018) verified stable assemblage of large, coalesced communities into groups of variable compositions with fewer number of species. While this reductive approach may prove successful for predicting *family-level community structures* (e.g. communities dominated by Enterobacteriaceae and Pseudomonadaceae), the approach fails for predictive comparisons at the subfamily (genus) level. Under this paradigm, selection is governed by the dominant and *most metabolically active* members of a certain metabolic family, which exert the most influence upon the environment and the nutrient resources available. A few key *dominant* species drive the community interaction and coalescence results.

At this stage, it remains unclear as to which paradigm of community selection is empirically observed. Is the determination of community coalescence more strongly governed by the collective metabolic presence of several *subdominant* species, or a small number of *dominant* species within each invasive and resident community? Previous work has shown that there may be a many highly able taxa within a community, rather than only a handful of individual species that dominate all community compositions it is present in (Goldford 2018). A tradeoff in metabolic diversity and resource exploration exists between these two possibilities. The valuation of one possibility over the other may describe the outcomes and relative abundances of coalesced communities, from the known relative abundances of species in an invasive and a resident community.

1.3 Predicting communities: challenges

There appear to be fundamental rules governing community assembly, seemingly based upon the metabolic abilities of various species (Goldford 2018). Metabolic capacities and the partitions of usable resources across a community is key in determining the outcomes of community coalescence. Although a community of generalist consumers and a community of specialized consumers could feasibly consume similar metabolic resources, coalescence outcomes may favor communities with decreased niche separations (Castledine 2020).

Higher-order interactions of cross-feeding implicate the necessary consideration of huge numbers of species interactions when predicting community assemblages. These metabolic considerations are nonnegligible, as up to 40% of a community's biomass has been observed to be produced from secreted metabolic by-products (Lu 2018). Among other factors, cross-feeding interactions are necessary for predicting both the presence of and relative abundances of species within a coalesced community.

Further, the histories of previously coalesced communities must be considered when collecting assemblages from the environment for experimental work. Communities that have previously undergone invasion and coalescence to form new native communities may have *pre-selected* for certain taxa to exist together (Sieroncinski 2017). As stated by Lu (2018):

Communities that are not randomly formed and have a previous history of coexistence may exhibit an emergent 'cohesiveness', which produces correlated invasional outcomes amongst species from the same community. For instance, the metabolically more efficient community

might be expected to collectively dominate after community coalescence, since it will overwhelmingly affect the environment and lead to conditions which its members can tolerate, but the members of the other community generally may not.

After considering the interactions of microbes with resources present in their environment, along with metabolic interactions of species with each other, we predict that *the invasion success of a given taxon (family of consumers) is determined by its community members* (Lu 2018).

To study this problem, we distill the question into qualitative comparisons. It is important to distinguish the driving force between higher-order coalescence. Is coalescence between community assemblages determined by a *few key individuals*, or the collective power of interactions among *multiple, less dominant* members of community? Using this framework, the competitive abilities of the individual dominant species within precursor communities must be able to invade in isolation, or with a cohort. Further, its success within a coalesced community must be related to success in direct competition with the opposing dominant species.

1.4 *In silico* analysis: opportunities and pitfalls

In silico analysis denotes experimentation performed on computer, or via computer simulations. Spearheaded by toxicological and pharmacological industry, the push toward experimentation performed with modern computing has been driven by a desire to minimize experimental risk while maximizing the possible scale of analysis (Raunio 2011).

Consideration of multiple factors in modulating community assemblage is difficult. Empirical verification is both time- and resource- intensive. *In silico* experiment and analysis lends itself to high-throughput analysis needed to manipulate these coalescence events in a targeted manner, rather than relying on cosmopolitan or already stabilized communities present *ex situ* (Rillig 2016). Experimental *in vitro* results continue to inspire and improve these coalescence models.

Simulation of community growth and coalescence greatly reduces the man-hours and tangible resources needed to perform and verify the same models. Work by Marsland (2020) yields a Python package for “simulating microbial population dynamics in a reproducible, transparent and scalable way.” Namely, the **community-simulator** provides efficient methods for creating, propagating, and coalescing communities from a diverse range of simulated resources and environments.

However, *in silico* analysis has its limitations. Comparison of results with real, empirical data can be difficult. While the **community-simulator** allows for great flexibility in choosing and extracting species data, the simplified model of bacterial metabolism can limit applicability of results. Similarly, empirical verification of any qualitative trends may prove difficult. Performing these experiments at the intersection of theory and experiment is difficult, due to the wide array of natural and ecological processes that work upon community assemblies in natural ecosystems (Goldford 2018). These environmental influences may be especially pronounced when a novel community invades a resident community and is at an abiotic disadvantage or advantage (Castledine 2020).

Further, *in silico* analysis lends itself to simplification of assumptions. While the natural world is exceeding complex and organisms persist under forces of natural selection, simulation

greatly simplifies forces present in the external world. With this knowledge, we hope to discern the driving forces of community selection that may exist separate from forces present in the external world: selection through the cooperation of multiple rare individuals within a community, or assemblage through attrition by single dominant species.

2. Methods

2.1 Empirical *in vitro* methods

Experimental data has previously been collected in the Sanchez Laboratory at the Microbial Sciences Institute, Yale University West Campus, West Haven CT. Eight environmental sources were sampled with sterile tools to isolate existing microbiota. Such communities were stabilized for 84 generations within carbon nutrient-limited media, containing either glutamine or citrate. These populations displayed stabilized growth and formed cross-feeding metabolic networks (Lu 2018). Microbial communities were plated into minimal media supplemented with 200 μ g/mL cycloheximide and 0.09 C-mol/L glutamine or sodium citrate as carbon sources in 96-well plates. Cultures were maintained at an incubation temperature of 30°C. Identification and propagation of cultures followed standard established protocols, described elsewhere (Goldford 2018).

Upon maintenance of stabilized communities, the identities *dominant* and other, *sub-dominant* species were determined. Pairwise competitions were performed between the dominant species in all eight populations. Mixtures of isolate were propagated for seven serial transfers on their native glucose or citrate environments.

In addition to pairwise, species-level competitions, coalescence was performed between communities. *Invasion* and *resident* communities were selected from the eight stabilized communities. Coalescence was similarly performed by mixing equal volumes of both isolate, then propagating for seven serial growth-dilution cycles. Coalescence was performed in two aerobic environments containing two distinct limiting carbon sources: citrate and glutamine. Frequencies and abundancies of each species within these coalesced communities was determined through 16S illumina sequencing (Lu 2018).

2.2 The community-simulator package

For this analysis, existing Python packages written by Marsland (2020) were used. The package itself provides Python scripts and functions for:

preparing the initial states and environmental conditions for a set of samples, automatic generation of dynamical equations based on a dictionary of modeling assumptions, random parameter sampling with tunable levels of metabolic and taxonomic structure, parallel integration of the dynamical equations, and support for metacommunity dynamics with migration between samples. (Marsland 2020)

Analytic tools utilized from this package included the capabilities of generating `Community` class objects, representing 96-well plates containing a unique community assemblage in each plate. Each `Community` is governed by a series of `assumptions`, which include parameters such as the following:

n_wells: number of wells in the virtual well plate, or the number of unique communities
Sgen: number of generalist species present in each community
response: functional response rate
SA: number of species in each family (consumer type)
MA: number of resources of each of 3 types
S: initial number of species per well

Other mutable parameters include the initial allocation of species and resources within each well plate. Initial allocation is performed with a random number generation providing floating decimal values within a preset range. This random number generation can be predictably changed or set by calling the `np.random.seed()` function on different integer-valued seeds.

The functional response rate describes the functional threshold at which the growth response is initiated. While nonlinear functional response allows for simpler modeling, empirical response rate within biological systems is most likely nonlinear. *In vivo* systems should demonstrate a low response to stimulus until a functional threshold is reached, after which response remains constant or the organism is desensitized to the stimulus. The **type I**, **type II**, and **type III** parameters model three such possible response rates: linear, exponential, and logistic models, respectively.

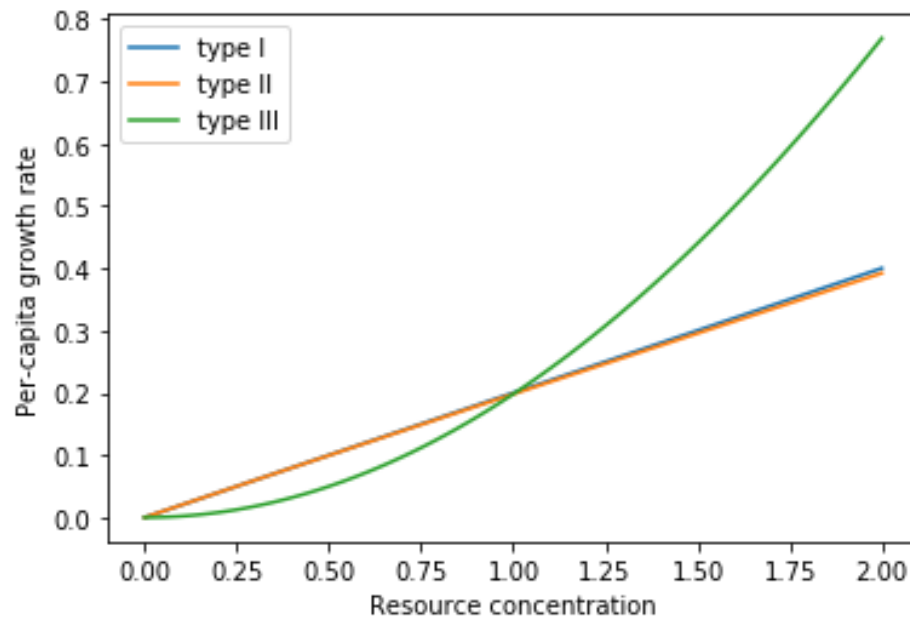


Figure 2.2.1: Three forms of functional response rates modeled within the `community-simulator` package. Units on axes are relative, with arbitrary units.

Once established, a **Community** may be allowed to **Propagate** for time, specified in hours. Parameters of growth time are tuned to resemble growth rates of real microbial populations. Communities within each well can also undergo a **Passage**, in which the composition of each well plate is diluted by a certain `transfer_fraction` and copied to a new plate. The newly **Passaged** plate replaces the precursor plate within the **Community** object. Used in conjunction, the **Propagate** and **Passage** functions approximate the growth-dilution cycles used in real experimental protocols.

2.3 *In silico*: community preparation

Several new functions and scripts were written in Jupyter Notebook (Python 3), using the `community-simulator` package (Marsland 2020). Pseudocode and implemented functions are provided in Appendix A.

Initial instantiation of a `regional pool`, or all possible species, was generated twice for each experiment: once for *invasive* communities and once for *resident* communities. 210 species were present within this regional pool: 60 of each 3 types of specialist consumer (modeled by accepting a certain type of resource), and 30 generalist consumers.

For each community, 210 values were sampled from a power-law distribution of probability density function:

$$ax^{a-1}$$

where $a=0.8$ is the power-law constant and x is on the range $[0,1]$. Values were then normalized to obtain relative species abundance for each of 210 total species that could be present. Values below a threshold of 10^{-4} were zeroed, effectively removing the species from the community pool.

Each *invasive* and *resident* Community was represented by the instantiation of a simulated 96-well plate. Each well contains a combination of species from the `regional pool`, also generated through a power-law distribution. These two communities then independently followed protocols for ten 24-hour growth cycles of `Propagate/Passage` to reach stable population equilibrium. Equilibrium within these communities was determined by plotting the biomass and species diversity within each well. Both metrics reached asymptotic equilibrium well before 10 cycles of `Propagate/Passage` were attained. (Verification is provided in section 3.2.)

These communities were generated for a series of mutable parameters, including response type and random seeding. `Response` was drawn from one of three possible types:

- type I: a simple linear response curve with no upper limits; lower limit set to 0
- type II: exponential response curve with an asymptotic upper limit k
- type III: logistic response curve with asymptotic upper limit k

Stochastic variability between communities with different response type was controlled between trials with identical random seeds. Adequate variance between trials was achieved by initializing species abundances with different random seeds.

2.4 *In silico*: community coalescence

For each pair of generated *invasive* and *resident* communities, a third *coalesced* community was created by joining the two communities together. Relative abundances of species within the *invasive* and *resident* communities were added, then normalized, to create a new 96-well plate. These communities then underwent cycling of `Propagate/Passage` identical to that described above.

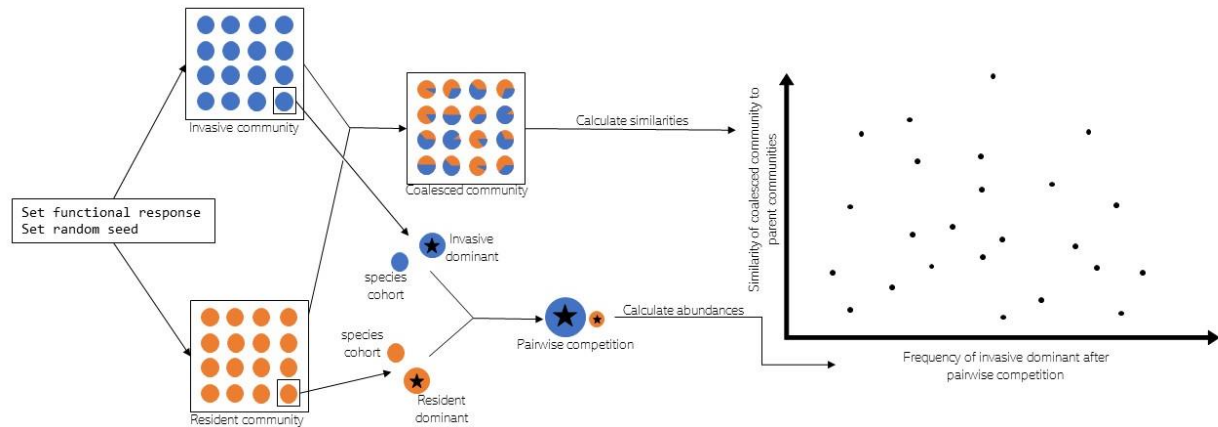


Figure 2.4.1: Recapitulation of *in silico* methods. Each schematized well plate contains 96 communities. Each community is propagated to stabilization before further analysis. Each point within the final scatterplot represents the outcome of one of 96 communities, or wells, in an experiment.

As population equilibrium was attained for these *coalesced* communities, species diversity and biomass were similarly tracked and plotted. Similarity between these *coalesced* communities and their precursor *invasive* and *resident* communities was quantified through three metrics: Jaccard index, Bray-Curtis index, and an “overlap” metric. Definitions of these metrics are found in Appendix B.

Within the previously generated *invasive* and *resident* communities, the dominant species was determined by identifying the species with the highest absolute abundance within the well. Using methods analogous to community coalescence simulation, pairwise competition between these two species was performed using novel Python functions (Chang 2019). Propagation of these pairwise competitions was also performed for ten generations at 24-hour cycles.

The relative abundance of the dominant species from the *invasive* community was calculated as a proportion over the total biomass from the dominant *invasive* and dominant *resident* species. As a result, each of the 96 coalesced communities were matched to this relative abundance based on the identity of the dominant *invasive* and dominant *resident* species prior to coalescence.

2.5 *In silico*: quantifying success

Upon identification of dominant species within both the *invasive* and *resident* communities, several coalescence/competition experiments were performed. These included: coalescence of two communities with each other; competition of two individual species against each other; and invasion of a single dominant species within a population. Within each of these results, the “success” of the dominant species within the invasive community was calculated by quantifying its relative abundance within the final species pool.

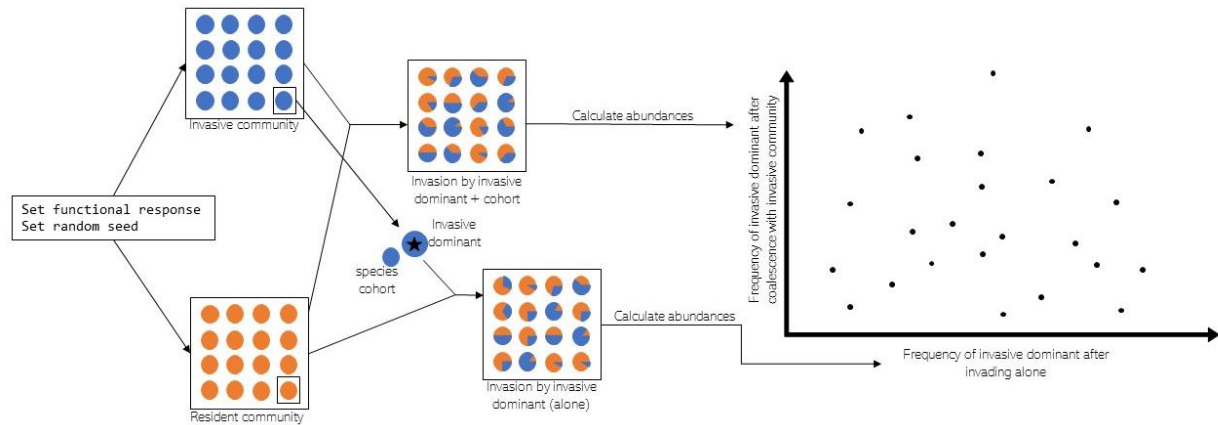


Figure 2.5.1: Comparing the relative success of dominant invasive species after invading the resident community with a cohort and invading the resident community alone. Each point within the final scatterplot represents the outcome of one of 96 communities, or wells, in an experiment.

3. Analysis and Results

3.1 Empirical *in vitro* results

In this study, we examined whether communities whether dominant species in two coalesced communities are directly placed in competitive pressure against one another, or whether their presence is primarily attributed to the presence of metabolic effects of a cohort of sub-dominant species. This aimed to describe the difference between positive reductive (“top-down”) selection, and positive inductive (“bottom-up”) selection.

Similarities between *invasive*, *resident*, and *coalesced* communities were quantified through various metrics (see Appendix B for descriptions). In quantifying invasion success, it is intuitive to compare the similarity of the resident (e.g. invaded) community) with the final, coalesced community. Relative abundances of the *invasive dominant* and *resident dominant* species were also determined for the final coalesced communities.

To further characterize the extent to which dominant competition governs community coalescence, we also determined the relative invasion success of the endemic invasive species (i.e. those that are present in the invasive but not the resident community). We found that the relative frequency of a dominant against another in head-to-head pairwise competition is only weakly predictive of the relative success of its endemic community members during community coalescence (linear regression, adjusted $R^2=0.23$, $P<0.001$). (Lu 2018)

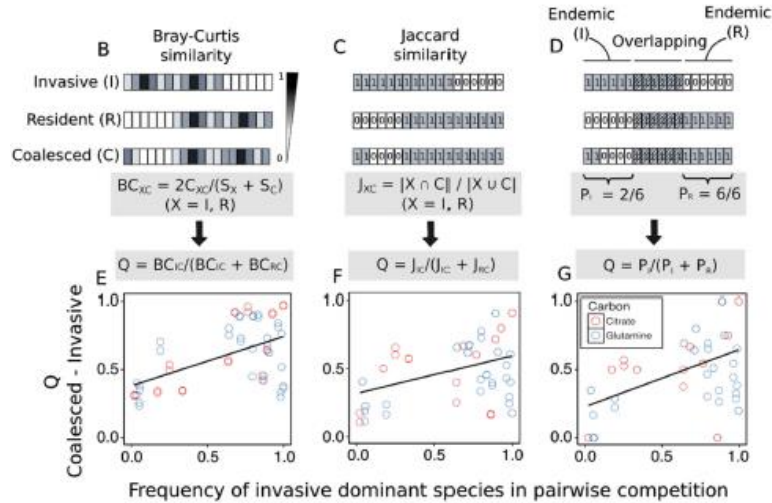


Figure 3.1.1: Experimental results of community coalescence and comparisons with relative frequencies of invasive dominant and resident dominant species. Qualitatively positive regression lines indicate that competitive abilities of dominant species are an indicative driving force in community coalescence. Figure adapted from Lu (2018).

The results of pairwise dominant competition also exhibited typical hierarchical structure. However, *transitivity* of hierarchy was violated, presumably due to metabolic effects of other species present within communities. Thus, the presence and metabolic activity of sub-dominant species is directly responsible for the relative abundance of dominant species within a community.

These results were further confirmed by comparing the relative success of individual species after invasion with a cohort of species, or invasion performed alone.

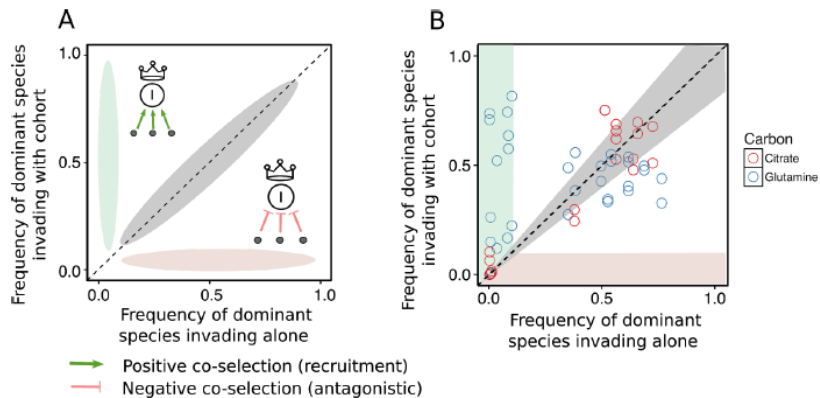


Figure 3.1.2: Experimental results of individual species abundance after invading with a cohort and after invading alone. The green region demarcates a regime where dominant species reach higher abundance within the final community only after invading with a cohort. The gray region demarcates a regime where invasive ability is unaffected by the presence of a cohort. Figure adapted from Lu (2018).

While we may conclude that the presence of sub-dominant species influences the success and behavior of a dominant species, there are other instances where the dominant species is unaffected by the presence of a cohort. Additionally, the presence of a cohort is strictly

beneficial: there are no regimes where the presence of a cohort negatively influences the success of the dominant species. This scenario is somewhat unlikely, as the corresponding precursor communities would exist in an unstable equilibrium and is subject to collapse.

3.2 *In silico* analysis

Using the `community-simulator` package, communities were simulated and propagated to reach population equilibrium. Biomass within each community was plotted during each Propagate/Passage cycling. Initial community masses should be below equilibrium value, due to random sampling and omission of species from communities. As each transfer is performed, species present should grow in abundance and stabilize within a growth period.

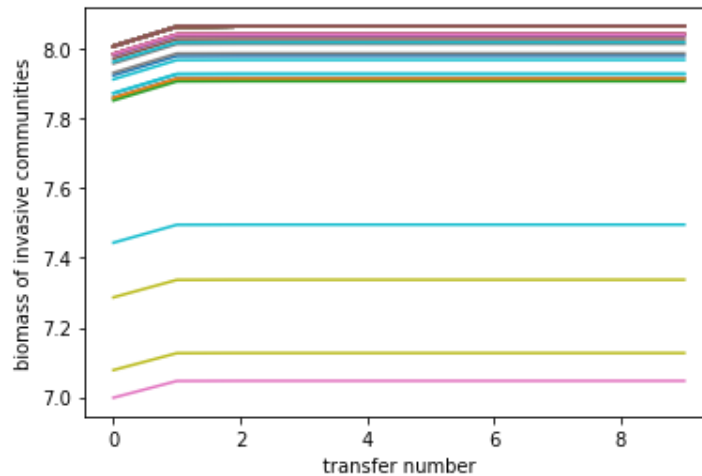


Figure 3.2.1: Sample plot of biomass present within a 96-well plate of invasive communities. Asymptotic growth at values between 7.0 and 8.4 indicate that the species present have reached a growth limit imposed by, and equal to, the quantity of usable resources available to the specialists and generalists present.

Additionally, the number of species within each community was plotted during Propagate/Passage cycling. Total species diversity should decrease as competitors between species are eliminated. Simply put: the number of species present must remain constant or decrease, but cannot increase. Equilibrium is attained when the number of species remains constant for each set of usable resources. For *in silico* approximation, we estimate this time period as the earliest time after the period of steepest decline in species diversity.

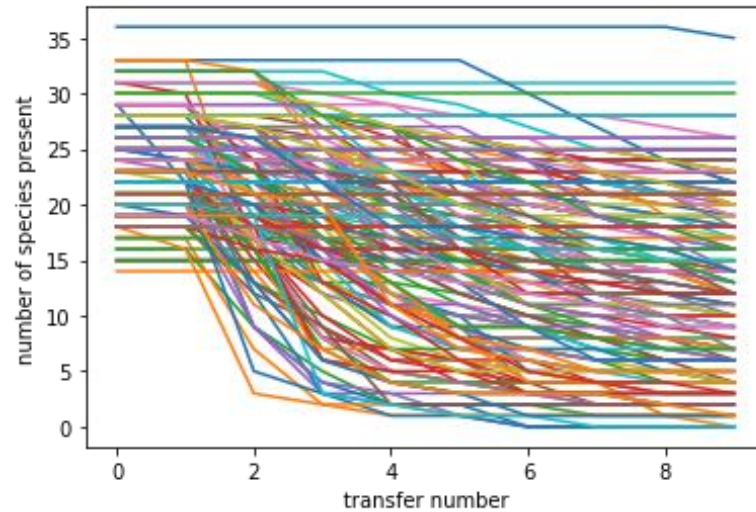


Figure 3.2.2: Control plot of species number within a pair of *invasive* and *resident* communities. Species diversity is asymptotic or near-asymptotic by the tenth passage.

Coalesced communities were compared to their precursor *invasive* and *resident* communities using a three metrics: Jaccard, Bray-Curtis (BC), and one quantifying the relative “overlap” between the union of species. (Definitions are available in Appendix B.) The results of pairwise competition between dominant species of these same precursor communities were quantified through the final relative abundance of the dominant *invasive* species. Linear regressions between these two values were performed for each set of 96 coalesced communities that resulted from each of 27 96-well plates holding *invasive* and *resident* communities. The slope and R^2 value for these linear regressions was calculated. The Pearson correlation coefficient was also calculated between these two quantities.

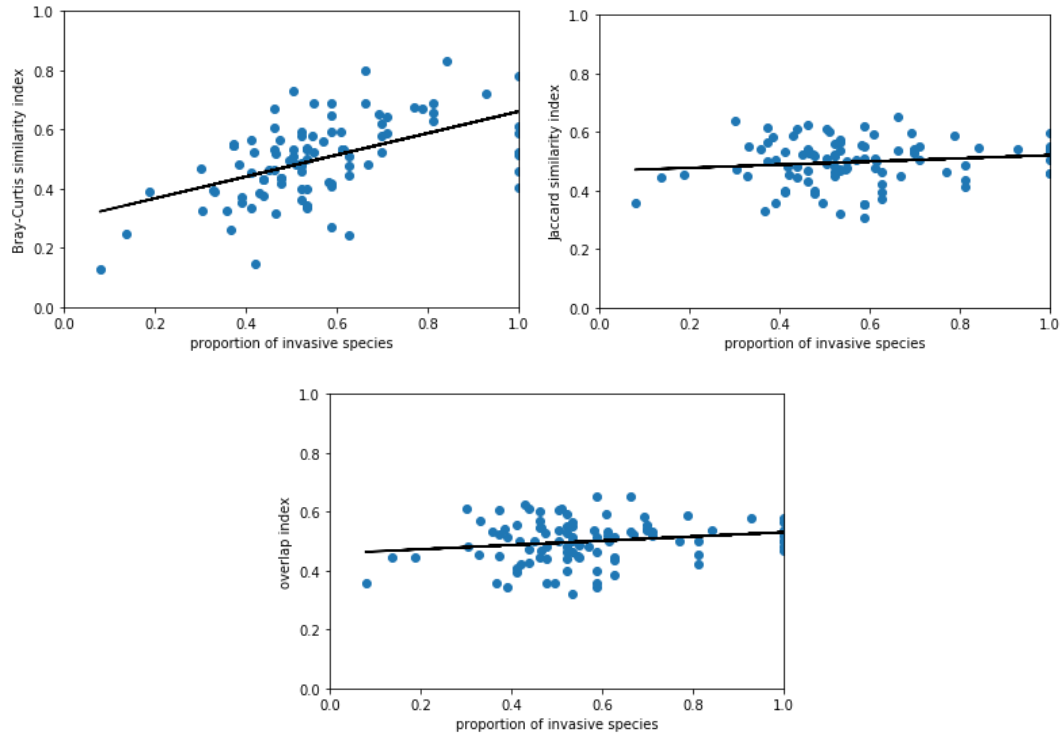


Figure 3.2.3: Sample plots and linear regressions between similarity indices and relative abundance of pairwise competitions. Sample community was generated with Type I response curve.

The final slopes and R^2 value resulting from linear regressions, and Pearson correlation coefficient, were further analyzed by functional response type (e.g. type I, type II, or type 3).

3.3 Implications for results of pairwise species competition

Twenty-seven experiments with **Community** objects were instantiated, each with unique functional response rates and random seeds. Complete *in silico* analysis was performed on each of these simulated experiments. Population equilibrium was attained in each of 27 simulated experiments, where resident and invasive communities were represented in 96-well plates. Coalescence and pairwise competition were also successfully achieved in their respective simulations. Wall clock runtime for each of these experiments was approximately 45 minutes.

Linear regressions for each plot of community coalescence similarity and abundance of the invasive species after pairwise competition were generally positive.

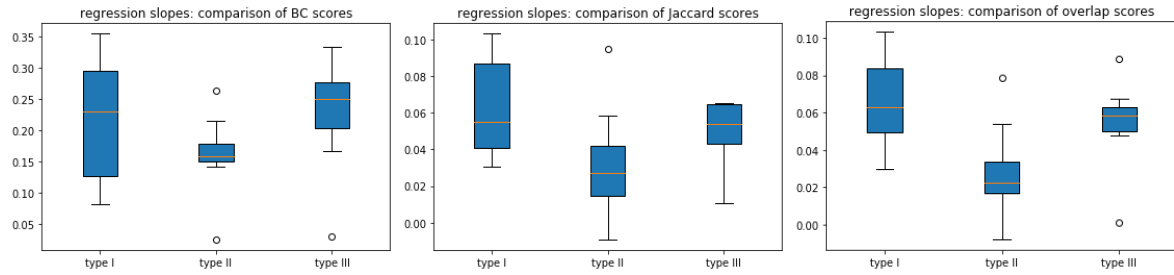


Figure 3.3.1: Box-and-whisker plots of slopes from linear regression. Nine data points are included per box-and-whisker plot of type I, type II, or type III functional response.

Two negative values were generated in the regression slopes, both the result of experiments conducted under type II functional response. A one-sample t-test was performed to test the statistical difference of these values from zero. Each sample was significantly different from zero at the 0.05 significance level.

Thus, there is a possible positive relationship between the relative abundance of the resident dominant species in pairwise competition, and the similarity of the coalesced community to the precursor communities. This was further verified using the Pearson correlation coefficient between relative abundance of the resident dominant species in pairwise competition and the similarity metric.

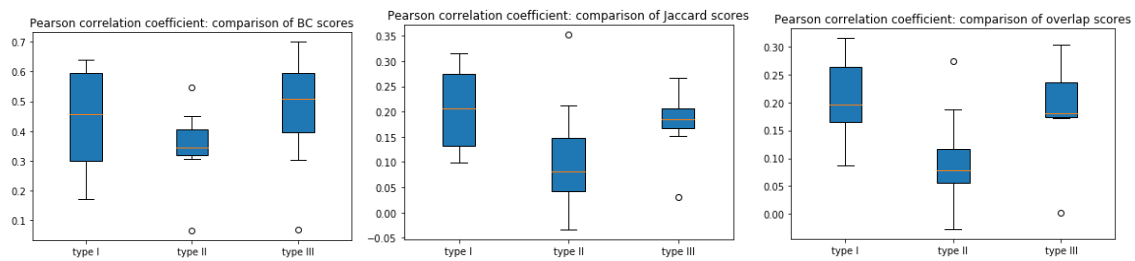


Figure 3.3.2: Box-and-whisker plots of the Pearson correlation coefficient between each of 27 datasets. Pearson correlation is calculated for 96 points within a dataset.

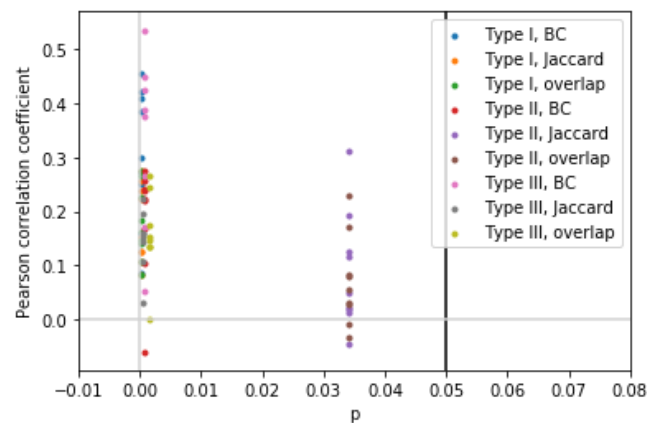


Figure 3.3.3: Pearson correlation coefficient between relative abundance of the resident dominant species in pairwise competition and the similarity metric, plotted against the p-values of a one-sample t-test of regression slopes tested against the value zero. The black line denotes $p=0.05$ significance level.

There is a positive relationship between these two values: the relative abundance of a dominant resident species after pairwise competition, and the similarity between a final coalesced community and its precursor communities. This statement holds true over all methods of quantifying similarity, and all three functional response rates tested.

This agrees with results from *in vitro* experimentation: namely, that the results of a community coalescence can be predicted from trends in pairwise species competition (Lu 2018). Heuristically, the similarity between a coalesced community and its precursor species is directly related to the success of pairwise competition between the two dominant species of the parent populations.

3.4 Implications for invasion by a single species

Invasion by a single invasive species was compared against successful invasions performed by a cohort. The *in silico* results reveal qualitatively similar final abundances of the invasive dominant species, suggesting that individual dominant species tend to perform equally well regardless of whether a cohort is present or not.

These results seem to support the presence of a few key dominant species within the ecosystem of microbial communities. These key dominant species do not require the metabolic contribution of other species within a community, as they may be able to fulfill the same metabolic niches equally well on their own.

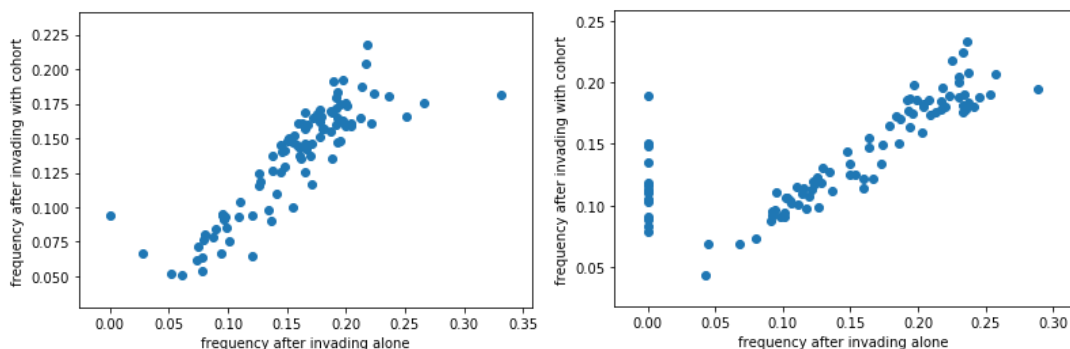


Figure 3.4.1: Abundance of a single species within the invasive community, after invasion as a member of an invasive community or after invasion invading alone. Abundances are calculated after reaching population equilibrium at 10 transfers. Solid line indicates the identity line.

left: Sample plot taken from populations grown under Type I (logistic) functional response.
right: Sample plot taken from populations grown under Type III (logistic) functional response.

Two regimes seem to be present: dominant invasive species that invade equally well regardless of the presence of their cohort, and dominant invasive species that require the presence of a cohort to perform well. From *in silico* simulations, species within the former category seem to be more prevalent. Relatively few species seem to exist within the category; these individual species are almost always unable to reach an appreciable frequency without their cohort, but perform well in the presence of their community.

Findings from *in vitro* experimentation seem to agree. While some dominant species seem to do well regardless of the presence of a cohort, it is more likely that species tend to positively co-

recruit other species during an invasion. This may be due to the increased likelihood of individual species to fail during *in vitro* invasion, or a higher diversity of species in the pool that does not include key dominants.

4. Discussion

While this study yields promising results in establishing tradeoffs between inductive and reductive models of community selection, further work is required to establish the full landscape of forces at work during coalescence. High-throughput work *in silico* provides an opportunity to precisely control initial species abundances and resource availabilities. Further work should consider carbon-limiting resources in the environment, and the combinatorial nature of cross-feeding.

Our preliminary results suggest that the results of two-community coalescence can be directly predicted by the results of individual pairwise competition. Further, the cohort of sub-dominant species within a stable community plays a large role within the stabilization of a coalesced community. Theoretical *in silico* work of community coalescence in removing the metabolic effects of a dominant species may be possible, while somehow preserving cross-feeding contributions of sub-dominant species. This experimentation should carefully consider the “cohesiveness” of these sub-dominant species both before and after coalescence.

Consideration of community-level competition between assemblages of mutually competing species may also be valuable. A quantitative approach to this may refer to the optimization of *characteristic functions* of various community assemblages (Tikhonov 2016).

Additional divergent trends of individual pairwise competitions also provide opportunities for the study of dominance within a community. Survivorship of a dominant species within a community may be modulated one, a few, or many sub-dominant species. Our empirical results indicate that invasion by a dominant species—performed in solitude—is often successful, despite the absence of a cohort. Characterization of these dependencies may shed light on a *critical size* of a cohort to successfully reside within a sterile environment, along with the necessary metabolic functions. Work performed by Sieroncinski (2017) with anaerobic communities has shown similarities in combinations of multiple dominant species both before and after coalescence, demonstrating that “certain combinations of taxa within a community might be co-selected as a result of coevolved interactions.”

Other parameters are also available for tuning in community selection simulations. Further work may increase the complexity of future studies “to also capture mixing ratios and modes, among other parameters (Rillig 2016).” *In silico* results also provide flexible controls for “pH changes, diauxic shifts, or the fact that glucose and citrate communities are more similar to each other than they are to those stabilized in leucine” (Goldford 2018).

5. References

- Castledine, M., Sierocinski, P., Padfield, D., Buckling, A. (2020). "Community coalescence: an eco-evolutionary perspective." Philosophical Transactions of the Royal Society B **375**(20190252).
- Chang, C. Y. (2019). Python package for community selection. .
- Friedman, J., Higgins, L.M., Gore, J. (2017). "Community structure follows simply assembly rules in microbial microcosms." Nature: Ecology & Evolution **1**.
- Goldford, J. E., Lu, N., Bajic, D., Estrela, S., Tikhonov, M., Sanchez-Gorostiaga, A., Segre, D., Mehta, P., Sanchez, A. (2018). "Emergent simplicity in microbial community assembly." Science **361**: 469-474.
- Lu, N., Sanchez-Gorostiaga, A., Tikhonov, M., Sanchez, A. (2018). "Cohesiveness in microbial community coalescence." bioRxiv preprint.
- Marsland, R., Cui, W., Goldford, H., Mehta, P (2020). "The Community Simulator: A Python package for microbial ecology." PLoS ONE **15**(e0230430).
- Raunio, H. (2011). "*In silico* toxicology--non-testing methods." Frontiers in Pharmacology **2**.
- Rillig, M. C., Antonovics, J., Caruso, T., Lehmann, A., Powell, J.R., Veresoglou, S.D., Verbruggen, E. (2015). "Interchange of entire communities: microbial community coalescence." Trends in Ecology & Evolution **30**.
- Rillig, M. C., Tsang, A., Roy, J. (2016). "Microbial community coalescence for microbiome engineering." Frontiers in Microbiology **7**.
- Sierocinski, P., Milferstedt, K., Bayer, F., Großkopf, T., Alston, M., Baskowski, S., Swarbreck, D., Hobbs, P.J., Soyer, O.S., Hamelin, J., Buckling, A. (2017). "A single community dominates structure and function of a mixture of multiple methanogenic communities." Current Biology **27**: 3390-3395.
- Simberloff, D., Von Holle, B. (1999). "Positive interactions of nonindigenous species: invasional meltdown?" Biological Invasions **1**: 21-32.
- Tikhonov, M. (2016). "Community-level coalescence without cooperation." eLife **5**.

Appendix A: Pseudocode and Python analysis

```
assumptions['n_wells'] = 96
assumptions['response'] = response_type
np.random.seed(seed)

invasive_plate, invasive_community_function = generate_plate(assumptions, dynamics)
np.random.seed(2**seed)
resident_plate, resident_community_function = generate_plate(assumptions, dynamics, G)
for G times:
    invasive_plate.Propagate(), invasive_plate.Passage()
    resident_plate.Propagate(), resident_plate.Passage()

coalescence_plate = invasive_plate + resident_plate
for G times:
    coalescence_plate.Propagate(), coalescence_plate.Passage()

invasive_dominants = get_dominant_species(invasive_plate)
resident_dominants = get_dominant_species(resident_plate)
pairwise_plate = make_synthetic_pairwise_community(invasive_dominants, resident_dominants)
for G times:
    pairwise_plate.Propagate(), pairwise_plate.Passage()
get_invasive_dominant_frequencies(pairwise_plate)

calculate_metrics(invasive_plate, resident_plate, coalescence_plate)

box_whisker_plots(calculated_metrics)
```

Appendix B: Metrics for similarity

B.1 Bray-Curtis index

$$BC_{X,C} = 1 - \frac{2C_{X,C}}{S_X + S_C}$$

$C_{X,C}$: the sum of minimum species abundancies for those species present in both communities

S_X, S_C : the sum of occurrences of species present in communities X and C , respectively

For the shown plots we use the complement of this value:

$$BC_{X,C} = \frac{2C_{X,C}}{S_X + S_C}$$

B.2 Jaccard index

$$J_{X,C} = \frac{|X \cap C|}{|X \cup C|}$$

$|X \cap C|$: the number of species common to both communities X and C

$|X \cup C|$: the number of species present in *either* community X and C

B.3 An “overlap” metric

$$O_{X,C} = \frac{||X \cup C| \setminus C|}{||X \cup C| \setminus C| + ||X \cup C| \setminus X|}$$

$||X \cup C| \setminus C|$: the number of species present in community C , but not in X

$||X \cup C| \setminus X|$: the number of species present in community X , but not in C