THE UNIVERSITY OF CHICAGO

INVESTIGATING MICROBIAL COMMUNITY ECOLOGY USING SYNTHETIC BACTERIAL COMMUNITIES

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON MICROBIOLOGY

BY

KEVEN DOOLEY

CHICAGO, ILLINOIS

JUNE 2023

Table of Contents

List of Figures

List of Tables

Abstract

The functions of microbial communities are crucial to sustaining life on Earth and have a pervasive impact on human society. Thus, there is great interest in developing a nuanced understanding of the processes that control microbial communities, with the hope that we can apply such knowledge to manipulate and potentially engineer microbial communities to carry out specific functions. However, a complex set of interacting biotic and abiotic factors influence the structure and function of microbial communities. The rise of 'omics technologies has offered excellent tools to obtain observational data from microbial communities and generate hypotheses regarding the ecology of these systems. However, experimental approaches are required to evaluate such hypotheses and ultimately unravel the complexity of microbial communities. Artificially constructed communities of bacteria, referred to as "synthetic communities", offer a powerful approach with which we can investigate ecological hypotheses in a controlled environment.

Here, I present work using synthetic bacterial communities to study interspecific interactions, coexistence, and ecological invasion. In the first chapter, I evaluated the assumption that the interaction between two members of a community is unaffected by the surrounding community context and found that changes in community richness and density were strong predictors of how interaction effects varied across contexts. In the second chapter, I decomposed a set of bacterial isolates into all pairwise and *n-2* communities to compare coexistence between these "bottom-up" and "top-down" contexts and found that pairwise observations of coexistence and exclusion were useful but incomplete predictors of the composition of complex assemblages. In the third chapter, I investigated how the timing of an ecological invasion affected the success of the invader and the impact on the resident community. I found evidence that the effect of timing on invasion outcome was associated with changes in resource use efficiency over the course of community assembly.

Introduction

The biochemical processes performed by microbial communities are vital to life on our planet and impact human society in countless ways. Our various human-associated microbial communities affect our health in many ways by performing vital nutritional, immunological, and even developmental functions (Lynch and Pederson, 2016). Soil and plant microbiomes play an important role in our supporting the crops that sustain our society (Trivedi et al., 2020). Microbial communities even help us process waste (Narihiro and Sekiguchi, 2007) and remediate polluted landscapes (Brune and Bayer, 2012). An appreciation of the widespread importance of microbial communities is relatively new to the field of microbiology and has been ushered in by the rapid advances in 'omics technologies. However, as our ability to characterize the structure (composition) and function of microbial communities has rapidly advanced, our understanding of the forces that shape these complex communities has not matched pace (Widder et al., 2016; Prosser 2020). There is great interest, however, in developing an advanced understanding of the forces that govern microbial communities, as many hope that such an understanding will enable us to modify, and even design, microbial communities to perform a desired function (Brenner et al., 2008; De Souza et al., 2020).

To do so will require us to understand the ecology of these complex systems. A diverse set of forces jointly determine microbial community structure and function. Like all ecological communities, environmental conditions have great impact on microbial communities. Abiotic factors like temperature (Biller et al., 2015) and pH (Lauber et al., 2009) are primary determinants of microbial community structure. Stochastic factors such as migration and disturbance are also relevant to microbial communities (Symons and Arnott, 2014; Fukami, 2015). Another important determinant of microbial community structure and function is interaction between community members, referred to as "interspecific interactions". Indeed, microbes are well known for their

metabolic interactions, especially "cross-feeding", a form a metabolic exchange which has received great attention in studies of gut microbiomes.

Ecologists have long been interested in interspecific interactions and have attempted to measure them as pairwise relationships between members of a community. Georgy Gause provides a famous and early example of such efforts, which were aptly performed in pairwise cocultures of two species of Paramecium, a genus of aquatic ciliate, or two species of yeast (1934). In such studies, he demonstrated that although the species of each pair could grow independently, when grown in coculture, one species would outcompete the other, and in the case of the Paramecia, lead to exclusion from the community. Such work represents one of the first efforts to quantitatively measure ecological interactions and were done so as an empirical evaluation of the theoretical work of Alfred Lotka (1920) and Vito Volterra (1926). Decades beyond the work of these pioneering scientists, our understanding of ecological interactions has progressed, but our basic framework for conceptualizing and estimating such interactions exists on the foundations they established.

Today, microbial interactions are estimated through a variety of approaches. High throughput amplicon and metagenomic sequencing opened the gates for efforts to catalogue the complexity of natural microbial communities, and following this flood of genetic data came the development of computational approaches which could use such data to infer microbial interactions. A very common approach is to infer ecological interactions from correlations in relative abundances across cross-sectional samples of microbial communities (Faust and Raes, 2012; Carr et al., 2019). However, this approach has drawn scrutiny for the limitations of using compositional data to perform such analyses (Weiss et al., 2016; Gloor et al., 2017; Röttjers and Faust, 2018). A more complex approach fits longitudinal data to a model of community dynamics that considers interspecific interactions, often the generalized Lotka-Volterra (gLV) model (Fisher and Mehta,

2014; Bucci et al., 2016; Clark et al., 2021). This approach, however, requires the additional estimation of absolute abundances (via colony counting, optical density, qPCR, flow cytometry, etc.) and dense timeseries, both of which can be challenging to collect. A final approach is simply the continuation and expansion of the empirical approach employed by Gause and others (e.g., Vandermeer, 1969). Namely, working with synthetic microbial communities assembled (and disassembled) in the laboratory. Synthetic communities are simpler than most natural microbial communities but have become increasingly popular as experimentally tractable systems for evaluating specific ecological hypotheses (Widder et al., 2016). Despite the differences between these computational and empirical approaches, all have been used to investigate how knowledge of interactions can be leveraged to predict the structure and function of microbial communities.

A multitude of approaches, ranging in complexity and capacity, exist for using interactions to predict features of microbial communities. An approach that has received much recent attention is an extension of the gLV approach described above, where the interactions inferred by fitting data to a gLV model of dynamics are used to predict the outcome of unobserved communities (Fisher and Mehta, 2014; Venturelli et al., 2018; Rao et al., 2021). A notable example of this approach included a simple functional model alongside the gLV model of community dynamics to accurately predict the target function (butyrate production) of unobserved communities (Clark et al., 2021). Although powerful, this approach is complex and data intensive, requiring multiple rounds of a design-test-learn cycle. A lower complexity, less data intensive approach was recently developed which offers a method for inferring interactions and predicting the composition of unobserved ecological communities without requiring assumptions of an underlying model of community dynamics (Maynard et al., 2020). This approach is exciting because it offers an experimentally efficient approach for estimating interactions in communities of moderate complexity, which may prove especially useful for working with synthetic microbial communities. A final approach worth

discussing is a so called "assembly rule" of microbial communities, which states multispecies assemblages will only coexist if all species can coexist in pairwise (Friedman et al., 2017). This naïve approach is limited in application but well studied and surprisingly useful for predicting coexistence in microbial communities (Meroz et al., 2021; Ortiz et al., 2021; Lax and Gore, 2022). The approaches described here, and the many others that exist, demonstrate that we can apply knowledge of interactions to predict features of unobserved microbial communities. However, as described earlier in this introduction, a multitude of forces shape the structure and function of microbial communities, and if we wish to better understand the assembly of these communities, we must grapple with those forces as well.

One such ecological force pertinent to many microbial systems is migration/invasion. Indeed, given the ubiquitousness and scale of microbial communities, commonplace events like a rain shower, the turning over of soil, or eating a meal can be thought of as largescale migration events. Almost any designed microbial community would thus need to be robust to migration and, conversely, efforts to change the composition of communities would be facilitated by understanding this process. The study of ecological invasions in macroecological systems has often focused on the relationships between community diversity, productivity, and invasibility. An early postulate of the field proposed that more diverse communities should be more robust to invasion (Elton 1958), with a possible mechanism being the increased productivity of diverse communities. In other words, more diverse communities are more productive because they are better able to use the available environmental resources and are thus more robust to invasion because they deny an invader the resources it requires (Tilman, 1999). These relationships have been investigated in the context of microbial communities as well. In fact, microbial communities represent an excellent system in which to study invasion because one can feasibly manipulate and study diverse communities in high replication over short time frames – which cannot be said for studies working in macroecological

systems, such as plant communities. As such, studies working with experimental microbial communities have identified several factors relevant to invasion of these systems, many of which align with studies in macroecological systems. Many studies have demonstrated that increased community diversity and productivity are associated with decreased invasibility of microbial systems, as in some macroecological systems (Hodgson et al., 2002; Van Elsas et al., 2012; Jones et al., 2021). Other factors like "propagule pressure" (i.e., invader density) and biotic interactions have also been identified as relevant to community invasibility (Acosta et al., 2015; Jones et al, 2017; Albright et al., 2020). We do not yet have a full understanding of the ecological invasion of microbial communities, but it is apparent that experimental microbial communities are an effective tool with which we can interrogate the determinants of invasibility.

In this dissertation, I present three separate studies in which I used synthetic bacterial communities to investigate several of the concepts briefly reviewed above. The three studies can be broadly summarized as focusing on interspecific interactions, coexistence, and ecological invasion, respectively. All three, however, are united in the experimental approach they take, namely, the use of synthetic bacterial communities. This approach offers a high level of experimental control and scalability, both factors vital to the work I sought to conduct.

In my first chapter, I evaluated a common simplifying assumption made when studying interspecific interactions in microbial communities; namely, that the relationship between two members of a community is unaffected by the surrounding community context. To investigate this assumption, I compared interactions across a set of synthetic bacterial communities. I found that interactions were generally weak and negative in effect, with interactions between bacteria of the same genus more negative than those between different genera. Community richness and density were strong predictors of interaction effects, which generally decreased as richness increased.

Despite this, the modest changes in interaction strength across similar community contexts could serve as predictors for those same interactions in different contexts, as long as they were not substantially different. This work suggests that by better understanding emergent patterns in microbial interactions, we can better predict the structure and function of microbial communities.

In my second chapter, I decomposed a previously observed synthetic community into all pairwise and *n-2* communities to compare coexistence between these "bottom-up" and "top-down" contexts. In doing so, I sought to evaluate the importance of emergent effects in determining the composition of simple microbial communities. I found that pairwise coexistence frequently did not persist in the top-down contexts, but this could generally be explained by observations of pairwise exclusion with other relevant community members. I also found that most isolates that did not coexist in pairwise also failed to coexist in more complex assemblages. Further, instances of context-dependent coexistence were largely attributable to a small set of isolates which frequently displayed such unexpected coexistence. Finally, I observed that there was a strong positive correlation between prevalence rank in pairwise and complex communities. These results demonstrate that pairwise observations of coexistence and exclusion can make accurate, but incomplete, predictions of coexistence in complex assemblages, a finding with useful implications for designing simple microbial communities.

And in my third chapter, I investigated how the timing of an ecological invasion affects the success of that invasion and the impact on the resident community. To investigate this, I invaded synthetic bacterial communities at three different stages of community assembly: the initial assembly of the community, 24 hours into community dynamics, and 7 days into community dynamics. I found that the timing of the invasion had a significant impact on its success and effect on the community, with early invasions being most successful and having the largest impact. This result was

statistically associated with the growth of the invaders on spent media. Further, invader success was positively associated with lower community richness and resource use efficiency. These results suggested that the effect of invasion timing was largely due to changes in resource use efficiency over the course of community assembly.

References

Acosta, Francisco, et al. "Dynamics of an experimental microbial invasion." *Proceedings of the National Academy of Sciences* 112.37 (2015): 11594-11599.

Albright, Michaeline BN, et al. "Biotic interactions are more important than propagule pressure in microbial community invasions." *Mbio* 11.5 (2020): e02089-20.

Biller, Steven J., et al. "Prochlorococcus: the structure and function of collective diversity." *Nature Reviews Microbiology* 13.1 (2015): 13-27.

Brenner, Katie, Lingchong You, and Frances H. Arnold. "Engineering microbial consortia: a new frontier in synthetic biology." *Trends in biotechnology* 26.9 (2008): 483-489.

Brune, Karl D., and Travis S. Bayer. "Engineering microbial consortia to enhance biomining and bioremediation." *Frontiers in microbiology* 3 (2012): 203.

Bucci, Vanni, et al. "MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses." *Genome biology* 17 (2016): 1-17.

Carr, Alex, et al. "Use and abuse of correlation analyses in microbial ecology." *The ISME journal* 13.11 (2019): 2647-2655.

Clark, Ryan L., et al. "Design of synthetic human gut microbiome assembly and butyrate production." *Nature communications* 12.1 (2021): 3254.

De Souza, Rafael Soares Correa, Jaderson Silveira Leite Armanhi, and Paulo Arruda. "From microbiome to traits: designing synthetic microbial communities for improved crop resiliency." *Frontiers in Plant Science* 11 (2020): 1179.

Elton, C. S. "The ecology of invasions by animals and plants." (Methuen, London, 1958).

Faust, Karoline, and Jeroen Raes. "Microbial interactions: from networks to models." *Nature Reviews Microbiology* 10.8 (2012): 538-550.

Fisher, Charles K., and Pankaj Mehta. "Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression." *PloS one* 9.7 (2014): e102451.

Gause, G. F. "The Struggle for Existence." (Williams & Wilkins, 1934).

Gloor, Gregory B., et al. "Microbiome datasets are compositional: and this is not optional." *Frontiers in microbiology* 8 (2017): 2224.

Hodgson, David J., Paul B. Rainey, and Angus Buckling. "Mechanisms linking diversity, productivity and invasibility in experimental bacterial communities." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269.1506 (2002): 2277-2283.

Jones, Matt L., et al. "Biotic resistance shapes the influence of propagule pressure on invasion success in bacterial communities." (2017): 1743-1749.

Jones, Matt Lloyd, et al. "Relationships between community composition, productivity and invasion resistance in semi-natural bacterial microcosms." *Elife* 10 (2021): e71811.

Lauber, Christian L., et al. "Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale." *Applied and environmental microbiology* 75.15 (2009): 5111-5120.

Lax, Simon, and Jeff Gore. "Strong Ethanol-and Frequency-Dependent Ecological Interactions in a Community of Wine-Fermenting Yeasts." *bioRxiv* (2022): 2022-09.

Lotka, Alfred J. "Analytical note on certain rhythmic relations in organic systems." *Proceedings of the National Academy of Sciences* 6.7 (1920): 410-415.

Lynch, Susan V., and Oluf Pedersen. "The human intestinal microbiome in health and disease." *New England Journal of Medicine* 375.24 (2016): 2369-2379.

Maynard, Daniel S., Zachary R. Miller, and Stefano Allesina. "Predicting coexistence in experimental ecological communities." *Nature ecology & evolution* 4.1 (2020): 91-100.

Meroz, Nittay, et al. "Community composition of microbial microcosms follows simple assembly rules at evolutionary timescales." *Nature communications* 12.1 (2021): 2891.

Narihiro, Takashi, and Yuji Sekiguchi. "Microbial communities in anaerobic digestion processes for waste and wastewater treatment: a microbiological update." *Current opinion in biotechnology* 18.3 (2007): 273-278.

Ortiz, Anthony, et al. "Interspecies bacterial competition regulates community assembly in the C. elegans intestine." *The ISME Journal* 15.7 (2021): 2131-2145.

Prosser, James I. "Putting science back into microbial ecology: a question of approach." *Philosophical Transactions of the Royal Society B* 375.1798 (2020): 20190240.

Rao, Chitong, et al. "Multi-kingdom ecological drivers of microbiota assembly in preterm infants." *Nature* 591.7851 (2021): 633-638.

Röttjers, Lisa, and Karoline Faust. "From hairballs to hypotheses–biological insights from microbial networks." *FEMS microbiology reviews* 42.6 (2018): 761-780.

Trivedi, Pankaj, et al. "Plant–microbiome interactions: from community assembly to plant health." *Nature reviews microbiology* 18.11 (2020): 607-621.

Vandermeer, John H. "The competitive structure of communities: an experimental approach with protozoa." *Ecology* 50.3 (1969): 362-371.

Van Elsas, Jan Dirk, et al. "Microbial diversity determines the invasion of soil by a bacterial pathogen." *Proceedings of the National Academy of Sciences* 109.4 (2012): 1159-1164.

Venturelli, Ophelia S., et al. "Deciphering microbial interactions in synthetic human gut microbiome communities." *Molecular systems biology* 14.6 (2018): e8157.

Volterra, Vito. "Fluctuations in the abundance of a species considered mathematically." *Nature* 118.2972 (1926): 558-560.

Weiss, Sophie, et al. "Correlation detection strategies in microbial data sets vary widely in sensitivity and precision." *The ISME journal* 10.7 (2016): 1669-1681.

Widder, Stefanie, et al. "Challenges in microbial ecology: building predictive understanding of community function and dynamics." *The ISME journal* 10.11 (2016): 2557-2568.

Chapter 1: Richness and density jointly determine context dependence in bacterial interactions

1.1 Abstract

The structure and function of a microbial community is in part determined by the network of interactions among its constituent members. To simplify the complexity of these networks, it is often assumed that an interaction between two community members is unaffected by the surrounding community context. Here, we investigate that assumption and search for patterns describing variation in the effect of pairwise interactions across contexts. To do so, we used shallow short-read sequencing to compare a set of interactions across several synthetic bacterial community contexts. We found that the full set of interactions was largely composed of weak interactions with a negative average effect. We also observed a phylogenetic effect, with interactions between bacteria of the same genus more strongly negative than inter-genus interactions. Community richness and total density emerged as strong predictors of interaction effects and contributed to a general attenuation of interaction effects as richness increased. This attenuation was observed in both population level and per-capita measures of interaction effects, suggesting factors beyond systematic changes in population size were involved; namely, changes to the interactions themselves. Despite the effect of community context in attenuating interaction strengths, the modest change in the strength of pairwise interactions across similar community contexts meant that they could serve as predictors of those same interactions across community contexts, provided the contexts were not substantially diverged. These results suggest we can advance our ability to predict the structure and function of microbial communities through a better understanding of the emergent properties of microbial interactions.

1.2 Introduction

Microbes are the engines of many biochemical processes that support life on Earth (Falkowski et al., 2008), and as such may serve as a tool to modify or design biological systems (Brenner et al., 2008; Liu et al., 2019). Importantly, however, microbes rarely perform these complex functions in isolation, instead acting within communities. Many efforts are thus underway to design microbial communities that perform desired functions, enabling us to co-opt these powers of chemical transformation and develop applications relevant to human health, agriculture, and industry. However, the intricate relationships underlying such complex functions provide a challenge that must be overcome, as interactions among members constrain the extent to which the abundance and distribution of a focal microbe can be manipulated. Overcoming this challenge will require understanding the forces that determine the structure and function of microbial communities.

Interactions between community members have long been known to affect community composition (Diamond, 1975; Gotelli and McCabe, 2002; Horner-Devine et al., 2007) and therefore the emergent functions performed by a community (Foster and Bell, 2012; Yu et al., 2019). Leveraging an understanding of interspecific interactions is a promising and actively researched approach for designing the structure and function of microbial communities (Clark et al., 2021; Connors et al., 2022). However, for such an approach to be effective, observations of interactions made in one community context must inform the extent of that interaction in another context.

Interactions are often modelled as a network of unchanging pairwise per-capita or proportional effects between members of a community (Wootton and Emmerson, 2005; Levine et al., 2017; Momeni et al., 2017). By assuming that it is appropriate to distill an interaction into a simple static relationship, we can reduce the complexity of interaction networks (Gonze et al., 2018), a practical necessity for diverse microbial communities, and apply knowledge of interactions gleaned from

other contexts to make predictions about unobserved communities (Maynard et al, 2020). However, a variety of known effects call this simplification into question. Interactions can be subject to higher order effects ("higher order interactions" or "HOIs") where a pairwise interaction is altered by the presence of one or more other community members (Wootton, 1993; Billick and Case, 1994; Bairey et al., 2016). Habitat modification can also affect microbial interactions (McNally and Brown, 2015), an example being environmental pH modification, which has been observed as a relevant factor in microbial community assembly (Amor et al., 2020; Aranda-Díaz et al., 2020; Ratzke et al., 2020). Due to effects such as these, knowledge of pairwise interaction strength or coexistence can have limited predictive power in complex communities (Friedman et al. 2017; Chang et al., 2022). Thus, advancing our understanding of what contributes to the variation of interactions between contexts stands to facilitate the rational design of microbial communities.

One potential solution to these complexities is to identify patterns in how pairwise interactions vary across contexts and uncover what gives rise to such patterns. Applying such an understanding stands to improve our predictions of how microbial interactions will change between community contexts. Encouragingly, recent work has demonstrated that stronger negative interactions are found at high nutrient concentrations (Ratzke et al., 2020), confirming the possibility of identifying broadly general patterns. By expanding our understanding of such patterns, we hope to improve the predictive power of pairwise interactions. Here, we use synthetic bacterial communities to observe how interactions vary across community contexts and identify patterns underlying that variation.

1.3 General Methods

Assembly of synthetic communities and measurement of interactions – We began by assembling a set of synthetic communities from a pool of 56 bacterial strains isolated from the leaves of wild and field-

grown *Arabidopsis thaliana* by randomly dividing isolates into seven pools of eight members. We then created all combinations of those pools (i.e., seven 8-member groups, all twenty-one 16-member groups, etc.) for a total of 127 unique communities (figure 1.1a). These communities were inoculated into a custom growth-medium derived from *A. thaliana* leaves (*Arabidopsis* leaf medium, ALM) (methods) at a consistent total community titer, with each member accounting for an equal proportion of the population given the initial richness (number of community members). To allow the communities to reach a steady state reflective of the long-term composition, we passaged each community for 6 days by performing a 1:100 dilution into fresh medium every 24 hours (figure 1.1b). This period was sufficiently long to allow the community composition to stabilize (supplementary figure 1.1). We characterized the compositions of these communities by mapping Illumina short reads against a nearly complete and high-quality genome assembled for each isolate (methods).

From this set of 127 communities, we then identified putative interactions by finding pairs of communities where a focal isolate was observed to be both excluded by and capable of coexisting alongside one or more specific isolates (figure 1.1c, supplementary figure 1.2). In these scenarios, we posited that the observed context-dependent coexistence was related to interactions between the focal isolate and/or its context-dependent excluder with additional members of the community. We thus focused on communities featuring this phenomenon. From all paired communities in which we observed context-dependent coexistence, we selected a set of ten pairs that would maximize compositional diversity.

To probe these potential interactions, we then decomposed these communities into assemblages varying by a single isolate (figure 1.1c), always including a focal isolate and/or its excluder. Our goal was to assemble a set of nested subcommunities with compositions varying by a single member.

With these sets of communities, we were able to measure a given interaction across multiple community contexts (figure 1.2a). In total, we assembled 245 communities and passaged them for 6 days, as previously described. A subset of communities was passaged for 12 days, with samples from days 6 and 12 sequenced to confirm that community composition was stable by day 6 (supplementary figure 1.3).

Absolute abundance information is required to measure interactions. Thus, we performed colony counting from serial dilutions of each day-6 sample to measure absolute community density, which was partitioned by the relative abundance sequencing data to estimate the absolute density of each isolate in a community.

With measures of absolute abundance in hand, we measured interactions by comparing abundances between pairs of communities that varied by a single member. For example, the interaction between a "focal" isolate A and an "interactor" isolate B was observed by comparing the abundance of A in a community lacking B to the abundance of A in a community where B was present (A'). Here, we measure interactions using two metrics (figure 1.2a, methods), and refer to the "strength" of an interaction as the absolute value of the interaction effect. Our first metric measures an interaction as the ratio of the abundance of the focal isolate in contexts with and without the interactor. This is a commonly used metric (Baichmann-Kass et al., 2022; Hsu et al., 2019; Kehe et al., 2021; Schafer et al., 22; Weiss et al., 2022), which represents an interaction as a population level effect on the focal isolate. Our second metric measures an interaction as the per-capita effect of an interactor B on the abundance of a focal A (Gonze et al., 2018). Conceptually, the population level effect of an interaction is a function of the per-capita effect of an interactor scaled by the density of that interactor in a given community context. This is relevant, as below we will

show that a general relationship between richness and density existed in our communities and contributed to the observed effect of interactions.

1.4 Results

Negative interactions were more common and stronger than positive interactions – We observed a total of 388 pairwise interactions across all community contexts (figure 1.2). Negative interactions were more common, representing 67% of population level interactions. We observed median values of -0.29 and -0.22, for population level and per-capita effects, respectively (figure 1.2). Negative interactions were statistically stronger than positive interactions for both measures (one-sided Wilcoxon rank sum test: p-values $<4e^{-8}$ and 0.002, respectively). We also observed phylogenetic effects for both population level and per capita level measures of interactions, where isolates belonging to the same genus tended to have a stronger negative interaction than those belonging to distinct genera (one-sided Wilcoxon rank sum test: p-value 0.004 and 0.017, respectively).

Individual interactions attenuated as richness increased – Our method of measuring interactions compares contexts that differ in richness by a single "interactor" isolate. To observe how interactions change between contexts, we extended this framework to compare the strength of a pairwise interaction between focal and interactor isolates across community contexts with or without a single "background" isolate (figure 1.2a). In this way, we compared how an interaction varied between two community contexts differing in richness by one. Use of our complete dataset enabled us to analyze 275 instances of such paired contexts. First focusing on the population level effects, we observed that interactions generally attenuated in strength when measured in a community with one additional background member (Wilcoxon signed rank test, p-value 0.003). When grouping interactions by their initial direction, the median positive and negative interaction became less positive and negative,

respectively (figure 1.3a). Interaction strength, however, was significantly weaker for initially negative interactions and significantly stronger for initially positive interactions (one-sided Wilcoxon rank sum test: p-values $<1e^{-5}$ and 0.007, respectively).

As previously stated, the population level effect of an interaction is a function of the per-capita effect of an interactor and the density of that interactor in each community context. Thus, the observed decrease in population level effects suggests a decrease in the strength of per-capita effects and/or a systematic decrease in interactor density. Indeed, when we consider the shift in per-capita interactions, we observe again that interaction effects attenuated as richness context increased (Wilcoxon signed rank test, p-value 0.003). As with the population level effects, per-capita effects grouped by initial direction showed consistent shifts (figure 1.3b). Interaction strength became significantly weaker for initially negative interactions; however, unlike at the population level, it remained statistically unchanged for positive interactions (one-sided Wilcoxon rank sum test: p-values 0.002 and 0.64, respectively). These results suggest that part of the decrease in the population level effects can be explained by a decrease in the per-capita effects. Next, we evaluate an alternative explanation by investigating the relationships between richness and density in our communities.

Relationships between richness and density help explain trends in population effects across richness –We observed that, as richness increased, average total density of communities gradually increased to a modest extent (figure 1.4a), while the average density of each member decreased before reaching an asymptote of about four (figure 1.4b). This relationship between richness and the density of community members was experienced consistently across individual isolates (supplementary figure 1.4). This general decrease in density with increasing richness helps explain the observed attenuation of interactions when measured as a population effect. Namely, as individual densities decreased with an increase in community richness, the population level effect of the interactor should decrease.

Indeed, interactor density explained a significant portion of variance in population level effects (linear regression, adjusted R^2: 0.09, p-value $<5e^{-10}$). Additionally, the relationship between the densities of community members and total community density meant that as richness increased, the absolute change in total density associated with an interaction decreased (figure 1.4c, Pearson's r: -0.23, p-value 0.002). In other words, on average, adding a given interactor to a community resulted in a smaller change to total community density in higher richness contexts.

Interactions are predictive between contexts – We next asked, what is the remaining predictive capacity of an observed interaction across community contexts? We attempted to answer this question by modelling the change in abundance of a focal isolate between community contexts (i.e., the effect of an interaction), informed by richness, total density, and interactions observed in different contexts.

First, while modelling the change in abundance associated with all 388 observed interactions, both richness and the change in total density emerged as highly explanatory variables associated with the change in focal isolate abundance (table 1.1). Richness context was less explanatory than the change in total density (adjusted R^2 of 0.08 and 0.33, respectively), and joint models including both variables and their interaction could explain 57% of variance in focal isolate change in abundance.

Next, we used the set of 275 paired interaction contexts differing by a richness of one to interrogate the predictive power of interactions across contexts (e.g., compare interaction effects between richness contexts 1=>2 and 2=>3). In this dataset, interactions in one context were able to describe ~16% of the variance in the change in density of a focal isolate (i.e., interaction effect) in the other context (table 1.2). A model using the change in total density of the context for which the interaction effect was being predicted (akin to the models above) explained ~27% of variance. A joint model of these two variables (change in total abundance of the predicted context and interaction effect in another context) explained ~42%, suggesting that the two variables are largely

independent. Expanding the dataset to consider comparisons between any observations of a given interaction (e.g., compare richness contexts 1=>2 and 4=>5) reduced the explanatory power of interactions between contexts to ~10% of variance, suggesting the predictive power of interactions decays as communities diverge in richness and composition (supplementary table 1.2). Ultimately, these results demonstrate the persistent but limited predictive power of interactions across contexts and highlight the relevance of community level properties in understanding the assembly of microbial communities.

1.5 Discussion

Here, we used synthetic bacterial communities to observe a large set of interactions across community contexts, ranging from the simplest possible community of two coexisting isolates to complex communities with up to seven isolates. These interactions were on average weakly negative and displayed a phylogenetic effect, in alignment with other studies of microbial interactions (Russel et al., 2017; Schafer and Vorholt., 2019; Kehe et al., 2021). However, positive interactions were not uncommon, an observation that has growing recent empirical support (Kehe et al., 2021; Baichmann-Kass et al., 2022). When comparing interactions across contexts of increasing richness, we observed a general attenuation of interactions, though this arose predominately due to a consistent shift in negative interactions (figure 1.3). Interestingly, we observed that much of this change can be explained by relationships between individual density and richness/total density (table 1.1, figure 1.4). Namely, as richness increased, the modest increase in total density resulted in a decrease in individual isolate density (figure 1.4b). This relationship can help explain the observed attenuation of population level effects, as decreased density of interactor isolates in higher richness contexts should lead to smaller effects and did, in fact, explain ~9% of variance in population level

effects. However, the per-capita effects also showed some decrease in strength with an increase in richness, at least for negative interactions, suggesting additional processes were present that imparted a systematic change on the interactions.

Why would per capita effects be attenuated at high richness, and why predominately among initially negative interactions? Previous observation of the attenuation of pairwise interactions in the zebrafish gut was attributed to the effect of higher order interactions (Sundarraman et al., 2020), though that study was unable to identify the mechanisms of such effects. We have a similarly limited mechanistic understanding of observed interactions and what underpins their variation between community contexts. The importance of HOIs in microbial community assembly remains an actively debated subject, with theoretical and empirical evidence to support both sides (Vandermeer, 1969; Foster and Bell, 2012; Bairey et al., 2016; Grilli et al., 2017; Venturelli et al., 2018; Mickalide and Kuehn, 2019; Sanchez-Gorostiaga et al., 2019). However, HOIs are challenging to appropriately identify (Case and Bender, 1981; Wootton, 1993 & 1994), and our lack of fully characterized interaction networks precludes us from determining their relevance here.

Another possible explanation for the attenuation of per-capita effects is non-additivity in interactions. In other words, overlap in the mechanisms underpinning how multiple interactors affect a given focal isolate could result in a reduced per-capita effect when multiple interactors are present. Such non-additivity has been recently reported (Baichmann-Kass et al., 2022). This effect would be likely if metabolic interaction (such as competition over labile carbon sources) predominately underlies interactions and community assembly, as has been shown in synthetic communities that were organized into functional guilds by preferred metabolic strategy (Goldford et al., 2018; Estrela et al., 2021; Estrela et al., 2022). In the context of how we measure and compare interactions here, such mechanistic overlap would be hypothesized to reduce the impact of a novel

interactor due to a function already being performed by a "background" isolate in the community. Such mechanistic redundance would be probabilistically more likely as richness increases.

Despite the limitations of our data, some interesting insights can be inferred by asking what gives rise to the relationships we observed between richness and individual or total density. The apparent modest increase in total density in higher richness communities might have emerged for two reasons: 1) larger initial pools of isolates entailed greater metabolic diversity, thus allowing the community to occupy more of the available niche space, or 2) larger initial pools may simply have had a greater chance of including one or more isolates with high fitness in the environment. Both possibilities would result in higher levels of community metabolic activity at higher levels of richness, which has been observed to have a positive effect on those community members with relatively low fitness as a result of cross-feeding or general metabolic leakiness (Medlock et al., 2018; Kehe et al., 2019; Kehe et al., 2021). In this way, positive effects absent in simpler contexts may have emerged in more complex settings. This hypothesis would address the fact that we predominately observed attenuation among negative interactions, as it would result in an apparent decrease in per-capita effect while actually representing an independent emergent positive effect.

A key finding here was that the relationship between individual isolate density and richness/total community density was informative for predicting the change in abundance of an isolate between community contexts (table 1.1). But why were changes in total density informative of changes in individual density? We suggest that this result arose because individual isolate density decreased as richness increased (figure 1.4b) due to the modest changes in total density (figure 1.4a). The associated attenuation of interactions in higher richness contexts was inherently observed as a decreased change in individual density but also a decreased change in total community density (figure 1.4c). This link between the two effects meant that the change in total density was an

informative predictor of change in individual density (i.e., interaction effect) as well. Nonetheless, interaction effects themselves were useful predictors across contexts (table 1.2), suggesting that context-dependency generally does not redefine an interaction, but instead changes interactions to varying degrees. Indeed, we observed the explanatory power of interactions decayed as the divergence between community contexts increased (supplementary table 1.2). Such an outcome is in line with results from other studies, as it has been shown that predictions of coexistence based on pairwise cultures decay as complexity of the predicted community increases (Friedman et al., 2017).

We sought to advance our understanding of microbial interactions by observing how they vary across contexts and identifying patterns in that variation. Our observation of the general attenuation of interactions as richness increased is a straightforward and potentially useful result. And our finding that the relationships between individual density, richness and total density could help explain changes in pairwise interactions demonstrates both the usefulness of understanding community level properties and the value of considering interactions from the per-capita perspective. The observation that negative per-capita interactions nonetheless generally attenuated with richness suggests that context-dependency of interactions is a common feature in microbial communities. Further study of the specific processes that give rise to such context-dependence would be a fruitful endeavor that, combined with the observed population level processes, may improve our ability to predict the structure and function of microbial communities.

1.6 Detailed Methods

Bacterial isolates and reference genomes – All bacterial isolates were originally isolated from the leaves of wild or field grown *Arabidopsis thaliana* in the midwestern states of the USA (IL, IN, MI). Reference genomes for the isolates used in these experiments were assembled from Illumina short

reads using Spades v3.13.0 (Bankevich et al., 2012) with the "careful" flag. Assembled genomes were then manually curated in the Anvi'o v6.2 (Eren et al., 2021) software platform, specifically using the interactive interface to remove outlier contigs assembled from contaminating sequences. Anvi'o was also used to estimate the completion and contamination of assembled genomes and assign taxonomy. The isolate names, taxonomy, and assembly information are presented in Supplementary Table 1.1.

Arabidopsis leaf medium (ALM) – *Arabidopsis thaliana* (KBS-Mac-74, accession 1741) plants were grown in the University of Chicago greenhouse from January to March 2020. Seeds were densely planted in 15-cell planting trays and thinned after germination to 4-5 plants per cell. Above ground plant material was harvested just before development of inflorescence stems. Plant material was coarsely shredded by hand before adding 100g to 400mL of 10mM $MgSO_4$ and autoclaving for 55 minutes. After cooling to room temperature, the medium was filtered through 0.2μm polyethersulfone membrane filters to maintain sterility and remove plant material. The medium was stored in the dark at 4°C. Before being used for culturing, the medium was diluted 1:10 in 10mM $MgSO_4$.

Experimental set up and culturing – Fresh bacterial stocks were prepared by first inoculating the isolates into 1mL of ALM shaking at 28°C and growing overnight. 100uL of these cultures were then used to inoculate 5mL of ALM shaking at 28°C. Once the cultures were visibly turbid, they were divided into 1mL aliquots with sterile DMSO added to a final concentration of 7% as a cryoprotectant. Stocks were stored at -80°C. Additionally at this time, stocks were diluted and plated to quantify density through colony counting.

To initiate an experiment, stocks were diluted to target densities determined by the initial community titer (~$1x10^6$ cells) and the number of initial members. For the preliminary synthetic

communities, isolates were first combined into 7-member pools, subsequently combined into all 127 combinations of pools, and then distributed into three randomly selected wells containing 600μL of ALM in sterile 1mL deep-well plates. Similarly, for the synthetic communities used to measure interactions, isolates were first combined into desired initial community compositions and then randomly distributed in triplicate into 1mL deep-well plates. All such manipulations were performed under an open atmosphere with a Tecan Freedom Evo liquid handling robot. Deep-well plates were covered with sterilized, loosely fitting plastic lids to allow air exchange. Plates were cultured in the dark at 28°C on high-speed orbital shakers capable of establishing a vortex in the deep-well plates to ensure that the cultures were well-mixed. After 24 hours, 6μL of each culture was manually transferred by multi-channel pipette into new plates containing 594μL of fresh ALM. The new plates were immediately returned to the incubator and the day-old plates were stored at -80°C. The sample plates from the final time point (day 6) were amended with 15% glycerol prior to storage in the freezer to preserve the cultures for subsequent colony counting.

DNA Extraction – DNA was extracted from synthetic communities using an enzymatic digestion and bead-based purification. Cell lysis began by adding 250μL of lysozyme buffer (TE + 100mM NaCl + 1.4U/μL lysozyme) to 300μL of thawed sample and incubating at room temperature for 30 minutes. Next, 200μL of proteinase K buffer (TE + 100mM NaCl + 2% SDS + 1mg/mL proteinase K) was added. This solution was incubated at 55°C for 4 hours and mixed by inversion every 30 minutes. After extraction, the samples were cooled to room temperature before adding 220μL of 5M NaCl to precipitate the SDS. The samples were then centrifuged at 3000 RCF for 5 minutes to pellet the SDS. A Tecan Freedom Evo liquid handler was used to remove 600μL of supernatant. The liquid handler was then used to isolate and purify the DNA using SPRI beads prepared as previously described (Rohland & Reich, 2012). Briefly, samples were incubated with 200μL of SPRI beads for 5 minutes before separation on a magnetic plate, followed by two washes

of freshly prepared 70% ethanol. Samples were then resuspended in 50μL ultrapure H2O, incubated for 5 minutes, separated on a magnetic plate, and supernatant was transferred to a clean PCR plate. Purified DNA was quantified using a Picogreen assay (ThermoFisher) and diluted to 0.5ng/μL with the aid of a liquid handler.

Sequencing library preparation – Libraries were prepared using Illumina Nextera XT kits and following a custom, scaled down protocol. This protocol differed from the published protocol in two ways: 1) the tagmentation reaction was scaled down such that 1μL of purified DNA, diluted to 0.5ng/μL, was added to a solution of 1uL buffer + 0.5μL tagmentase, and 2) a KAPA HiFi PCR kit (Roche) was used to perform the amplification in place of the reagents included in the Nextera XT kit. PCR mastermix (per reaction) was composed of: 3μL 5X buffer, 0.45μL 10mM dNTPs, 1.5μL i5/i7 index adapters, respectively, 0.3μL polymerase, and 5.75μL ultrapure H2O. The PCR protocol was performed as follows: 3 minutes at 72 °; 13 cycles of 95 °C for 10 seconds, 55 °C for 30 seconds, 72 °C for 30 seconds; 5 minutes at 72 °C; hold at 10 °C. Sequencing libraries were manually purified by adding 15μL of SPRI beads and following the previously described approach, eluting into 12μL of ultrapure H2O. Libraries were quantified by Picogreen assay, and a subset of libraries were run on an Agilent 4200 TapeStation system to confirm that the fragment size distributions were of acceptable quality. The libraries were then diluted to a normalized concentration with the aid of a liquid handler and pooled. The pooled libraries were concentrated on a vacuum concentrator prior to size selection for a 300-600bp range on a Blue Pippin (Sage Science). The distribution of size-selected fragments was measured by TapeStation. Size-selected pool libraries were quantified by Picogreen assay and qPCR (KAPA Library Quantification Kit).

Sequencing – Leveraging our previously assembled genomes, we characterized the compositions of our synthetic communities with a shallow metagenomics approach. We chose this approach as

opposed to 16S amplicon sequencing as some of our isolates had identical 16S sequences and preliminary work with mock synthetic communities demonstrated that amplicon sequencing yielded less accurate characterizations of community composition. Initial synthetic community samples were sequenced on a HiSeq 4000 platform while follow up synthetic community samples were sequenced on a NovaSeq 6000 platform. Reads were quality filtered and adapter/phiX sequences were removed using BBDuk from the BBTools suite. Reads were mapped to reference genomes using Seal (BBTools) twice, once with the "ambig" flag set to "toss" (where ambiguously mapped reads were left out) and once with the "ambig" flag set to "random" (where ambiguously mapped reads were randomly distributed to equally likely references). By comparing the results between these two strategies, we identified sets of reference genomes which resulted in high numbers of ambiguous reads (due to similarity) and corrected for such ambiguity by reallocating "tossed" reads based on proportions of unambiguous reads mapped in each sample containing a given set. To avoid mischaracterizing the composition of our synthetic communities due to contamination or non-specific mapping, for a given sample, isolates with less than 1% of total mapped reads were ignored.

Estimating absolute abundance – Absolute density of each community culture was measured by counting colonies from serial dilutions of the cultures (specifics below). Specifically, glycerol preserved final timepoint samples were plated on 1X tryptic soy agar (TSA) plates, in triplicate serial dilution (3e-5, 1e-6, and 3e-6 dilutions), and cultured at room temperature. Colony forming units (CFU) were counted by eye over the course of the next few days.

Calculation of interactions – Population level effects were calculated as the ratio of focal isolate density with and without the interactor and presented as (ratio - 1) for ease of interpretation. The per-capita effects were calculated as the change in focal isolate density between contexts with and without the interactor, divided by the density of the interactor from the context in which it was

present. To remove spurious interactions that arise from the presence of low abundance isolates close to the 1% relative abundance threshold, we pruned interactions to only include those within one standard deviation from the mean for both the population level and per-capita effect measures.

Statistical analysis and data visualization – Statistical analysis and figure generation was performed in R v4.0.2 with aid from the following packages: tidyverse (Wickham et al., 2019), reshape2 (Wickham, 2007), and car (Fox and Weisberg, 2018). Linear regression was performed with the "lm" function in R. All scripts are provided in the supplementary materials. Figures were made using Microsoft Paint 3D and PowerPoint.

1.7 Figures and Tables

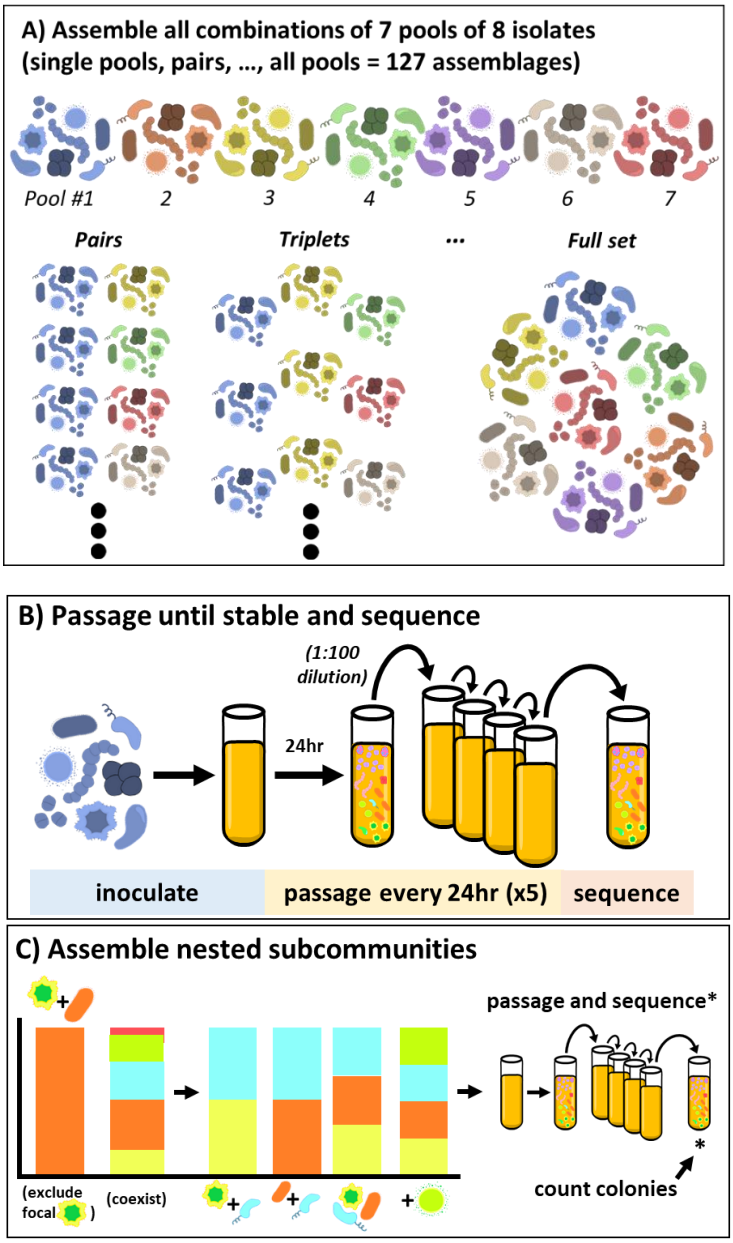


Figure 1.1: Experimental outline – A) A set of 56 isolates representing 21 genera were randomly pooled into 7 pools. All combinations of those pools were assembled at equal titers, with respective densities scaled to the total number of isolates initially present. B) These combinations were inoculated in triplicate into a custom medium derived from *Arabidopsis* leaves (ALM) and passaged daily into fresh medium at a 1:100 dilution for 5 days. To characterize the community compositions, the day-6 samples were sequenced, and short-reads were mapped to reference genomes. C) Ten communities displaying context-dependent coexistence were decomposed into nested subcommunities containing the focal isolate and/or putative excluder isolate. These communities were assembled, passaged, and sequenced as described for the previous communities. To provide the absolute abundance information necessary to measure interactions, the final timepoint (day-6) was quantified by counting colonies on 1X TSA plates.

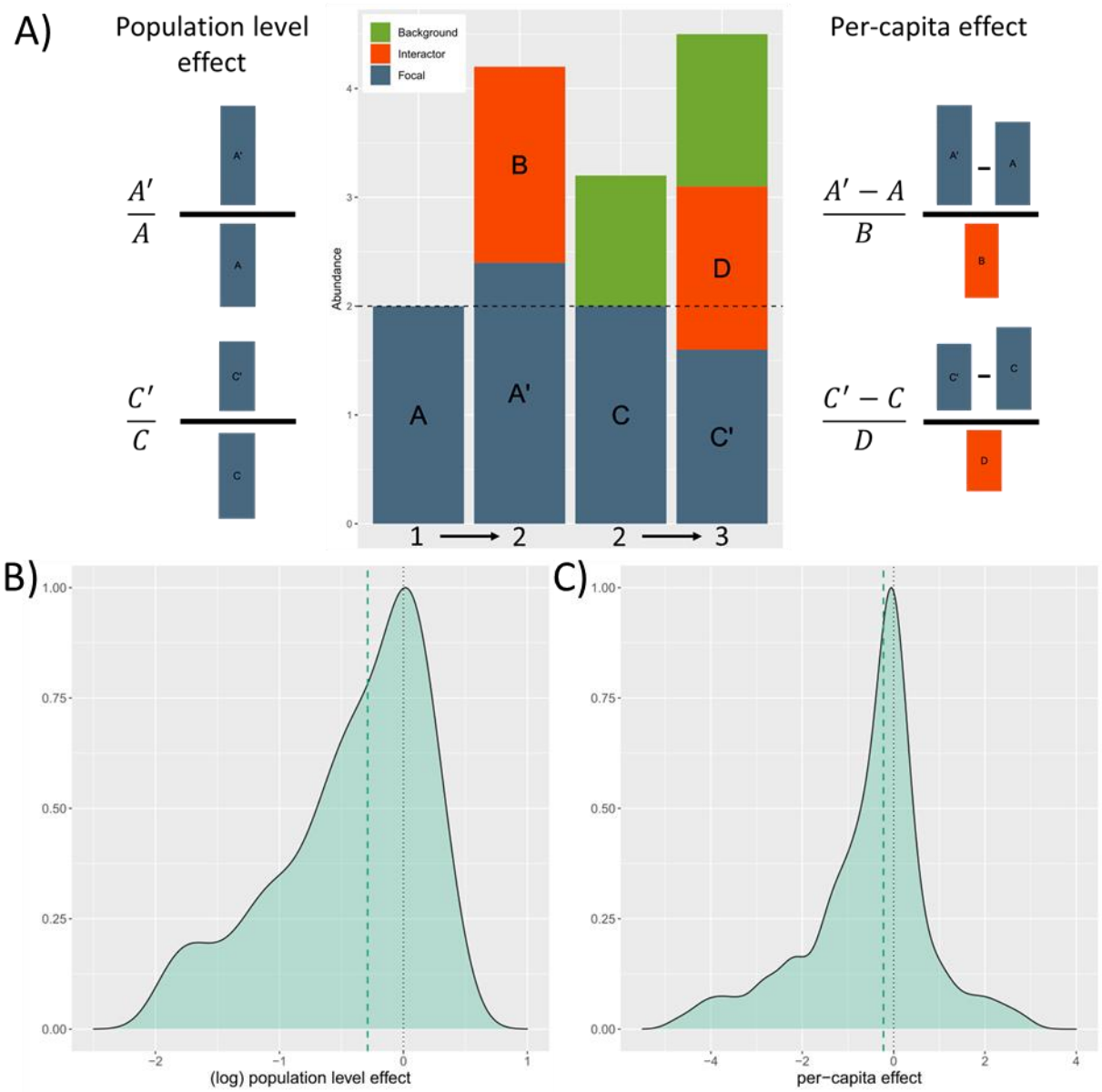


Figure 1.2: Distributions of observed interactions A) Interactions between a "focal" isolate and "interactor" isolate were calculated as two measures, 1) a population effect, calculated as the ratio of the focal isolate's density with and without the interactor present, and 2) a per-capita effect, calculated as the change in density of the focal isolate between contexts with and without the interactor, scaled by the abundance of the interactor. Interactions were always calculated between communities varying by a single isolate – the interactor. However, additional isolates ("background" isolates) could also be present in the compared communities. The "richness context" of an interaction refers to the richness of the pairs of community contexts from which an interaction is observed (e.g., 1=>2 for the first example interaction, 2=>3 for the second example interaction including a "background" isolate). C) the distribution of all observed interactions, as population level effects, (natural-log transformed to symmetrize ratios), C) the distribution of all observed interactions, as per-capita effects

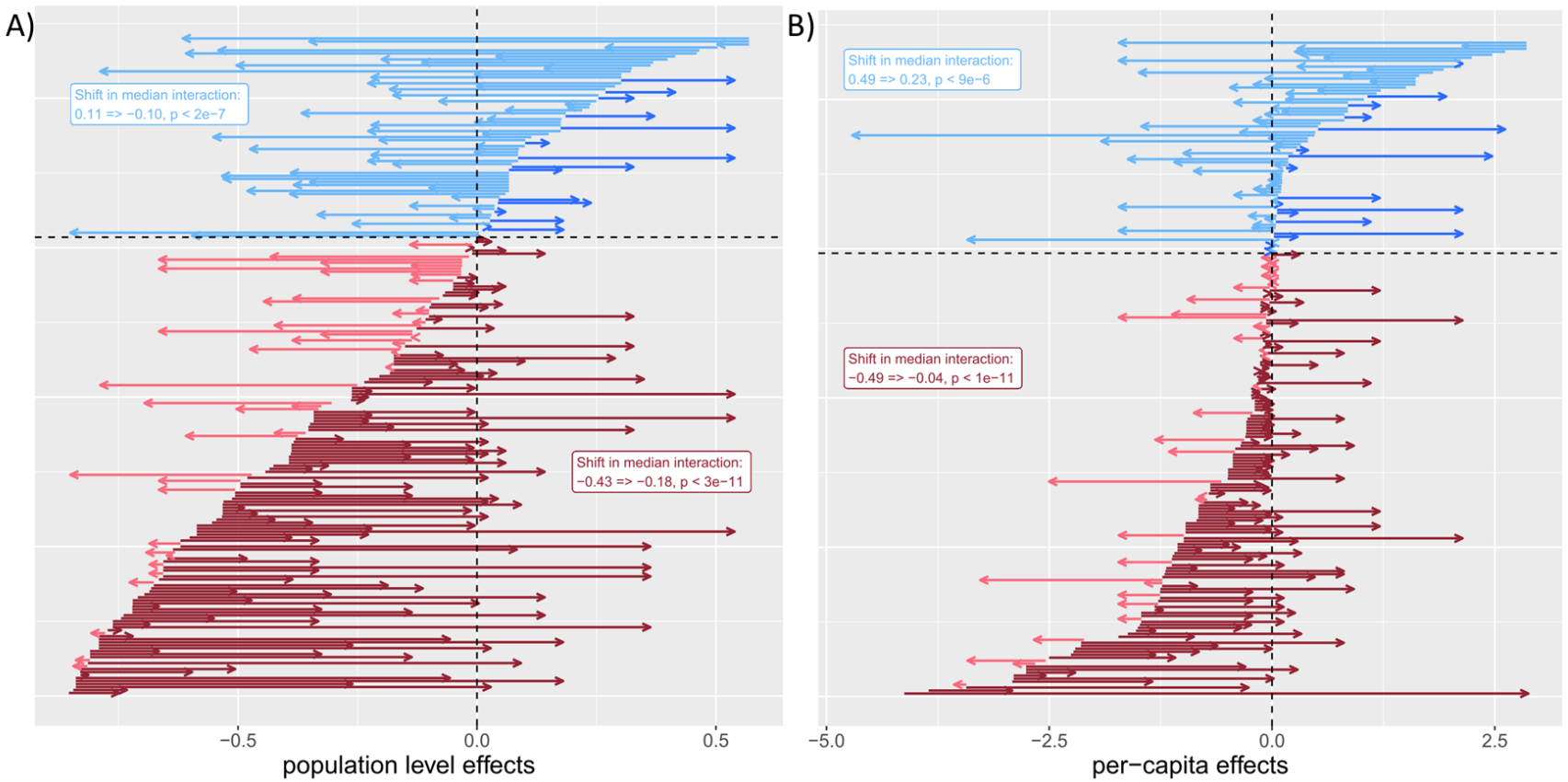


Figure 1.3: Interactions attenuated as richness increased – The shift in interactions between richness contexts varying by a richness of one for A) population level, and B) per-capita effects. All comparisons of an interaction that could be compared between richness contexts varying in richness by a single isolate. Each arrow on the plots represents an interaction observed in two separate richness contexts (e.g., 1=>2 and 2=>3 community members), with the tail of the arrow representing the value of the interaction in the lower richness context and the head of the arrow representing the value observed in the higher richness context. Arrows are colored by the initial direction of the interaction (positive or negative) and the shift in direction of the interaction (positive or negative). Labels display the shift in median interactions for initially positive and negative interactions, respectively, with p-values summarizing the outcome of Wilcoxon signed-rank tests to determine if the shift in interaction value represented a significant change.

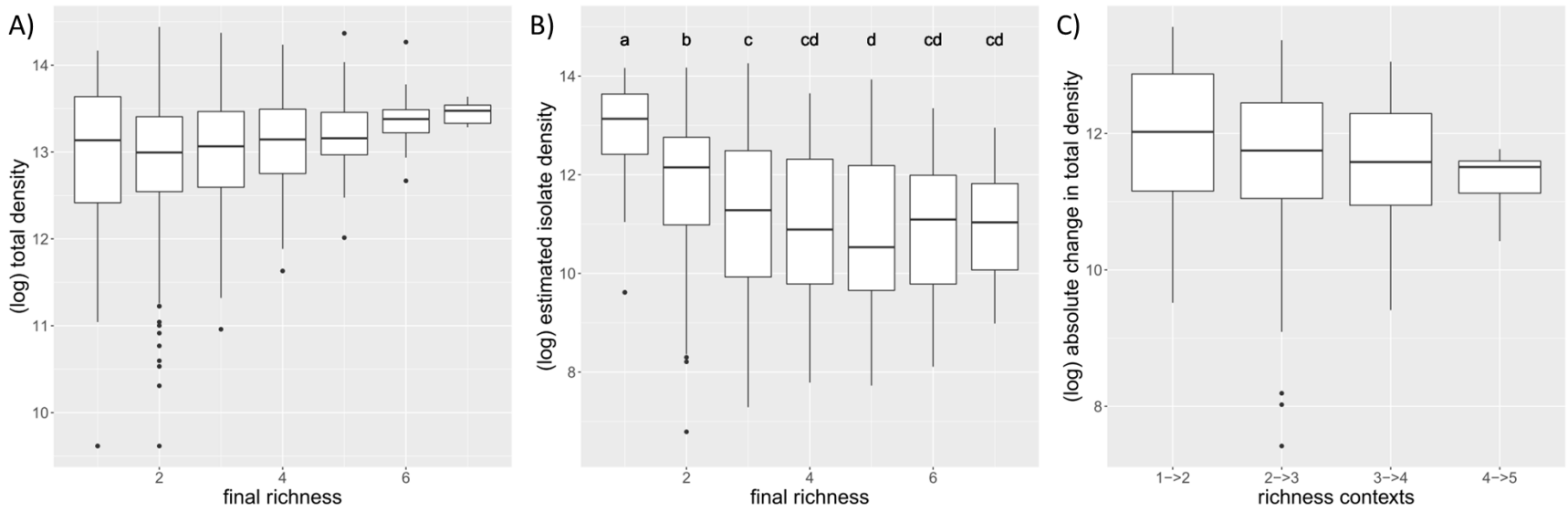


Figure 1.4: Relationships between richness and density – A) the relationship between final community richness and total community density, B) the relationship between richness and individual isolate density (significantly distinct groups were determined through post hoc pairwise t-tests using the Holm method for multiple testing correction), C) the absolute change in total community density associated with an interaction, grouped by the richness context in which interactions were observed (interactions from richness contexts with fewer than 3 observations were removed). For all plots, density is plotted on a log scale.

model	df	adjusted R^2	p-value
focal change ~ total change	1	0.3336	$< 2.2e^{-16}$
focal change ~ n-context	5	0.0829	$2.2e^{-7}$
focal change ~ total change + n-context	6	0.4293	$< 2.2e^{-16}$
focal change ~ total change : n-context	6	0.5011	$< 2.2e^{-16}$
focal change ~ total change * n-context	11	0.5652	$< 2.2e^{-16}$

Table 1.1: Summary of linear regressions modelling the change in density of a focal isolate (i.e., the effect of an interaction) by emergent community properties. "Focal change" indicates the change in density of the focal isolate in the predicted context. "Total change" indicates the change in total density between the community contexts of the interaction. "n-context" indicates the richness contexts over which the interaction was observed. "+" models separate variables with no interaction, ":" models only an interaction, "*" models separate variables with an interaction term.

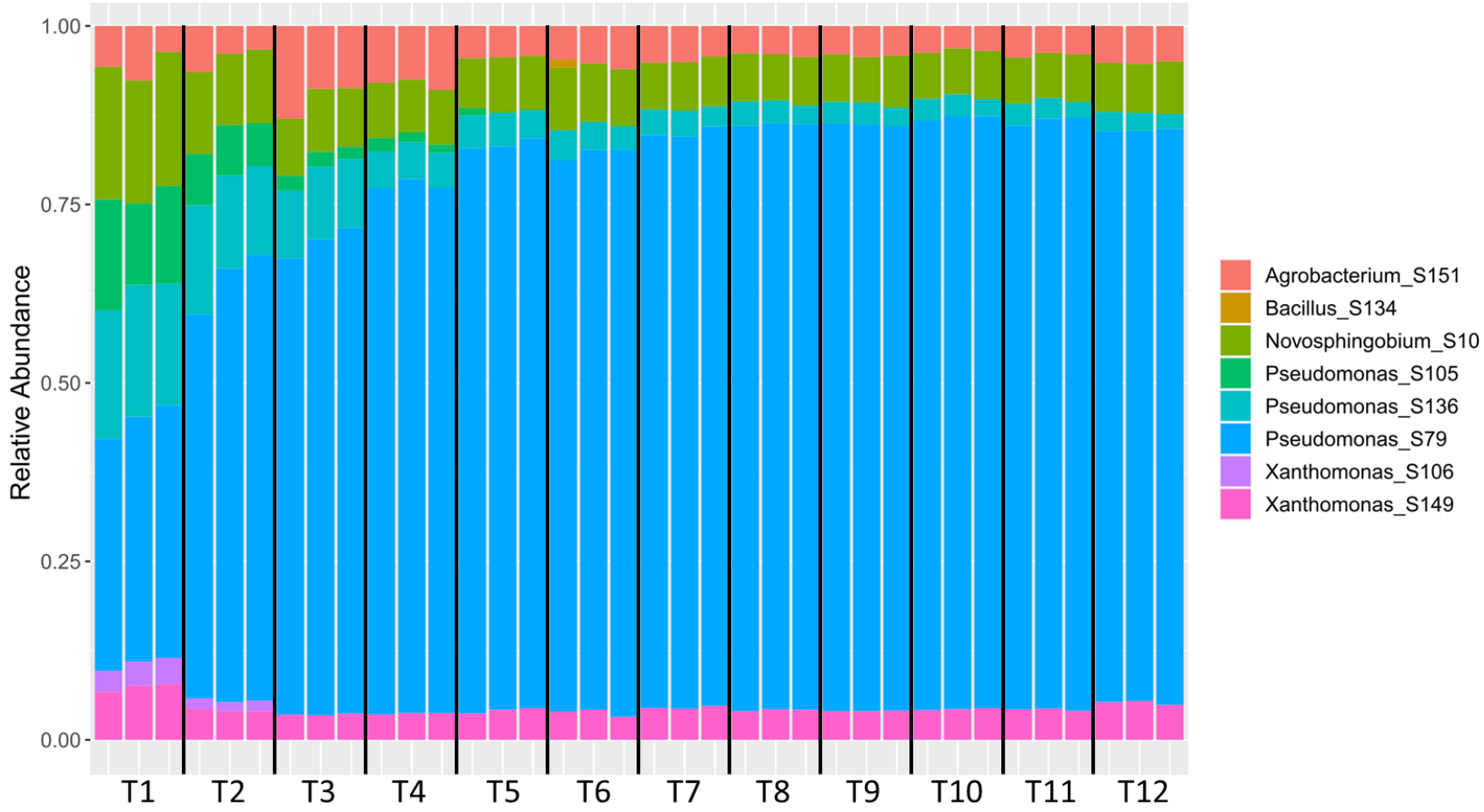
model	df	adjusted R^2	p-value
focal change ~ total change	1	0.273	$< 2.2e^{-16}$
focal change ~ n-contexts	4	0.0226	0.035
focal change ~ interaction effect	1	0.1588	$1.988e^{-12}$
focal change ~ total change + interaction effect	2	0.4169	$< 2.2e^{-16}$
focal change ~ total change * interaction effect	3	0.421	$< 2.2e^{-16}$

Table 1.2: Summary of linear regressions modelling the predictive power of interactions between contexts differing in richness context by a single community member. "Focal change" indicates the change in density of the focal isolate in the predicted context. "Total change" indicates the change in total density between the community contexts of the interaction. "n-contexts" indicates the richness contexts from the pair of interactions (e.g., 1=>2 & 2=>3). "Interaction effect" indicates the change in density of the focal isolate in the interaction context which was not being predicted. We considered predictive power of interactions from the bottom-up, i.e., "interaction effects" came from the lower richness context (continuing the example above, 1=>2), while "total change" came from the predicted higher richness context (2=>3), as in the models described in Table 1.1. "+" models separate variables with no interaction, "*" models separate variables with an interaction term.

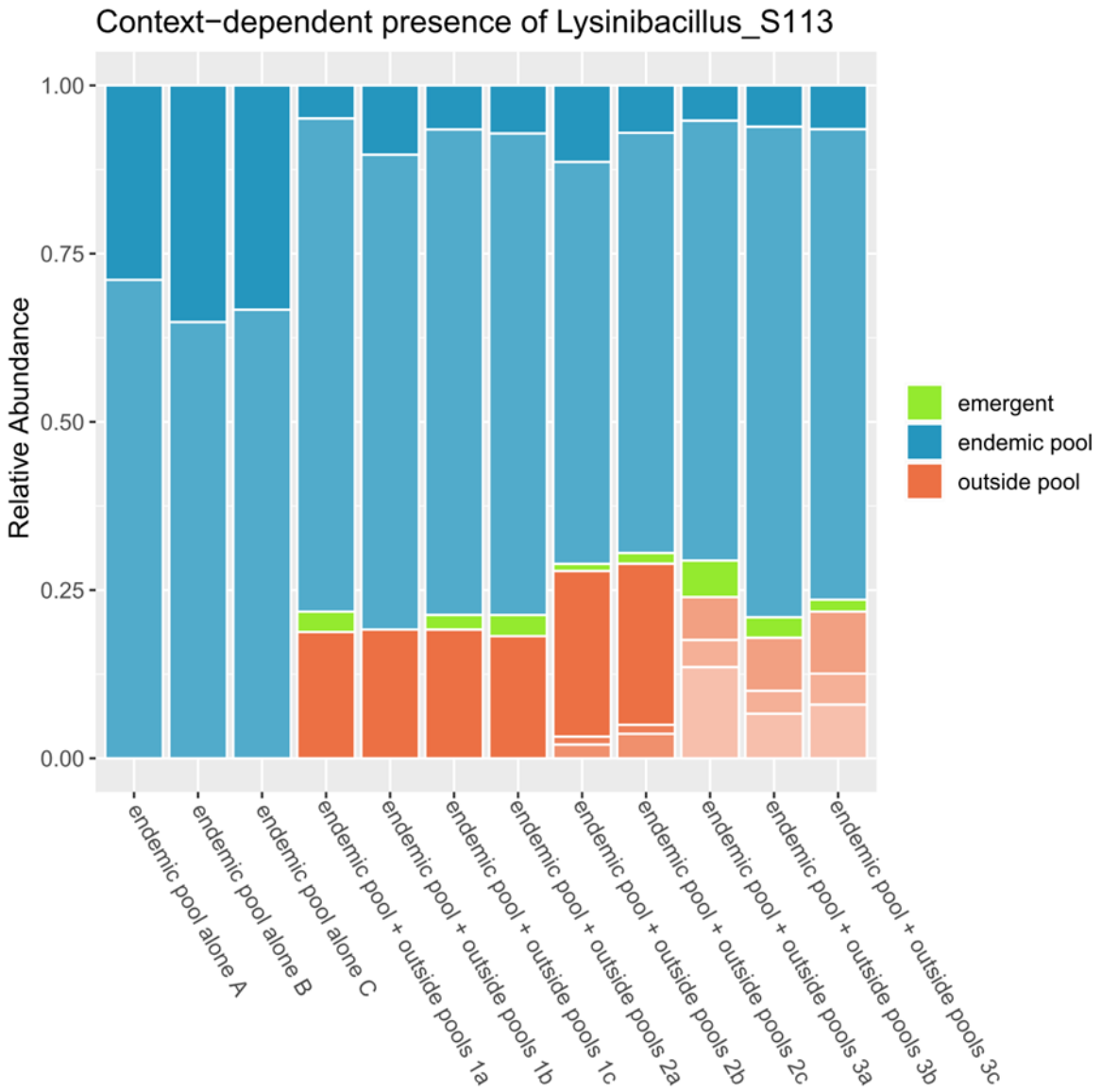
1.8 Supplementary figures and tables

name	in followup communities	genome length (bp)	number contigs	N50	GC content	percent completion	percent redundancy	Genus
CFBP2214_S67	1	5577436	33	662387	59.1	100.0	0.0	Agrobacterium
fls2-241-TYG-188a_S156	1	5347584	16	1492510	57.1	100.0	0.0	Agrobacterium
lyk4-40-TYG-31_S151	1	5137870	25	466280	57.5	100.0	1.4	Agrobacterium
efr-133-R2A-120_S25	1	5124622	81	124276	63.7	98.6	1.4	Arthrobacter
efr-133-R2A-63_S85	0	4425045	59	134882	63.9	98.6	1.4	Arthrobacter
efr-133-TYG-118_S22	0	5279483	50	176600	63.6	98.6	5.6	Arthrobacter
efr-133-TYG-120_S21	0	4970673	55	159873	63.7	98.6	1.4	Arthrobacter
fls2-241-R2A-200_S19	1	3813896	27	255168	62.6	100.0	0.0	Arthrobacter
lyk4-R2A-2_S134	1	3638498	19	764156	42.0	100.0	0.0	Bacillus
MEB006b_S157	1	2943144	11	455680	66.3	100.0	0.0	Brevundimonas
lyk4-R2A-23_S48	0	7707676	60	202001	66.6	100.0	2.8	Burkholderia
MEB011_S37	0	3658679	20	352868	70.9	100.0	0.0	Curtobacterium
ME-Dv--P-122a_S32	0	3769316	20	285580	71.1	100.0	0.0	Curtobacterium
efr-133-R2A-84_S152	1	4445808	23	399529	52.9	100.0	0.0	Enterobacter
MEB024_S129	1	4770259	18	465211	54.6	100.0	0.0	Enterobacter
ME-P-080_S24	0	3275705	11	588193	71.5	100.0	2.8	Frigoribacterium
fls2-241-R2A-40a_S122	0	3753110	4	3749333	69.5	98.6	1.4	Leifsonia
ME-Dv--P-043b_S39	0	4411491	50	171549	66.2	100.0	1.4	Luteibacter
fls2-241-R2A-57_S113	1	5275670	229	35215	36.4	100.0	1.4	Lysinibacillus
LMC-P-041_S56	0	3701438	14	543905	67.8	100.0	0.0	Microbacterium
lyk4-40-TSB-66_S53	0	3570227	18	477020	70.3	100.0	0.0	Microbacterium
lyk4-40-TYG-92_S29	0	6170652	34	342268	66.8	98.6	1.4	Mycolicibacterium
fls2-241-R2A-195_S10	1	5738748	61	187846	65.2	100.0	2.8	Novosphingobium
LMC-P-059a_S33	0	3889587	11	597505	69.7	100.0	0.0	Plantibacter
lyk4-40-MEA-4_S41	0	3969825	13	626676	69.5	100.0	0.0	Plantibacter
ME-Dv--P-095_S58	0	4005071	21	458290	69.6	100.0	0.0	Plantibacter
ME-Dv--P-122b_S64	0	3920291	14	374919	69.3	100.0	4.2	Plantibacter
MEB111_S57	0	6233258	35	311840	72.5	98.6	5.6	Promicromonospora
efr-133-R2A-89_S7	0	4553351	18	526635	66.2	98.6	1.4	Pseudarthrobacter
fls2-241-R2A-127_S5	0	4696413	59	170107	65.6	98.6	2.8	Pseudarthrobacter
fls2-241-R2A-168_S45	0	4884402	64	178784	64.4	98.6	1.4	Pseudarthrobacter
lyk4-40-TYG-27_S50	0	4557169	24	542361	66.2	98.6	1.4	Pseudarthrobacter
CFBP2511_S1	0	6188923	48	215684	58.6	100.0	1.4	Pseudomonas
efr-133-R2A-59_S78	0	6207741	41	223084	58.9	100.0	0.0	Pseudomonas
efr-133-TYG-103a_S91	1	6115300	51	206697	60.3	98.6	0.0	Pseudomonas
efr-133-TYG-23_S107	1	5676264	58	159512	60.7	100.0	0.0	Pseudomonas
efr-133-TYG-5_S79	1	6116959	13	793520	62.1	94.4	1.4	Pseudomonas
fls2-241-R2A-110_S136	1	6839427	53	218702	59.0	98.6	0.0	Pseudomonas
fls2-241-TYG-175_S105	1	5975177	64	153260	59.3	98.6	1.4	Pseudomonas
FR229a_S127	1	5926019	46	231831	60.2	98.6	1.4	Pseudomonas
lyk4-40-TSB-59a_S125	1	6355326	26	392868	60.5	98.6	1.4	Pseudomonas
lyk4-R2A-10_S69	1	6325665	30	494191	60.6	98.6	1.4	Pseudomonas
lyk4-R2A-8_S103	1	5480925	44	211405	61.1	97.2	0.0	Pseudomonas
lyk4-TYG-107_S51	1	6149173	50	260186	60.1	97.2	1.4	Pseudomonas
MEB105_S97	1	6091259	24	328361	60.3	100.0	0.0	Pseudomonas
MEJ086_S119	1	5850881	64	164366	62.3	100.0	0.0	Pseudomonas
MEJ108_S148	0	5910147	87	130449	59.2	100.0	1.4	Pseudomonas
ME-P-057_S98	0	5703857	59	159821	60.1	98.6	0.0	Pseudomonas
MEB032_S102	0	7154115	67	202086	62.3	98.6	2.8	Rhodococcus
MEB041_S88	0	4571639	15	1089805	68.5	98.6	1.4	Rhodococcus
efr-133-TYG-104_S8	0	3849934	38	178207	62.6	100.0	0.0	Arthrobacter
fls2-241-R2A-172_S17	0	4473686	27	565181	62.9	100.0	2.8	Arthrobacter
MEB009_S83	0	4281525	28	228093	66.2	98.6	1.4	Pseudarthrobacter
efr-133-TYG-130_S11	1	7080647	64	245655	67.2	100.0	2.8	Variovorax
fls2-241-TYG-148_S106	1	4965705	32	273625	65.4	100.0	0.0	Xanthomonas
LMC-A-07_S149	1	4937033	31	289890	65.3	100.0	1.4	Xanthomonas

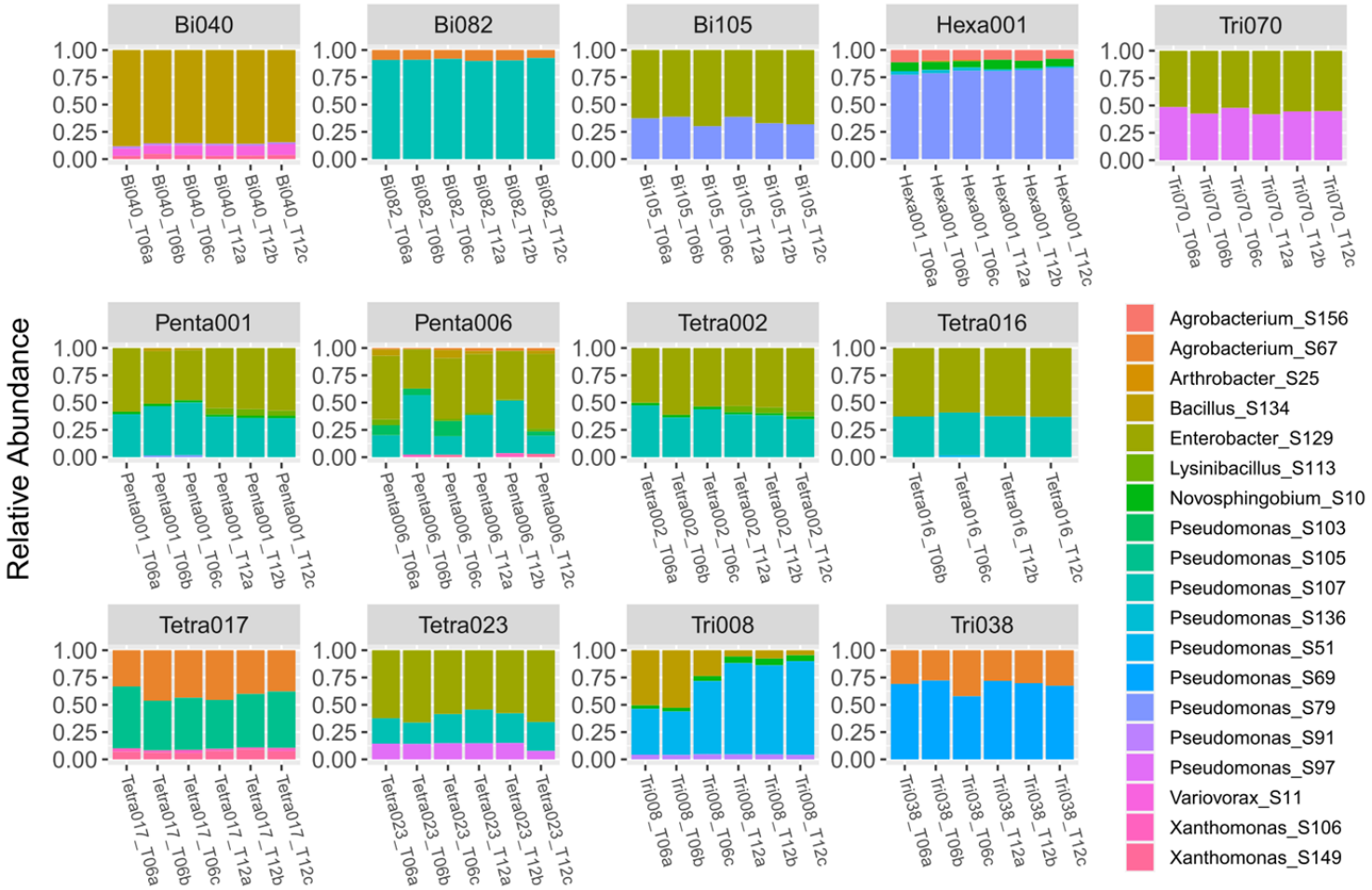
Supplementary Table 1.1: isolate detail



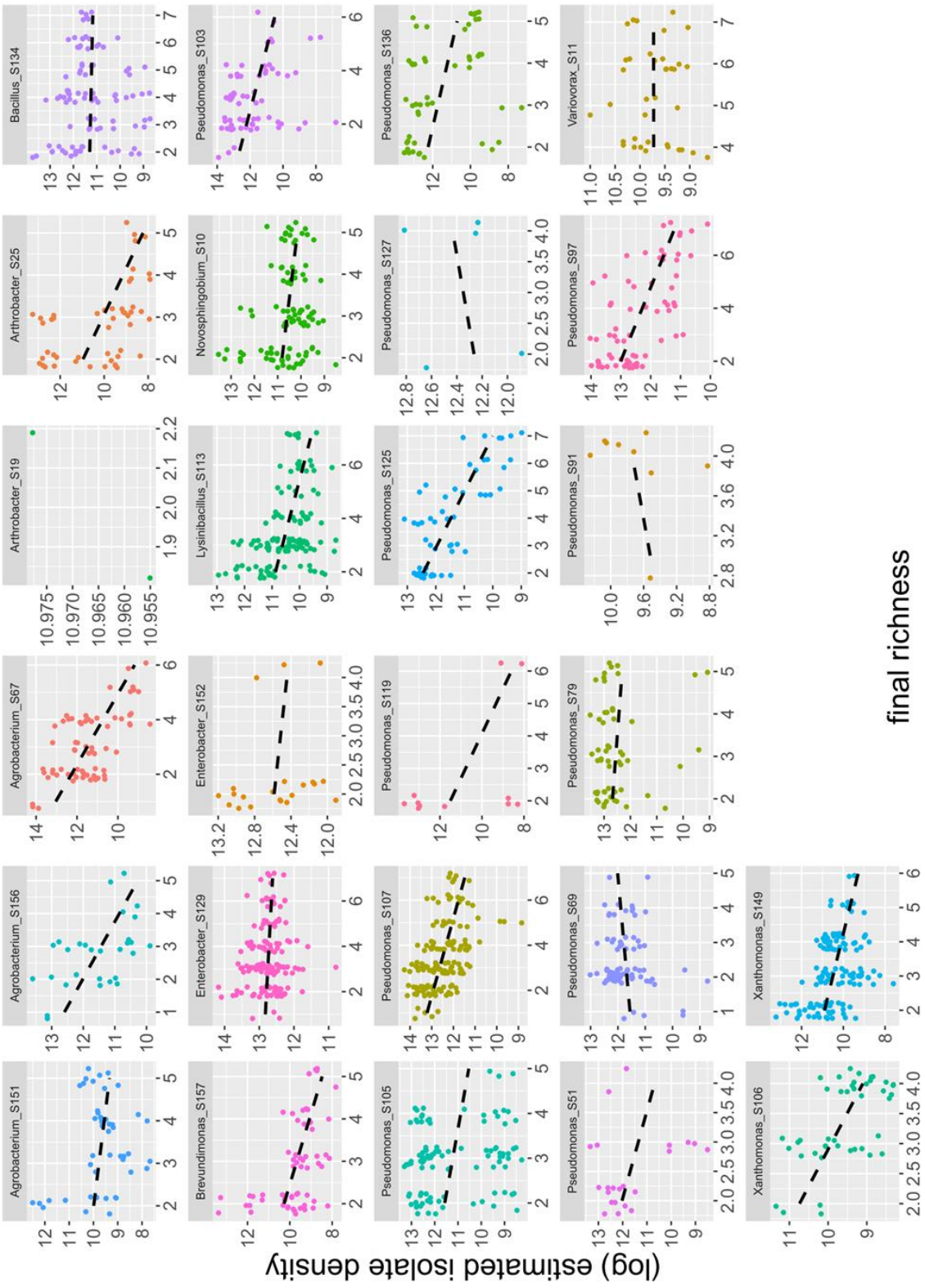
Supplementary Figure 1.1: A time course of community dynamics for an 8-member synthetic community. Each set of three stacked bars represents the composition of three replicates 24 hours after initial assembly or passaging for 12 days (T1 – T12). Dynamics play out quickly within the first four days and settle into a stable composition to until at least 12 days post initial assembly.



Supplementary Figure 1.2: An example of context-dependent coexistence of a Lysinibacillus isolate ("emergent" isolate) from the initial set of synthetic communities. That Lysinibacillus isolate was excluded by the other members of the 8-member pool to which it belonged ("endemic pool alone A-C"). However, in the three additional contexts shown here ("endemic pool + outside pools 1-3") which were assembled from that pool of 8 and at least one other pool, the Lysinibacillus isolate persisted to 6-days. The relative abundances of the non-emergent isolates are depicted by stacked bars in shades of blue (if from the "endemic pool" to which the Lysinibacillus isolate belonged) or shades of orange (if from pools other than the endemic pool).



Supplementary Figure 1.3: A comparison of day-6 and day-12 compositions for 13 different communities assembled for measuring interactions. Timepoint is displayed in the x-axis text for each stacked bar. Stacked bars are colored by isolate identity. As seen in the full time-series (supplementary figure 1.1), community compositions after 6 days is generally representative of community compositions after 12 days. These communities were assembled from pools varying in initial richness (2-8 isolates), but all show generally consistent compositions between days 6 and 12, suggesting assemblages with higher initial richness do not require a longer period to reach a stable composition.



Supplementary Figure 1.4: Individual isolate density generally decreased as richness increased. Density is plotted on a log scale, and its relationship with final community richness is summarized with a line of best fit.

model	df	adjusted R^2	p-value
all contexts			
focal change ~ total change	1	0.3521	$< 2.2e^{-16}$
focal change ~ interaction effect	1	0.09991	$3.428e^{-14}$
focal change ~ total change + interaction effect	2	0.4298	$< 2.2e^{-16}$

Supplementary Table 1.2: Summary of linear regressions modelling the predictive power of interactions across any richness contexts (e.g., 1=>2 & 4=>5). "Focal change" indicates the change in density of the focal isolate in the predicted context. "Total change" indicates the change in total density between the community contexts of the interaction. "Interaction effect" indicates the change in density of the focal isolate in the interaction context which was not being predicted. We considered predictive power of interactions from the bottom-up, i.e., "interaction effects" came from the lower richness context (continuing the example above, 1=>2), while "total change" came from the predicted higher richness context (4=>5). "+" models separate variables with no interaction, "*" models separate variables with an interaction term.

1.9 References

Amor, Daniel R., Christoph Ratzke, and Jeff Gore. "Transient invaders can induce shifts between alternative stable states of microbial communities." *Science advances* 6.8 (2020): eaay8676.

Aranda-Díaz, Andrés, et al. "Bacterial interspecies interactions modulate pH-mediated antibiotic tolerance." *Elife* 9 (2020): e51493.

Baichman-Kass, Amichai, Tingting Song, and Jonathan Friedman. "Interactions between culturable bacteria are highly non-additive." *bioRxiv* (2022): 2022-09.

Bairey, Eyal, Eric D. Kelsic, and Roy Kishony. "High-order species interactions shape ecosystem diversity." *Nature communications* 7.1 (2016): 1-7.

Bankevich, Anton, et al. "SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing." *Journal of computational biology* 19.5 (2012): 455-477.

BBMap – Bushnell B. – sourceforge.net/projects/bbmap/

Brenner, Katie, Lingchong You, and Frances H. Arnold. "Engineering microbial consortia: a new frontier in synthetic biology." *Trends in biotechnology* 26.9 (2008): 483-489.

Billick, Ian, and Ted J. Case. "Higher order interactions in ecological communities: what are they and how can they be detected?." *Ecology* 75.6 (1994): 1529-1543.

Case, Ted J., and Edward A. Bender. "Testing for higher order interactions." *The American Naturalist* 118.6 (1981): 920-929.

Chang, Chang-Yu, et al. "Emergent coexistence in multispecies microbial communities." *bioRxiv* (2022).

Connors, Bryce M., et al. "Model-guided design of the diversity of a synthetic human gut community." *bioRxiv* (2022): 2022-03.

Clark, Ryan L., et al. "Design of synthetic human gut microbiome assembly and butyrate production." *Nature communications* 12.1 (2021): 3254.

Diamond, Jared M. "Assembly of species communities." *Ecology and evolution of communities* (1975): 342-444.

Eren, A. Murat, et al. "Community-led, integrated, reproducible multi-omics with anvi'o." *Nature microbiology* 6.1 (2021): 3-6.

Estrela, Sylvie, et al. "Nutrient dominance governs the assembly of microbial communities in mixed nutrient environments." *Elife* 10 (2021): e65948.

Estrela, Sylvie, et al. "Functional attractors in microbial community assembly." *Cell Systems* 13.1 (2022): 29-42.

Falkowski, Paul G., Tom Fenchel, and Edward F. Delong. "The microbial engines that drive Earth's biogeochemical cycles." *Science* 320.5879 (2008): 1034-1039.

Foster, Kevin R., and Thomas Bell. "Competition, not cooperation, dominates interactions among culturable microbial species." *Current biology* 22.19 (2012): 1845-1850.

Fox, John, and Sanford Weisberg. *An R companion to applied regression*. Sage publications, 2018.

Friedman, Jonathan, Logan M. Higgins, and Jeff Gore. "Community structure follows simple assembly rules in microbial microcosms." *Nature ecology & evolution* 1.5 (2017): 1-7.

Goldford, Joshua E., et al. "Emergent simplicity in microbial community assembly." *Science* 361.6401 (2018): 469-474.

Gonze, Didier, et al. "Microbial communities as dynamical systems." *Current opinion in microbiology* 44 (2018): 41-49.

Gotelli, Nicholas J., and Declan J. McCabe. "Species co-occurrence: a meta-analysis of JM Diamond's assembly rules model." *Ecology* 83.8 (2002): 2091-2096.

Grilli, Jacopo, et al. "Higher-order interactions stabilize dynamics in competitive network models." *Nature* 548.7666 (2017): 210-213.

Horner-Devine, M. Claire, et al. "A comparison of taxon co-occurrence patterns for macro-and microorganisms." *Ecology* 88.6 (2007): 1345-1353.

Hsu, Ryan H., et al. "Microbial interaction network inference in microfluidic droplets." *Cell systems* 9.3 (2019): 229-242.

Kehe, Jared, et al. "Massively parallel screening of synthetic microbial communities." *Proceedings of the National Academy of Sciences* 116.26 (2019): 12804-12809.

Kehe, Jared, et al. "Positive interactions are common among culturable bacteria." *Science advances* 7.45 (2021): eabi7159.

Liu, Yong-Xin, Yuan Qin, and Yang Bai. "Reductionist synthetic community approaches in root microbiome research." *Current Opinion in Microbiology* 49 (2019): 97-102.

Maynard, Daniel S., Zachary R. Miller, and Stefano Allesina. "Predicting coexistence in experimental ecological communities." *Nature ecology & evolution* 4.1 (2020): 91-100.

McNally, Luke, and Sam P. Brown. "Building the microbiome in health and disease: niche construction and social conflict in bacteria." *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1675 (2015): 20140298.

Medlock, Gregory L., et al. "Inferring metabolic mechanisms of interaction within a defined gut microbiota." *Cell systems* 7.3 (2018): 245-257.

Mickalide, Harry, and Seppe Kuehn. "Higher-order interaction between species inhibits bacterial invasion of a phototroph-predator microbial community." *Cell systems* 9.6 (2019): 521-533.

Momeni, Babak, Li Xie, and Wenying Shou. "Lotka-Volterra pairwise modeling fails to capture diverse pairwise microbial interactions." *Elife* 6 (2017): e25051.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ratzke, Christoph, Julien Barrere, and Jeff Gore. "Strength of species interactions determines biodiversity and stability in microbial communities." *Nature ecology & evolution* 4.3 (2020): 376-383.

Rohland, Nadin, and David Reich. "Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture." *Genome research* 22.5 (2012): 939-946.

Russel, Jakob, et al. "Antagonism correlates with metabolic similarity in diverse bacteria." *Proceedings of the National Academy of Sciences* 114.40 (2017): 10684-10688.

Sanchez-Gorostiaga, Alicia, et al. "High-order interactions distort the functional landscape of microbial consortia." *PLoS Biology* 17.12 (2019): e3000550.

Schäfer, Martin, et al. "Mapping phyllosphere microbiota interactions in planta to establish genotype–phenotype relationships." *Nature microbiology* 7.6 (2022): 856-867.

Sundarraman, Deepika, et al. "Higher-order interactions dampen pairwise competition in the zebrafish gut microbiome." *MBio* 11.5 (2020): e01667-20.

Vandermeer, John H. "The competitive structure of communities: an experimental approach with protozoa." *Ecology* 50.3 (1969): 362-371.

Venturelli, Ophelia S., et al. "Deciphering microbial interactions in synthetic human gut microbiome communities." *Molecular systems biology* 14.6 (2018): e8157.

Quinn, Thomas P., et al. "propr: an R-package for identifying proportionally abundant features using compositional data analysis." *Scientific reports* 7.1 (2017): 1-9.

Weiss, Anna S., et al. "In vitro interaction network of a synthetic gut bacterial community." *The ISME journal* 16.4 (2022): 1095-1109.

Wickham, Hadley. "Reshaping data with the reshape package." *Journal of statistical software* 21 (2007): 1-20.

Wickham, Hadley, et al. "Welcome to the Tidyverse." *Journal of open source software* 4.43 (2019): 1686.

Wootton, J. Timothy. "Indirect effects and habitat use in an intertidal community: interaction chains and interaction modifications." *The American Naturalist* 141.1 (1993): 71-89.

Wootton, J. Timothy. "Predicting direct and indirect effects: an integrated approach using experiments and path analysis." *Ecology* 75.1 (1994): 151-165.

Wootton, J. Timothy, and Mark Emmerson. "Measurement of interaction strength in nature." *Annu. Rev. Ecol. Evol. Syst.* 36 (2005): 419-444.

Yu, Xiaoqian, Martin F. Polz, and Eric J. Alm. "Interactions in self-assembled microbial communities saturate with diversity." *The ISME journal* 13.6 (2019): 1602-1617.

Chapter 2: Pairwise observations of coexistence make accurate, but incomplete, predictions in synthetic bacterial communities

2.1 Abstract

Ecological communities are often studied from the "bottom-up", as an assemblage described by a network of pairwise relationships. However, some communities possess features that cannot be explained from a pairwise perspective, provoking the perspective that we should study communities from the "top-down", where emergent properties have the potential to present themselves. Here, we assembled all pairs and all *n-2* communities from a set of 21 bacterial isolates to compare observations of coexistence between pairwise and 19-member "complex" contexts. In doing so, we evaluated the importance of emergent effects in determining the composition of simple microbial communities. We observed that isolates that coexisted in pairwise frequently did not coexist in more complex contexts, but that this failure to coexist was generally predictable based on observations of pairwise exclusion with other community members. Further, we observed that most isolates that were excluded in pairwise were also excluded by that same competitor in more complex assemblages. And although context-dependent coexistence was frequently observed in complex contexts, most instances of such unexpected persistence involved only a small set of frequently observed isolates. Finally, we found that prevalence ranks when measured in pairwise and complex contexts had a strong positive correlation. Thus, pairwise observations of coexistence and exclusion can go a far way in explaining the composition of complex assemblages, a result with useful implications for the rational design of simple microbial communities.

2.2 Introduction

Coexistence of species in an ecological community is often studied from the "bottom-up" by decomposing communities into constituent interactions between community members (Gause, 1934; May 1972; Chesson, 2000). In microbial ecology, this perspective is embodied by efforts to use observations of pairs of community members studied in isolation to describe or predict more complex assemblages. This reductionist approach has proven useful for predicting the composition and function of ecological communities (e.g., Vandermeer, 1969; Richmond et al., 1975; Foster and Bell, 2012; Venturelli et al., 2018). However, it has also been demonstrated that this bottom-up approach can provide an incomplete understanding of ecological communities, especially as diversity increases (Friedman et al., 2017).

It is unsurprising that such a reductionist approach can fail to accurately describe microbial communities, especially when one considers the disparity in complexity between a pairwise coculture and a natural microbial community. A change in the interaction between two species across different community contexts is referred to as a "higher order interaction" (HOI), and many mechanisms can give rise to such an effect in microbial communities. Microbes can have profound impacts on their environments, and through mechanisms such as biofilm formation (Elias and Banin, 2012; Lee et al., 2014), environmental pH change (Aranda-Díaz et al., 2020; Ratzke et al., 2020) and production of inhibitory metabolites (Lax and Gore, 2022), the activity of one microbe might change the relationship between others. Additionally, given that many microbial interactions are mediated through metabolic competition or exchange (cross-feeding) (Gralka et al., 2020), changes in community context that alter the functional potential of a microbial community can change the available nutrient environment, with widespread effects on the community (Pascual-García et al., 2020; Dal Bello et al., 2021; Diaz-Colunga et al., 2022).

Debate as to the importance of HOIs in ecological systems is longstanding (Case and Bender, 1981; Pomerantz, 1981; Wootton, 1993) and ongoing in microbial systems, with evidence both supporting (Mickalide and Kuehn, 2019; Sundarraman, 2020; Schäfer et al., 2022) and minimizing (Vandermeer, 1969; Foster and Bell, 2012; Venturelli et al., 2018) their importance. Part of the debate stems from the challenge of identifying HOIs, and specifically distinguishing them from indirect effects, where the effect of one species on another is mediated by a density change in an intermediate species (Billick and Case, 1994). Insight into the importance of HOIs and indirect effects may be gleaned by studying microbial communities from the "top-down", i.e., focusing on how community members behave and interact when embedded in a complex community of more than two members (Carlström et al., 2019; Chang et al., 2022; Romdhane et al., 2022). As with the bottom-up approach, top-down community data has been used to predict the composition and function of microbial communities (Maynard et al., 2020; Clark et al., 2021; Schäfer et al., 2022).

Whether HOIs are important will strongly impact efforts to design microbial communities to assume a desired composition or perform a desired function (Brenner et al., 2008; Liu et al., 2019; Clark et al., 2021). With such goals in mind, one might then ask whether it would be most effective to study microbial communities from the bottom-up or top-down. The top-down approach is appealing from an experimental perspective because a single observation of a complex community can provide information about the relationships between multiple members (Maynard et al., 2020). Indeed, experimentally studying all subcommunities required to thoroughly investigate HOIs quickly becomes experimentally infeasible for anything but the simplest natural microbial communities. Only if HOIs have relatively little impact is a bottom-up approach feasible because then pairwise communities alone are needed for direct measurement of interactions. Additionally, it has been suggested that leveraging observations from both perspectives can improve predictions of microbial community composition and function (Maynard et al., 2020).

Ultimately, we must ask, what are the limitations of pairwise observations in describing more complex communities? In an effort to relate how observations of coexistence in pairwise and complex contexts relate to one another, we decomposed a 21-member synthetic bacterial community into all possible pairs and *n-2* subcommunities to compare coexistence observed from the bottom-up and top-down.

2.3 General Methods

Assembly of synthetic communities – We began with a set of 21 bacterial strains isolated from the leaves of wild and field-grown *Arabidopsis thaliana* (supplementary table 2.1). From this set, we assembled all two-member and 19-member (*n-2*) communities, representing "pairwise" and "complex" communities, respectively (figure 2.1). Each community was inoculated in triplicate into a custom growth-medium derived from *A. thaliana* leaves (*Arabidopsis* leaf medium, ALM) (methods) at a consistent total community titer, with each member accounting for an equal proportion of the population given the initial richness (number of community members). To allow the communities to reach a steady state reflective of the long-term composition, we passaged each community for 7 days by performing a 1:100 dilution into fresh medium every 24 hours. This period was sufficiently long to allow the community composition to stabilize (supplementary figure 2.1). We characterized the composition of all synthetic communities by mapping Illumina short reads against a nearly complete and high-quality genome assembled for each isolate (methods). These final, top-down communities, ranged in richness from 2 to 14 isolates.

2.4 Results

Coexistence in complex communities was a good predictor of pairwise coexistence – We began our comparisons of coexistence between complex and pairwise contexts by evaluating how well coexistence in complex contexts predicted coexistence in pairwise coculture. To do this, we decomposed the final composition of each complex community into its constituent pairs and determined what proportion of those pairs coexisted in pairwise (figure 2.2a). We observed that a high proportion of pairs coexisting in complex communities also coexisted in pairwise (median ~98%), demonstrating that coexistence in a complex context was a strong predictor of coexistence in pairwise.

Only a small set of isolates that failed to coexist in pairwise coexisted in complex communities – Despite our observation that ~98% of pairs coexisting in complex communities also coexisted in pairwise, 68% of complex communities contained at least one pair of species that did not coexist when alone. These examples of unexpected coexistence based on pairwise results were a subset of the 55 isolate pairs that failed to coexist in our measures of all pairwise interactions. For each of these 55 pairs, we asked if, and how often, that pair coexisted in complex contexts (figure 2.2b). We observed that ~93% of pairs that failed to coexist in pairwise nonetheless coexisted in at least one complex community context. However, when considering all complex communities in which each pair was initially present, these pairs persisted in a median of only 3% of communities. Surprisingly, this general rule did not apply to a small set of pairs (n = 12) that were able to coexist in at least 10% of complex contexts; these 12 pairs represented ~76% of all instances of context-dependent coexistence (figure 2.2d). Thus, although context-dependent coexistence was broadly observed in complex contexts, it was largely represented by a small set of isolates observed very frequently.

Isolates that coexisted in pairwise often failed to coexist in complex communities – We next evaluated how well pairwise coexistence translated to complex community contexts. For each pair that coexisted in

coculture (n = 155, ~74% of all pairs), we identified the proportion of complex contexts (conditioned on whether those isolates were initially present) where the pair coexisted (figure 2.2c). We observed that most of these pairs (n = 110) only coexisted rarely in complex contexts (<20% of contexts). Another group of pairs (n = 30) were more successful, coexisting in ~20-60% of contexts. And a final set of pairs (n = 29) showed robust coexistence, coexisting in >60% of complex contexts.

In investigating the compositions of these rough groupings, we observed that the set of 29 pairs demonstrating robust coexistence were composed of the most common isolates in the complex contexts, representing a set of high prevalence "core" isolates (with prevalence defined as the proportion of communities in which an isolate persisted relative to all communities in which it was initially present). We found that the set of pairs which exhibited middling frequencies of coexistence in complex contexts were generally composed of one "core" isolate and one isolate with intermediate complex-context prevalence. And those pairs that coexisted in the lowest number of top-down contexts were generally composed of isolates with low or middling top-down prevalence. Thus, we observed that aside from the set of pairs containing only core isolates, most pairs observed to coexist in pairwise showed rare or variable coexistence in complex communities. An obvious explanation for the failure of pairs that coexisted in pairwise to coexist in a complex context is that one or both isolates is excluded by other community members in the complex context. To address this, we asked how well our pairwise observations of exclusion can explain the composition of complex communities in the section that follows.

Expectations of persistence in complex contexts from pairwise exclusion were generally accurate – We evaluated how well the final composition of each complex community aligned with coexistence as expected from the pairwise contexts. To do this, we asked whether the final composition of each complex

community could be predicted by knowledge of pairwise competitive exclusion in isolation. As an extremely strict test of this hypothesis, we asked whether membership in each complex assemblage included only isolates that were not excluded by any other isolate from the initial pool when cocultured in pairwise. For example, suppose that in pairwise contexts, isolate A excludes isolate B, which in turn excludes isolate C (figure 2.3a). With our strict application of pairwise results, only A would be expected to persist, as B and C are both excluded in pairwise by another isolate. This is a naïve approach as it assumes the absence of higher order interactions and ignores indirect effects, i.e., how being embedded in the full network of interactions might alter pairwise outcomes.

Aware of this limitation, we assessed how well this approach performed by calculating the number of isolates expected to persist that were excluded (unexpected exclusion) and the number of isolates expected to be excluded that persisted (unexpected coexistence). Expectations of persistence were relatively accurate, with all isolates expected to persist successfully persisting in ~69% of compositions and low levels of unexpected exclusion generally (figure 2.3b). However, this approach often failed to identify isolates that would persist, as demonstrated by the fact that every prediction contained at least one instance of unexpected coexistence (figure 2.3c).

Isolate prevalence was consistent between pairwise and complex contexts – Finally, in further evaluating the consistency of coexistence outcomes between pairwise and complex contexts, we compared isolate prevalence in both datasets. We observed a strong positive correlation in prevalence ranks between the two contexts (Spearman's Rho = 0.87, p-value < $5e^{-7}$, figure 2.4). Consistent with this, when considering instances of competitive exclusion in the pairwise assays, we observed that the isolate that was excluded from a given pair was of a lower prevalence rank among complex communities ~98% of time.

2.5 Discussion

An ecological community can be conceptualized from two perspectives: from the bottom-up, as the sum of its parts, or from the top-down, as a product of emergent community-level properties. Here, we used synthetic bacterial communities to investigate how observations of coexistence taken from either perspective relate to one another, and in doing so, evaluated the importance of community-context-dependence on coexistence. Our results tend to support the proposed "assembly rule" that posits the coexistence of a multispecies assemblage requires coexistence between each pair of species (Friedman et al., 2017; Meroz et al., 2021). First off, in our system, ~98% of pairs present in final complex communities were pairs that coexisted in isolation (figure 2.2a). Further, a failure to coexist in pairwise generally translated into a failure to coexist in more complex contexts (figure 2.2b), demonstrating that a complex context only rarely enabled, coexistence of isolates that could not coexist in a pairwise setting. And on the flip side, unexpected failure to coexist in complex communities (figure 2.2c) was generally due to exclusion that could be explained by the pairwise interactions between excluded isolates and other members of the initial pool (figure 2.5b). Thus, pairwise cooccurrence assays were largely effective in predicting species capable of coexistence in more complex contexts.

Nevertheless, we commonly observed instances of context-dependent coexistence in our complex communities. Indeed, 68% of complex communities contained a pair which did not coexist in pairwise (figure 2.2a). These instances of unexpected coexistence were largely restricted to a set of just 12 pairs, which together constituted ~76% of all instances (figure 2.2d). Thus, although context-dependent coexistence was common, it was limited to few isolate pairs.

The frequency of context-dependent coexistence raises the question of what gave rise to these instances of unexpected coexistence? It is likely that such unexpected coexistence results from

indirect and/or higher order interactions, possibilities that are excluded by our naïve approach to predicting the composition of final communities. Indirect interactions are a likely cause; our approach treats all observations of pairwise exclusion as independent binary effects but, of course, the specific strengths of competitive interactions and dynamics of the community will determine its final composition. Returning to our previous example where isolate A persisted while B and C were excluded (figure 2.3a), it could instead be the case that the exclusion of B by A prevented the exclusion of C by B, resulting in a final composition of A and C. Such indirect effects are well documented as relevant to community composition (Paine, 1966; Calbet and Landry, 1993; Wootton, 1994). Additionally, higher order interactions may have attenuated the effect of some competitive interactions, resulting in unexpected coexistence.

A recent study on synthetic bacterial communities (Chang et al. 2022) identified many examples of pairs of strains from multispecies communities that could not coexist in a pairwise context, suggesting that coexistence was often an emergent community-level property. Although we also found context-dependent coexistence, Chang et al. observed this to a considerably higher degree. In their study, only ~40% of pairs from a set of 13 synthetic communities could coexist in pairwise, while we found that ~74% of pairwise assemblages resulted in coexistence and that, on average, ~98% of pairs from a complex context coexisted in pairwise. One important difference between our experimental systems is that Chang and colleagues cultured their communities in a minimal medium with glucose as a single carbon source, while we used a complex leaf-based medium containing many potential carbon sources at varying concentrations (supplementary table 2.2). The relative complexity of our medium might have facilitated coexistence by enabling multiple bacteria to pursue non-overlapping metabolic strategies, thus decreasing competition through niche partitioning (Baran et al., 2015; Brochet et al., 2021). Thus, if competition was less severe in our experimental setting, the higher levels of coexistence that we observed would have been expected.

Our results further suggest that the relative competitive ability of an isolate was a strong determinant of its fate in a community, irrespective of community context. We observed differences in the general competitive abilities of the isolates in our communities, as some isolates persisted in most communities whereas others showed variable or consistently low persistence (figure 2.4). Interestingly, we observed a strong relationship between prevalence ranks when measured in pairwise or complex settings, indicating that an isolate's ability to persist in a community was consistent across community contexts, despite variation in the number and nature of interactions it faced. This finding is in line with other studies that have found that relative competitive ranking is a strong predictor of microbial community composition (Higgins et al., 2017; Ortiz et al., 2021; Chang et al., 2022).

In summary, we found that the complex communities we assembled were largely composed of isolates that coexisted in pairwise. And although context-dependent coexistence was common, it was mainly exhibited by a small set of isolates in high frequency. Thus, predictions of isolate persistence in complex contexts inferred from pairwise outcomes generally had high accuracy. Interestingly, ranked prevalence between pairwise and complex contexts were strongly positively correlated, nicely demonstrating the consistencies we observed across community types. These results can help guide efforts to predict or design the composition of simple microbial communities.

2.6 Detailed Methods

Bacterial isolates and reference genomes – All bacterial isolates were originally isolated from the leaves of wild or field grown *Arabidopsis thaliana* in the midwestern states of the USA (IL, IN, MI). Reference genomes were assembled as previously described (methods, chapter 1). The isolate names, taxonomic information, and assembly information are presented in Supplementary Table 2.1.

Arabidopsis leaf medium (ALM) – *Arabidopsis thaliana* (KBS-Mac-74, accession 1741) plants were grown in the University of Chicago greenhouse from January to March 2020. Seeds were densely planted in 15-cell planting trays and thinned after germination to 4-5 plants per cell. Above ground plant material was harvested just before development of inflorescence stems. Plant material was coarsely shredded by hand before adding 100g to 400mL of 10mM $MgSO_4$ and autoclaving for 55 minutes. After cooling to room temperature, the medium was filtered through 0.2μm polyethersulfone membrane filters to maintain sterility and remove plant material. The medium was stored in the dark at 4°C. Before being used for culturing, the medium was diluted 1:10 in 10mM $MgSO_4$.

Assembly and culturing of synthetic communities – Fresh bacterial stocks were prepared by first inoculating the isolates into 1mL of ALM shaking at 28°C and growing overnight. 100uL of these cultures were then used to inoculate 5mL of ALM shaking at 28°C. Once the cultures were visibly turbid, they were divided into 1mL aliquots with sterile DMSO added to a final concentration of 7% as a cryoprotectant. Stocks were stored at -80°C. To initiate an experiment, stocks were diluted to target densities determined by the target initial community titer (~$3x10^6$ cells) and the number of initial members. Diluted stocks were then combined into all pairs ("pairwise") and all *n-2* ("complex") communities in triplicate with the aid of a liquid handling robot (Freedom Evo, liquid handling robot). Pools of stocks were used to inoculate 600μL of ALM in sterile 1mL deep-well plates, in triplicate. Deep-well plates were covered with sterilized, loosely fitting plastic lids to allow air exchange. Plates were cultured in the dark at 28°C on high-speed orbital shakers capable of establishing a vortex in the deep-well plates to ensure that the cultures were well-mixed. After 24 hours, 6μL of each culture was manually transferred by multi-channel pipette into new plates containing 594μL of fresh ALM. The new plates were immediately returned to the incubator and the

day-old plates were stored at -80°C. The pairwise and complex communities were passaged for 7 days.

DNA Extraction – DNA was extracted from synthetic communities using an enzymatic digestion and bead-based purification. Cell lysis began by adding 250μL of lysozyme buffer (TE + 100mM NaCl + 1.4U/μL lysozyme) to 300μL of thawed sample and incubating at room temperature for 30 minutes. Next, 200μL of proteinase K buffer (TE + 100mM NaCl + 2% SDS + 1mg/mL proteinase K) was added. This solution was incubated at 55°C for 4 hours and mixed by inversion every 30 minutes. After extraction, the samples were cooled to room temperature before adding 220μL of 5M NaCl to precipitate the SDS. The samples were then centrifuged at 3000 RCF for 5 minutes to pellet the SDS. A liquid handling robot was used to remove 600μL of supernatant. The liquid handling robot was then used to isolate and purify the DNA using SPRI beads prepared as previously described (Rohland & Reich, 2012). Briefly, samples were incubated with 200μL of SPRI beads for 5 minutes before separation on a magnetic plate, followed by two washes of freshly prepared 70% ethanol. Samples were then resuspended in 50μL ultrapure H2O, incubated for 5 minutes, separated on a magnetic plate, and supernatant was transferred to a clean PCR plate. Purified DNA was quantified using a Picogreen assay (ThermoFisher) and diluted to 0.5ng/μL with the aid of a liquid handler.

Sequencing library preparation – Libraries were prepared using Illumina Nextera XT kits. Our protocol differed from the published protocol in two ways: 1) the tagmentation reaction was scaled down such that 1μL of purified DNA, diluted to 0.5ng/μL, was added to a solution of 1uL buffer + 0.5μL tagmentase, and 2) a KAPA HiFi PCR kit (Roche) was used to perform the amplification in place of the reagents included in the Nextera XT kit. PCR mastermix (per reaction) was composed of: 3μL 5X buffer, 0.45μL 10mM dNTPs, 1.5μL i5/i7 index adapters, respectively, 0.3μL

polymerase, and 5.75μL ultrapure H2O. The PCR protocol was performed as follows: 3 minutes at 72 °C; 13 cycles of 95 °C for 10 seconds, 55 °C for 30 seconds, 72 °C for 30 seconds; 5 minutes at 72 °C; hold at 10 °C. Sequencing libraries were manually purified by adding 15μL of SPRI beads and following the previously described approach, eluting into 12μL of ultrapure H2O. Libraries were quantified by Picogreen assay, and a subset of libraries were run on an Agilent 4200 TapeStation system to confirm that the fragment size distributions were of acceptable quality. The libraries were then diluted to a normalized concentration with the aid of a liquid handler and pooled. The pooled libraries were concentrated on a vacuum concentrator prior to size selection for a 300-600bp range on a Blue Pippin (Sage Science). The distribution of size-selected fragments was measured by TapeStation. Size-selected pool libraries were quantified by Picogreen assay and qPCR (KAPA Library Quantification Kit).

Sequencing – We characterized the compositions of our synthetic communities with a shallow metagenomics approach. Pairwise and complex synthetic communities were sequenced on a HiSeq 4000 platform. Reads were quality filtered and adapter/phiX sequences were removed using BBDuk from the BBTools suite. Reads were mapped to reference genomes using Seal (BBTools) twice, once with the "ambig" flag set to "toss" (where ambiguously mapped reads were left out) and once with the "ambig" flag set to "random" (where ambiguously mapped reads were randomly distributed to equally likely references). By comparing the results between these two strategies, we identified sets of reference genomes which resulted in high numbers of ambiguous reads (due to similarity) and corrected for such ambiguity by reallocating "tossed" reads based on proportions of unambiguous reads mapped in each sample containing a given set. To avoid mischaracterizing the composition of our synthetic communities due to contamination or non-specific mapping, for a given sample, isolates with less than 1% of total mapped reads were ignored.

Statistical analysis and data visualization – Statistical analysis and figure generation was performed in R v4.0.2 with aid from the following packages: tidyverse (Wickham et al., 2019), reshape2 (Wickham, 2007), car (Fox and Weisberg, 2018), and margins (Leeper, 2021). All scripts are provided in the supplementary materials.

2.7 Figures and tables

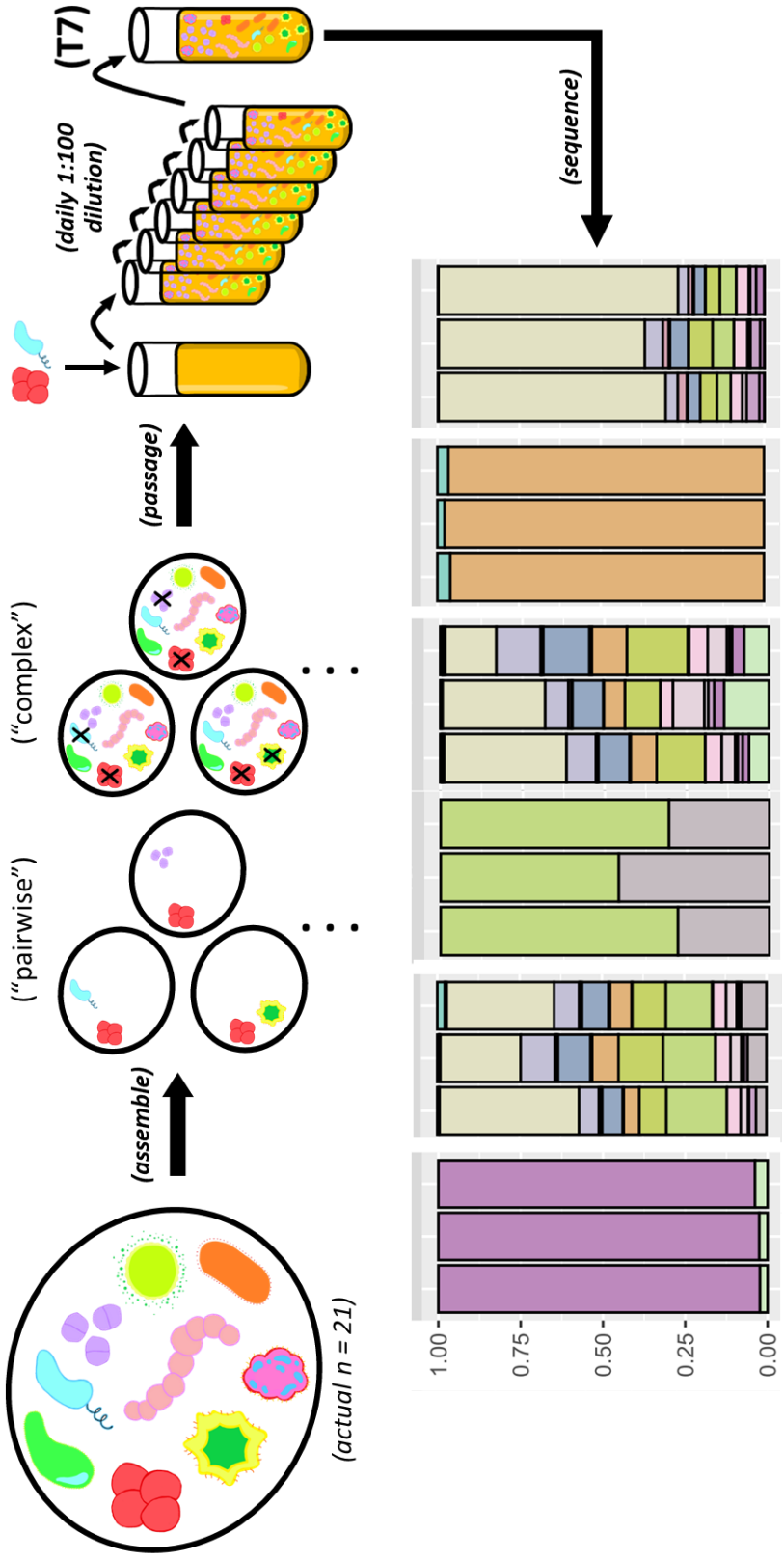


Figure 2.1 – Experimental design: A previously characterized 21-member synthetic bacterial community was decomposed into all 2-member (bottom-up) pairwise communities and all (*n*-2)-member (top-down) communities. These communities were inoculated (in triplicate) into a sterile, leaf-derived medium and passaged every 24hr (1:100 dilution) for a total of 7 periods of growth to allow the community to reach a stable composition. The composition of these samples were characterized by mapping short-reads against reference genomes. From these samples, we then assembled a set of 216 follow-up communities of intermediate complexity (2-11 initial members) to determine how well coexistence observed in the bottom-up and top-down communities aligned with coexistence in those assemblages. These communities were assembled, passaged, and sequenced as described for the communities in the initial experiment.

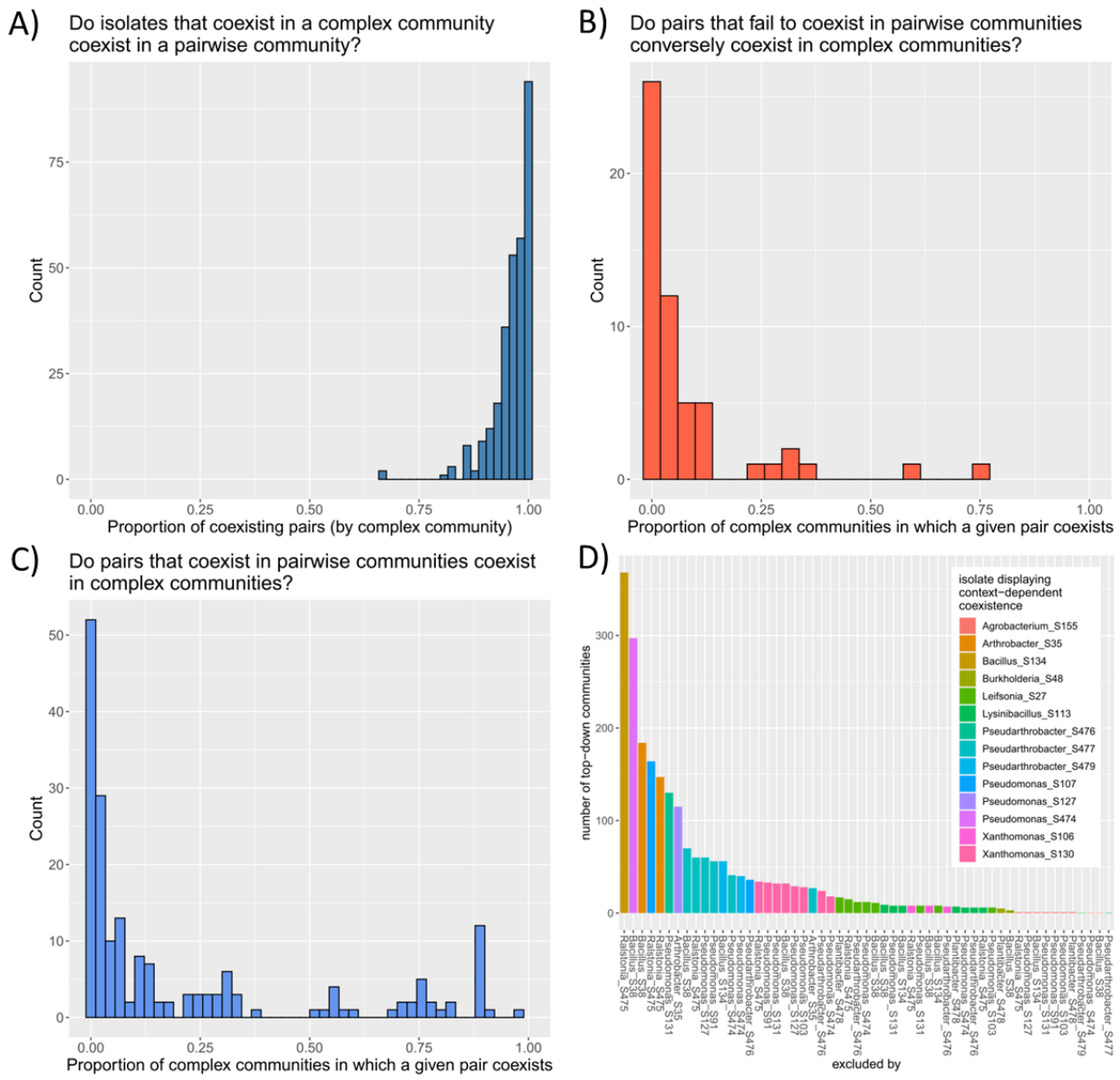


Figure 2.2 – Relating coexistence between bottom-up and top-down contexts: A) Each unique final composition of a top-down community was decomposed into all possible pairs and the proportion of those pairs that coexisted in pairwise was calculated. The distribution of these proportions is represented as a histogram. B) For each pair of isolates that did not coexist in pairwise, the proportion of top-down community contexts in which that pair coexisted was calculated. The distribution of these proportions for all such pairs is represented as a histogram. C) For each pair of isolates observed to coexist in pairwise, the proportion of top-down community contexts in which that pair coexisted was calculated. The distribution of these proportions for all such pairs is represented as a histogram. D) For each pair that failed to coexist in a bottom-up context, the number of top-down samples in which that pair was conversely observed to coexist is represented as a column in the bar plot.

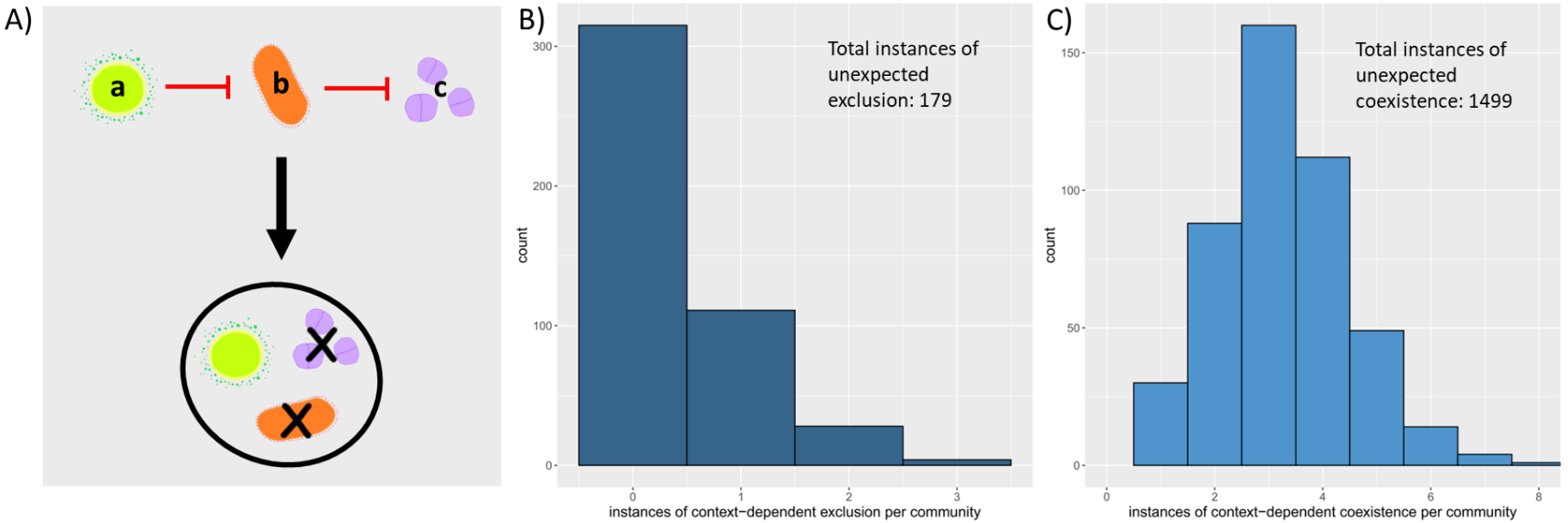


Figure 2.3 – Context-dependent exclusion was less common than context-dependent coexistence: A) Example depicting how pairwise outcomes were used to predict complex community structures. Here, isolate A excludes isolate B in pairwise, and isolate B excludes isolate C. Thus, when cultured together, only isolate A is expected to persist. B) Context-dependent exclusion was defined as an isolate failing to persist in a community context despite no isolate being present in the initial pool that excluded it in a pairwise context. The distribution of the number of instances of such exclusion per predicted community are presented as a histogram. C) Context-dependent coexistence was defined as an isolate persisting in a community context despite one or more isolates being present in the initial pool that excluded it in a pairwise context. The distribution of the number of instances of such coexistence per predicted community are presented as a histogram.

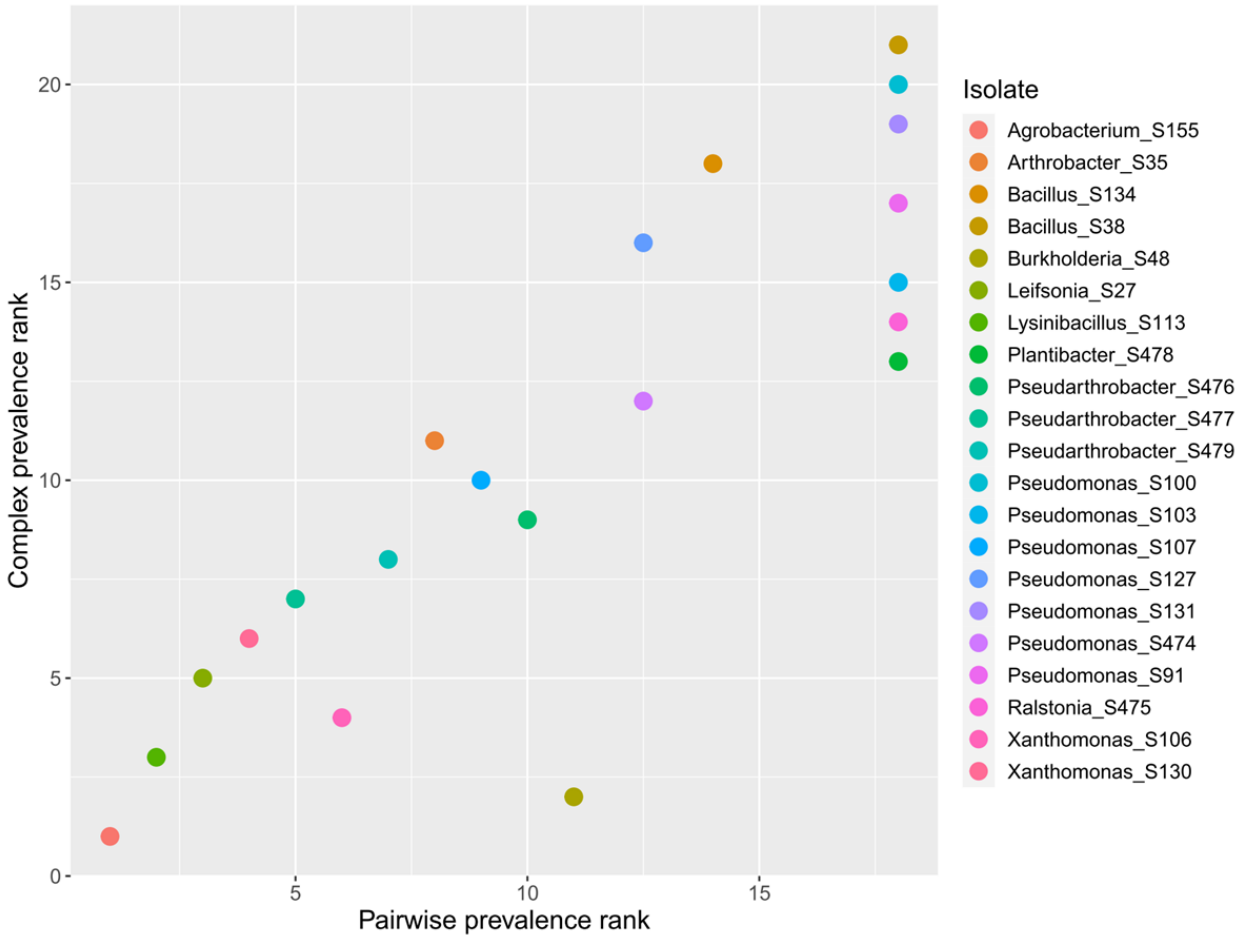
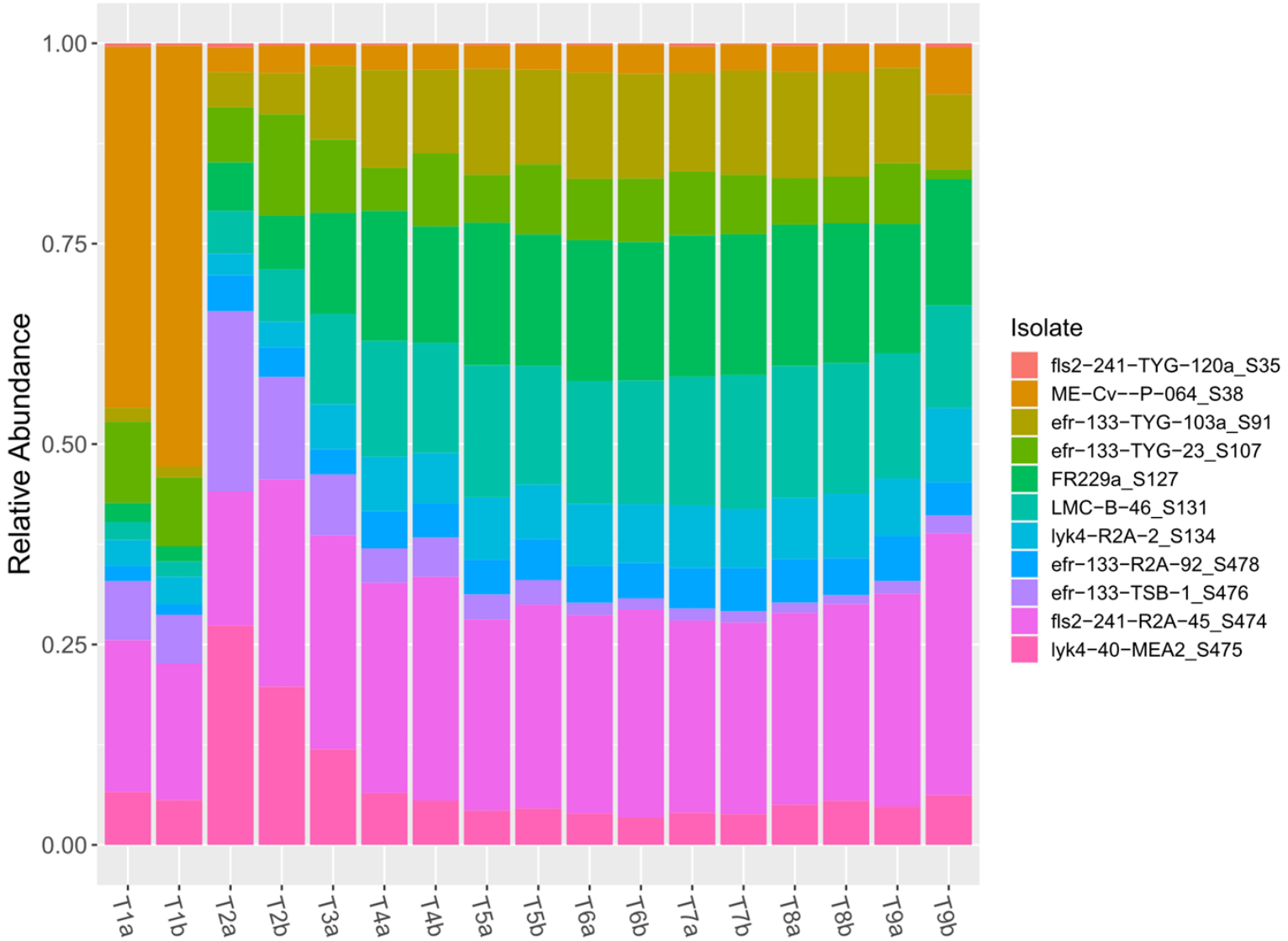


Figure 2.4 – Pairwise and complex context prevalence rank were positively correlated: Prevalence at either level was calculated as the number of communities in which an isolate persisted over the total number of communities in which that isolate was initially present.

2.8 Supplementary Figures and Tables

name	genome length (bp)	number contigs	N50	GC content	percent completion	percent redundancy	genus
lyk4_R2A_8	5.48E+06	44	2.11E+05	61.1	97.2	0.0	Pseudomonas
lyk4_R2A_2	3.64E+06	19	7.64E+05	42.0	100.0	0.0	Bacillus
fls2_241_TYG_120a	4.55E+06	128	5.92E+04	63.9	98.6	4.2	Arthrobacter
lyk4_TYG_106	3.75E+06	12	6.55E+05	69.8	98.6	4.2	Leifsonia
fls2_241_TYG_148	4.97E+06	32	2.74E+05	65.4	100.0	0.0	Xanthomonas
efr_133_TYG_23	5.68E+06	58	1.60E+05	60.7	100.0	0.0	Pseudomonas
lyk4_R2A_23	7.71E+06	60	2.02E+05	66.6	100.0	2.8	Burkholderia
fls2_241_R2A_45	5.67E+06	421	2.62E+04	61.4	100.0	0.0	Pseudomonas
fls2_241_TYG_188b	3.69E+06	7	9.21E+05	57.3	100.0	0.0	Agrobacterium
fls2_241_R2A_57	5.28E+06	229	3.52E+04	36.4	100.0	1.4	Lysinibacillus
lyk4_40_MEA_2	5.32E+06	605	1.44E+04	64.4	100.0	5.6	Ralstonia
efr_133_TYG_103a	6.12E+06	51	2.07E+05	60.3	98.6	0.0	Pseudomonas
efr_133_TSB_1	4.11E+06	1150	4.67E+03	66.0	97.2	5.6	Pseudarthrobacter
lyk4_R2A_7	4.16E+06	912	6.07E+03	66.6	95.8	5.6	Pseudarthrobacter
efr_133_R2A_92	2.78E+06	1056	3.02E+03	69.0	91.5	0.0	Plantibacter
ME(Cv)_P.064	3.53E+06	29	5.06E+05	41.6	100.0	0.0	Bacillus
LMC_B.46	6.27E+06	33	3.65E+05	60.4	98.6	0.0	Pseudomonas
LMC_A.08	5.93E+06	51	2.60E+05	59.2	98.6	1.4	Pseudomonas
ME_P.043	4.17E+06	876	6.81E+03	66.6	97.2	2.8	Pseudarthrobacter
FR229a	5.93E+06	46	2.32E+05	60.2	98.6	1.4	Pseudomonas
LMC_P.011	5.04E+06	24	3.15E+05	65.4	100.0	0.0	Xanthomonas

Supplementary Table 2.1: isolate details



Supplementary Figure 2.1 – Community assembly across passages: Communities stabilized in composition within the first few passages.

Compound	peak area				
Caffeic acid	0.000	Catechol	0.008	Aminocaproic acid	0.112
Indole-3-pyruvic acid	0.000	L-(+)-Ornithine	0.008	Valerate	0.127
Chlorogenic acid	0.000	Caffeine	0.008	Glycolic acid	0.136
Rosmarinic Acid	0.000	2-hydroxyisocaproate	0.008	Pyruvic acid	0.138
Dihydrocaffeicacid	0.000	Malonic acid	0.008	Vanillin	0.149
Indole-3-lactic acid	0.000	Asparagine	0.009	2-hydroxyhexanoic acid	0.154
Cystine	0.000	Desaminotyrosine	0.009	Phenylacetic acid	0.155
Testosterone	0.000	2-deoxy-D-Ribose	0.010	1,2-Propanediol	0.174
Coprostanol	0.000	Indole-3-propionic acid	0.010	Dulcitol	0.175
Phenyllactic acid	0.001	Glutamine	0.012	Fumarate	0.233
3,4-Dihydroxyphenylacetic acid	0.001	Cresol	0.012	Lysine	0.237
Oxaloacetic acid	0.001	2-hydroxy-3-methylbutyric acid	0.012	Ethanolamine	0.256
trans-ferulic acid	0.001	Tryptophol	0.012	Glycine	0.289
Dihydrotestosterone	0.001	Indole-3-acryclic acid	0.013	2-Hydroxybutyric acid	0.318
Cholesterol	0.001	alpha-Ketobutyrate	0.017	Threonine	0.436
3-(3-hydroxyphenyl)propionic acid	0.002	4-Hydroxyphenylacetic acid	0.017	L-Arabinose	0.488
Sinapic acid	0.002	p-Toluate	0.018	gamma-Aminobutric acid (GABA)	0.517
Alpha-Ketoglutaric acid	0.002	Indole-3-acetic acid	0.018	Benzoic acid	0.602
Glutamic acid	0.002	Tryptophan	0.021	DL-Methionine Sulfoxide	0.617
Melatonin	0.003	Imidazoleacetic acid	0.022	Cadaverine	0.637
O-Phosphorylethanolamine	0.003	Glutaric acid	0.028	D-Gluconic acid	0.695
m-Coumaric acid	0.003	3-aminoisobutyric acid	0.032	Tyrosine	0.792
Dihydroferulic acid	0.003	Histidine	0.042	4-Ethylphenol	0.840
Tryptamine	0.003	Linoleic acid	0.044	N-Acetylglucosamine (GlcNAc)	0.864
Dihydrocholesterol	0.003	Maltitol	0.046	Succinic acid	0.885
Phenylpyruvic acid	0.003	Maleic acid	0.046	Oxalate	0.895
Urocanic acid	0.003	4-methylvalerate	0.051	Citric acid	1.073
5-aminovalerate	0.004	Rhamnose	0.054	(2R,3R)-(-)-2,3-Butanediol	1.290
Imidazole Propionic acid	0.004	meso-Erythritol	0.058	Malic acid	1.370
Shikimic acid	0.004	Urea	0.063	Aspartic acid	1.423
DL-5-Hydroxylysine	0.004	Deoxycarnitine	0.066	Serine	1.759
Creatinine/Creatine	0.004	L-2-Aminobutyric acid	0.066	Lactic acid	1.793
Quinolinic acid	0.005	p-Coumaric acid	0.068	Valine	1.824
Methylsuccinic acid	0.005	Phenylalanine	0.071	Indole-3-acetamide	1.975
Hydrocinnamic acid	0.005	Indole	0.073	trans-4-Hydroxy-L-Proline	2.320
Aconitic acid	0.005	Histamine	0.073	Palmitic acid	2.452
D-(-)-Tartaric acid	0.006	Picolinic acid	0.074	Proline	3.226
(+-)-Octopamine	0.006	Dopamine	0.079	Indole-3-carboxaldehyde	3.867
Tyramine	0.006	Cysteine	0.079	Isoleucine	4.678
Quinic acid	0.006	trans-Cinnamic acid	0.081	Sucrose	8.419
Itaconate	0.007	D-Sorbitol	0.083	Alanine	12.947
Niacin	0.007	Hexanoate	0.088	Leucine	14.896
Methionine	0.007	Phenethylamine	0.089	D-Galactose	15.722

Supplementary Table 2.2: Compounds detected in Arabidopsis Leaf Medium using GC/MS profiling and a standard targeted panel of metabolites (with the Duchossois Family Institute Host-Microbe Metabolomics Facility, University of Chicago). Peak areas are included to provide semi-quantitative references for the relative amounts of the metabolites detected from the targeted panel.

2.9 References

Aranda-Díaz, Andrés, et al. "Bacterial interspecies interactions modulate pH-mediated antibiotic tolerance." *Elife* 9 (2020): e51493.

Baran, Richard, et al. "Exometabolite niche partitioning among sympatric soil bacteria." *Nature communications* 6.1 (2015): 8289.

Billick, Ian, and Ted J. Case. "Higher order interactions in ecological communities: what are they and how can they be detected?." *Ecology* 75.6 (1994): 1529-1543.

Brenner, Katie, Lingchong You, and Frances H. Arnold. "Engineering microbial consortia: a new frontier in synthetic biology." *Trends in biotechnology* 26.9 (2008): 483-489.

Brochet, Silvia, et al. "Niche partitioning facilitates coexistence of closely related honey bee gut bacteria." *Elife* 10 (2021): e68583.

Calbet, Albert, and Michael R. Landry. "Mesozooplankton influences on the microbial food web: direct and indirect trophic interactions in the oligotrophic open ocean." *Limnology and Oceanography* 44.6 (1999): 1370-1380.

Carlström, Charlotte I., et al. "Synthetic microbiota reveal priority effects and keystone strains in the Arabidopsis phyllosphere." *Nature Ecology & Evolution* 3.10 (2019): 1445-1454.

Case, Ted J., and Edward A. Bender. "Testing for higher order interactions." *The American Naturalist* 118.6 (1981): 920-929.

Chang, Chang-Yu, et al. "Emergent coexistence in multispecies microbial communities." *bioRxiv* (2022): 2022-05.

Chesson, Peter. "MECHANISMS OF MAINTENANCE OF SPECIES DIVERSITY." *Annu. Rev. Ecol. Syst* 31 (2000): 343-66.

Clark, Ryan L., et al. "Design of synthetic human gut microbiome assembly and butyrate production." *Nature communications* 12.1 (2021): 3254.

Dal Bello, Martina, et al. "Resource–diversity relationships in bacterial communities reflect the network structure of microbial metabolism." *Nature Ecology & Evolution* 5.10 (2021): 1424-1434.

Diaz-Colunga, Juan, et al. "Top-down and bottom-up cohesiveness in microbial community coalescence." *Proceedings of the National Academy of Sciences* 119.6 (2022): e2111261119.

Elias, Sivan, and Ehud Banin. "Multi-species biofilms: living with friendly neighbors." *FEMS microbiology reviews* 36.5 (2012): 990-1004.

Fox, John, and Sanford Weisberg. *An R companion to applied regression*. Sage publications, 2018.

Foster, Kevin R., and Thomas Bell. "Competition, not cooperation, dominates interactions among culturable microbial species." *Current biology* 22.19 (2012): 1845-1850.

Friedman, Jonathan, Logan M. Higgins, and Jeff Gore. "Community structure follows simple assembly rules in microbial microcosms." *Nature ecology & evolution* 1.5 (2017): 0109.

Gause, G. F. The Struggle for Existence (Williams & Wilkins, 1934).

Gralka, Matti, et al. "Trophic interactions and the drivers of microbial community assembly." *Current Biology* 30.19 (2020): R1176-R1188.

Higgins, Logan M., et al. "Co-occurring soil bacteria exhibit a robust competitive hierarchy and lack of non-transitive interactions." *BioRxiv* (2017): 175737.

Lax, Simon, and Jeff Gore. "Strong Ethanol-and Frequency-Dependent Ecological Interactions in a Community of Wine-Fermenting Yeasts." *bioRxiv* (2022): 2022-09.

Lee, Kai Wei Kelvin, et al. "Biofilm development and enhanced stress resistance of a model, mixed-species community biofilm." *The ISME journal* 8.4 (2014): 894-907.

Leeper TJ (2021). *margins: Marginal Effects for Model Objects*. R package version 0.3.26.

Liu, Yong-Xin, Yuan Qin, and Yang Bai. "Reductionist synthetic community approaches in root microbiome research." *Current Opinion in Microbiology* 49 (2019): 97-102.

May, Robert M. "Will a large complex system be stable?." *Nature* 238 (1972): 413-414.

Maynard, Daniel S., Zachary R. Miller, and Stefano Allesina. "Predicting coexistence in experimental ecological communities." *Nature ecology & evolution* 4.1 (2020): 91-100.

Meroz, Nittay, et al. "Community composition of microbial microcosms follows simple assembly rules at evolutionary timescales." *Nature communications* 12.1 (2021): 2891.

Mickalide, Harry, and Seppe Kuehn. "Higher-order interaction between species inhibits bacterial invasion of a phototroph-predator microbial community." *Cell systems* 9.6 (2019): 521-533.

Ortiz, Anthony, et al. "Interspecies bacterial competition regulates community assembly in the C. elegans intestine." *The ISME Journal* 15.7 (2021): 2131-2145.

Paine, Robert T. "Food web complexity and species diversity." *The American Naturalist* 100.910 (1966): 65-75.

Pascual-García, Alberto, Sebastian Bonhoeffer, and Thomas Bell. "Metabolically cohesive microbial consortia and ecosystem functioning." *Philosophical Transactions of the Royal Society B* 375.1798 (2020): 20190245.

Pomerantz, Mark J. "Do" Higher Order Interactions" in Competition Systems Really Exist?." *The American Naturalist* 117.4 (1981): 583-591.

Schäfer, Martin, et al. "Mapping phyllosphere microbiota interactions in planta to establish genotype–phenotype relationships." *Nature microbiology* 7.6 (2022): 856-867.

Sundarraman, Deepika, et al. "Higher-order interactions dampen pairwise competition in the zebrafish gut microbiome." *MBio* 11.5 (2020): e01667-20.

Ratzke, Christoph, Julien Barrere, and Jeff Gore. "Strength of species interactions determines biodiversity and stability in microbial communities." *Nature ecology & evolution* 4.3 (2020): 376-383.

Richmond, Rollin C., et al. "A search for emergent competitive phenomena: the dynamics of multispecies Drosophila systems." *Ecology* 56.3 (1975): 709-714.

Rivett, Damian W., et al. "Elevated success of multispecies bacterial invasions impacts community composition during ecological succession." *Ecology Letters* 21.4 (2018): 516-524.

Rohland, Nadin, and David Reich. "Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture." *Genome research* 22.5 (2012): 939-946.

Romdhane, Sana, et al. "Unraveling negative biotic interactions determining soil microbial community assembly and functioning." *The ISME Journal* 16.1 (2022): 296-306.

Venturelli, Ophelia S., et al. "Deciphering microbial interactions in synthetic human gut microbiome communities." *Molecular systems biology* 14.6 (2018): e8157.

Wickham, Hadley. "Reshaping data with the reshape package." *Journal of statistical software* 21 (2007): 1-20.

Wickham, Hadley, et al. "Welcome to the Tidyverse." *Journal of open source software* 4.43 (2019): 1686.

Wootton, J. Timothy. "Indirect effects and habitat use in an intertidal community: interaction chains and interaction modifications." *The American Naturalist* 141.1 (1993): 71-89.

Wootton, J. Timothy. "The nature and consequences of indirect effects in ecological communities." *Annual review of ecology and systematics* 25.1 (1994): 443-466.

Chapter 3: Invasion timing affects invasion outcome in synthetic bacterial communities

3.1 Abstract

Microbial communities regularly experience ecological invasions that can change their composition and function. And some factors thought to affect the outcome of such invasions can vary over the course of community assembly. Here, we use synthetic bacterial communities to evaluate how the timing of an invasion, relative to the community assembly process, affects its success and impact on the invaded community. We invaded 15 distinct communities with 3 bacterial invaders at the initial assembly of the community ("initial"), 24 hours into community dynamics ("early") and 7 days into community dynamics ("late"). Communities were passaged every 24 hours and characterized through shallow short-read sequencing once reaching a stable composition. We observed that communities with lower richness or resource use efficiency were less robust against invasion. We also observed that even unsuccessful invasion could have large impacts on invaded communities, most strongly affecting the most abundant communities members. Invasions were most successful and had the largest effect on community composition in the "early" treatment. This result was statistically associated with invader growth on spent media, suggesting it was related to changes in resource use efficiency over the course of community assembly. Our results demonstrate that invasion timing can affect the outcome of that invasion, a finding relevant to efforts to adjust or maintain the composition of microbial communities.

3.2 Introduction

Microbial communities are ubiquitous and of great importance to human health (Clark et al. 2021), agriculture (Liu et al., 2019), and industry (Zhou et al., 2015). As such, many efforts are underway to design or modify microbial communities that assume a desired composition or perform a desired function. But, like all ecological communities, microbial communities are exposed to environmental fluctuations and migration that can affect community composition and function (Burns et al., 2017; Nguyen et al., 2021). This represents a challenge for our efforts to manipulate microbial communities to serve our own ends.

Take, for example, efforts to design a microbial community that performs a desired function. As has often been observed, such a community might perform as designed *in vitro* or in a host under well controlled conditions but shift in composition and function when exposed to the complexity of a natural environment (Burns et al., 2017; Morella et al., 2020; Bergelson et al., 2021). Or, as an inverse example, consider the engraftment of a probiotic bacterium in the gut microbiome. The potential positive effect of that probiotic on host health is irrelevant if it cannot persist in the host microbiome (Santos et al., 2006; Robinson et al., 2010). Both scenarios deal with the complexities of ecological invasions.

Invasion ecology concerns the establishment and impact of outside species on ecological communities. A primary factor that has been demonstrated to affect the outcome of ecological invasions is the relationship between diversity and resource use efficiency (Stachowicz et al., 1997; Tilman, 1999; Hodgson et al., 2002). This relationship suggests that a successful invasion is more likely to occur when an invader has greater access to resources, and that a positive relationship between diversity and resource use efficiency deprives invaders of such access. Previous work has also related invasion timing and outcome as a function of change in resource availability, namely,

through synchronization of invasion with periods of decreased community productivity (Li and Stevens, 2012: Mallon et al., 2015). However, communities arise from a dynamic assembly process, during which composition (diversity) and function (resource use) are in flux. Thus, we posit that the timing of an invasion, relative to the community assembly process, is likely relevant to its outcome.

Here, we used synthetic bacterial communities to ask the question: as community assembly progresses towards a stable state, how does the timing of an invasion affect its outcome? We consider the "outcome" of an invasion as both the success/failure of the invader to persist in a community over time and the effect of invasion on the composition of the invaded community.

3.3 General Methods

Experimental design and assembly of synthetic communities – To study invasion across multiple community contexts, we assembled a set of 15 synthetic bacterial communities from a pool of 48 bacterial strains, representing 24 genera, originally isolated from the leaves of wild and field-grown *Arabidopsis thaliana* (supplementary table 3.1). These communities ranged in initial richness from 8 to 48 strains. We inoculated these pools into a plant-based medium derived from the leaves of greenhouse grown *A. thaliana* (methods). Communities were inoculated at an initially consistent total cell density, with each member at an equal density reduced in proportion to the richness of the pool. To ensure we analyzed these communities in a state reflective of their long-term compositions, we passaged each community into fresh medium (1:100 dilution) every 24 hours for 7 days (figure 3.1). We have previously demonstrated that such a period is sufficient to allow communities in this system to reach an ecologically stable state (chapter 1).

To test the effect of invasion timing on community assembly, we invaded each of these 15 communities with three different bacterial invaders at each of three time points in the community assembly process (figure 3.1). We chose to work with multiple community contexts and invaders to seek general patterns in the effect of invasion timing, rather than outcomes specific to a certain invader and/or community context. We used a strain of *Pseudomonas poae* (Pseudomonas_MEJ082), a strain of *Pseudomonas viridiflava* (Pseudomonas_RMX31.b), and a strain of *Xanthomonas campestris* (Xanthomonas_S130) as invaders. Each invader was added at a density determined in preliminary experiments to be sufficient to lead to successful invasion in some, but not all, community contexts (methods). These densities were 0.1% of estimated total community density for *X. campestris* and 10% for both *P. poae* and *P. viridiflava*.

We assessed three invasion timing treatments that targeted distinct phases of the community dynamics. In the "initial" invasion treatment, invaders were added to the community during initial community assembly (T0), when the assembly process has just begun. In the "early" invasion treatment, invaders were added immediately after the first round of passaging (24 hours, T1), which is an especially dynamic point in the assembly process (chapter 1). And in the "late" invasion treatment, invaders were added after 7 days of growth (entailing 6 rounds of passaging, T7), when community dynamics had reached a steady state. After adding invaders, communities were passaged for an additional 7 days. We used NGS to characterize the composition of these communities; specifically, by mapping short reads back to previously assembled reference genomes (methods).

3.4 Results

Invasions were commonly unsuccessful, and more so in high richness communities – We define a successful invasion as one where an invader persists in the community in which they invade. In general,

successful invasions were uncommon. Across all invaders and all invaded communities, successful invasion was only observed ~24% of the time. Some invaders were more successful than others, while some invaded communities were more likely to be invaded. Specifically, there was a statistical association between invader identity and invasion success (Chi-square test of independence, p-value $< 2e^{-9}$), with *X. campestris* displaying the highest success rate at 39% successful invasions and *P. viridiflava* showing the lowest at 6% (table 3.1, supplementary figure 3.1a). Interestingly, these two invaders showed an opposite relationship in relative abundances, with the relative abundance of *P. viridiflava* in the successfully invaded communities reaching the highest levels (±SD) at 0.27±0.12 and *X. campestris* showing the lowest at 0.03±0.02.

We also observed a statistical association between invaded community and invasion success, with communities 3,4 and 6 experiencing notably high invasion success rates (table 3.1, supplementary figure 3.1b). Given that our communities varied in richness across invasion treatments, we used logistic regression to analyze the relationship between invasion success or failure and average richness of an invaded community prior to invasion (supplementary table 3.2). We found that successful invasion was less likely as community richness increased, but with a low average marginal effect, with an increase in richness of one being associated with a 3.3% reduction in the probability of an invasion attempt being successful (figure 3.2a).

Resource use efficiency was associated with invasion success – If invaders are more successful when invading communities with low resource use efficiency and/or with an abundance of metabolites suitable for cross-feeding, then we would expect invasion success to be related to an invader's growth on the nutrients available in each invaded community. To examine this possibility, we evaluated how well invaders were able to grow on "spent media" of the uninvaded communities (methods), and whether growth on spent media was predictive of invasion outcome. Briefly, we filtered each of the

communities after one day and seven days of growth (representing the communities immediately prior to the "early" invasion and "late" invasion treatments, respectively) through 0.2μm polytetrafluoroethylene filters to remove bacterial cells and isolate sterile "spent" medium. We then assessed the growth of each invader by culturing each of the three invader species in each spent medium (amended with M9 salts, final concentration 0.3X) and measuring OD600 after 48 hours.

Extensive growth was rare and generally restricted to *P. poae* (supplementary figure 3.2), which was the invader with an intermediate level of invasion success at 27%. We used logistic regression to analyze the relationship between invader growth on spent media and invasion outcome and found that *P. poae* displayed a significant, positive relationship (supplementary table 3.2). For that invader, we observed that growth on spent media was a significant predictor of invasion outcome, with a 0.1 increase in optical density associated with a ~40% average increase in the probability of a successful invasion (supplementary table 3.2, figure 3.2b).

We observed no difference in growth between the filtrates from day 1 versus day 7 for any of the invaders (two sample t-tests, all p-values > 0.4). However, to assess if community resource use efficiency may have changed over the course of community assembly, we compared the initial (T1) and late stage (T6) densities (OD600) of each uninvaded community. Using logistic regression, we observed a significant relationship between the change in density and invasion success, with an increase in density of 0.1 associated with a 7% average decrease in the probability of invasion success (supplementary table 3.2, figure 3.2c).

Invaders had an outsized effect on the most abundant members of invaded communities – We next asked if there were any consistent patterns in how community compositions shifted after invasion. For each member detected in a given community prior to invasion, we related its pre-invasion rank abundance to the log-ratio of its relative abundances in the post-invasion and pre-invasion contexts

(disregarding isolates not detected after invasion). We observed a significant difference between ranks (one-way ANOVA: $F_{8,532} = 9.37$, p-value $< 4e^{-11}$), with post-hoc testing (Tukey's honest significance test) identifying the most abundant isolates as significantly different from all other ranks (figure 3.3a). Interestingly, only the highest rank abundance community members showed a significant decrease in relative abundance (one-sample t-test: p-value $< 2.2e^{-16}$). In some cases, in which an invader persisted, the change in community composition was largely due to the presence of the invader (figure 3.3b). However, there were also instances where, despite occupying a low relative abundance (figure 3.3c) or even failing to invade (figure 3.3d), an invader had a large effect on community composition.

Invasion timing affected invasion outcome but can be explained by invader growth on spent media – To assess the effect of invasion timing on invasion outcome, we first compared the invasion success rate of each invasion treatment. A Chi-square test of independence demonstrated there was a significant association between invasion treatment and invasion outcome (success/failure) (p-value $< 6e^{-5}$), with successful invasions more common than expected in the early invasion treatment and less common than expected in the initial invasion treatment (figure 3.4a) when all invaders were considered jointly. To further test if invasion timing affected invasion outcome, we measured the effect of each treatment on the composition of invaded communities. Specifically, for a given invasion treatment, we calculated the average Bray-Curtis dissimilarity between samples of that treatment and uninvaded control communities, for each of the 15 communities separately and disregarding an invader if it was present (figure 3.4b). We then used pairwise single-factor PERMANOVA tests to determine if the invasion treatments resulted in community compositions distinct from the uninvaded controls (methods). We observed that all invasion timing treatments resulted in significantly distinct community compositions, regardless of the success of the invasion (all p-values ≤ 0.006 after multiple testing correction).

Pursuing these results further, we sought to determine if there were differences in the extent to which invasion timing treatments affected community composition. Thus, we compared the Bray-Curtis dissimilarities of each treatment (relative to the uninvaded controls) to each other. A one-way ANOVA found that the average dissimilarity from the uninvaded controls differed between the invasion treatments ($F_{2,384}$ = 5.02, p-value 0.007, supplementary table 3.3). Post-hoc tests showed strong support for a significant difference in dissimilarity between the early and initial invasion treatments with an average difference of 0.08 (Tukey's honest significant test: 95% CI [0.02, 0.14], p-value = 0.006, supplementary table 3.3). There was also marginal statistical support for a difference between the early and late invasion treatments (Tukey's honest significant test: 95% CI [-0.007, 0.11], p-value = 0.092, supplementary table 3.3).

Given that we had previously identified associations between invasion outcome and community richness, invader identity, and invader growth on spent media, we performed additional analyses including these variables in the model (supplementary table 3.4). Neither richness nor invader identity had significant effects in the model or on the effect of invasion timing. However, invader growth on spent media entered the model as a significant covariate and, importantly, reduced the effect of invasion timing. Note, however, that this model only considered the "early" and "late" invasion treatments, as those are the only treatments for which we could measure growth on spent media prior to invasion.

3.5 Discussion

We hypothesized that invasion timing would be relevant to invasion outcome because factors affecting ecological invasions have the potential to change over the course of the community assembly process. Community diversity is one such dynamic factor. Indeed, the relationship between

diversity and community invasibility has long been studied in plant communities (Peart and Foin, 1985; Tilman 1999; Kennedy et al., 2002; Fridley et al., 2007) as well as experimental bacterial communities (Hodgson et al., 2002; Eisenhauer et al., 2012) and in the context of enteric pathogens (Dillon et al., 2005; Britton and Young, 2014). It has been demonstrated that the relationship between diversity and invasibility is related to the relationship between diversity and resource use efficiency; namely, that increased diversity can result in more complete occupancy of available niche space, thus increasing community resource use efficiency and reducing the availability of resources an invader needs to successfully invade (Tilman, 1999; Loreau and Hector, 2001; Hodgson et al., 2002). Our results are in general alignment with this effect, as we observed that higher richness (more diverse) communities were less likely to be successfully invaded (figure 3.2a). We also observed that *P. poae* growth on spent medium was positively associated with invasion success (figure 3.2b), suggesting communities that left higher levels of remaining resources after 24 hours of growth were more likely to be invaded by that invader. And finally, in our analysis of the effect of invasion timing on compositional divergence relative to uninvaded controls, the effect of invasion timing was reduced when we incorporated invader growth on spent media (supplementary table 3.4), suggesting that differences in the vulnerability of communities to invasion over time were, at least in part, modulated by available resources.

Another relevant factor is community composition, which inherently changes as community members are filtered out during assembly. This is relevant to invasibility in that invaders can be excluded if they compete with species that share similar nutrient requirements (Innerebner et al., 2011; Yang et al., 2017). This mechanism, referred to as a "sampling effect", is also related to richness, as higher richness increases the chance that a community contains species capable of excluding an invader through competition (Tilman, 1999; Mallon et al., 2015; Wei et al., 2015). Thus, as community composition changes over the course of the assembly process, invasibility, as

mediated through this mechanism, will vary. And, in particular, communities should become more vulnerable to invasion as they settle into stable communities of reduced complexity.

We can use the effects described above to contextualize the differences we observed between invasion timing treatments. First, we observed that successful invasions were under-represented in the initial invasion treatment (figure 3.4a). This may have resulted from a "sampling effect", where the increased richness encountered by invaders in the initial invasion treatment (as no competitive exclusion could yet have occurred) increased the chances that there was a community member present which could exclude the invader, thus decreasing the chance of successful invasion.

Further, when comparing the impact of each treatment on community composition, we observed the greatest differences between the early and the initial/late invasions treatments (supplementary table 3.3), suggesting the late invasion treatment (like the initial invasion treatment) was relatively robust to invasion as compared to the early treatment. We demonstrated that this treatment effect was associated with invader growth on spent media (supplementary table 3.4) and may have resulted from increased community resource use efficiency later in the assembly process (figure 3.2c). Resource use efficiency might increase over time because communities early in the assembly process can exhibit non-optimal resource use due to high levels of early competition (Rivett et al., 2016; Jones et al., 2017), but achieve higher resource use efficiency once reaching a stable state (Bittleston et al, 2020), through mechanisms such as changes in metabolic regulation or evolution that result in community members better able to utilize available resources (Lawrence et al., 2012; Fiegna et al., 2015).

Previous work investigating the importance of invasion timing on invasion outcome has focused on the synchronization of invasions with periods of increased resource availability (Li and Stevens, 2012: Mallon et al., 2015). This is also related to work investigating the relationship between

disturbance and invasibility, which has posited that disruptions in community resource use efficiency due to disturbance provide a greater opportunity for invasion (Burke and Grime, 1996; Clark and Johnston, 2011; Symons and Arnott, 2014). Our results are in general alignment with these perspectives, as they rely on periods of increased relative resource availability as predictors of invasion outcome.

Consistent with our results, Rivett et al. (2018) reported decreased success of invasions later in the assembly process and identified change in resource availability across community assembly as a mechanism underlying invasion outcome. In that study, however, assembly occurred in a static environment with no nutrient replenishment. Our method of assembly through passaging represents a distinct assembly process akin to environments with higher sustained metabolic activity resulting from periodic influxes of resources (e.g., the gut environment). Despite the difference in nutrient dynamics between our two systems, the convergence of our results suggests the relationship between invasion timing and outcome is robust across different environments.

An interesting finding of this work is that even unsuccessful invasions could have large effects on the final composition of the communities from which they were excluded (figure 3.3d). Transient invasions were observed to have profound impacts on the composition of simple synthetic bacterial communities (Amor et al., 2020). More generally, it has been shown that priority effects (Martinez et al., 2018; Carlström et al., 2019) and minor differences in transient states early in the assembly of bacterial communities can lead to divergent final community compositions (Bittleston et al., 2020). These results highlight the importance of transient community states during the assembly process and can help us understand why the initial and early invasion treatments led to such distinct community compositions (supplementary table 3.3). Namely, the community compositions at T0

and T1 were sufficiently distinct that it is unsurprising disruption via invasion would drive distinct paths of further assembly.

Overall, we show that invasion of a synthetic bacterial community at different points of the community assembly process affects invasion success and the impact of the invaders on the resident community. While differences in invaders were identified, we found rather consistent increases in the rates and impact of invasion during the community assembly process that were consistent with the effects of community composition, and resource use efficiency. These results further our understanding of the factors affecting the invasibility of microbial communities, with implications relevant to human health, agriculture, and industry.

3.6 Detailed Methods

Bacterial isolates and reference genomes – All bacterial isolates were originally isolated from the leaves of wild or field grown *Arabidopsis thaliana* in the midwestern states of the USA (IL, IN, MI). Reference genomes were assembled as previously described (methods, chapter 1). The isolate names, taxonomic information, and assembly information are presented in Supplementary Table 3.1.

Arabidopsis leaf medium (ALM) – *Arabidopsis thaliana* (KBS-Mac-74, accession 1741) plants were grown in the University of Chicago greenhouse from January to March 2020. Seeds were densely planted in 15-cell planting trays and thinned after germination to 4-5 plants per cell. Above ground plant material was harvested just before development of inflorescence stems. Plant material was coarsely shredded by hand before adding 100g to 400mL of 10mM $MgSO_4$ and autoclaving for 55 minutes. After cooling to room temperature, the medium was filtered through 0.2μm polyethersulfone membrane filters to maintain sterility and remove plant material. The medium was

stored in the dark at 4°C. Before being used for culturing, the medium was diluted 1:10 in 10mM $MgSO_4$.

Assembly and culturing of synthetic communities – Fresh bacterial stocks were prepared by first inoculating the isolates into 1mL of ALM shaking at 28°C and growing overnight. 100uL of these cultures were then used to inoculate 5mL of ALM shaking at 28°C. Once the cultures were visibly turbid, they were divided into 1mL aliquots with sterile DMSO added to a final concentration of 7% as a cryoprotectant. Stocks were stored at -80°C. To initiate an experiment, stocks were diluted to target densities determined by the target initial community titer (~$1x10^6$ cells) and the number of initial members. Diluted stocks were then combined into 15 unique pools ranging in initial richness from 8-48 isolates. These pools were used to inoculate 600μL of ALM in sterile 1mL deep-well plates, in triplicate. Deep-well plates were covered with sterilized, loosely fitting plastic lids to allow air exchange. Plates were cultured in the dark at 28°C on high-speed orbital shakers capable of establishing a vortex in the deep-well plates to ensure that the cultures were well-mixed. After 24 hours, 6μL of each culture was manually transferred by multi-channel pipette into new plates containing 594μL of fresh ALM. The new plates were immediately returned to the incubator and the day-old plates were stored at -80°C.

Determining appropriate invading densities for invaders – Given that we aimed to assess the effect of invaders on the composition of invaded communities, we wanted to avoid instances where invader density was too low to impact a community or so high such that an invader would dominate every invaded community. To identify an appropriate invading density for each isolate, 8 of the communities (#8-15) were invaded with each invader across a range of initial densities (0.01%, 0.1%, 1%, 10%, 25%, 100% of estimated invaded community density). These cultures were passaged as described for the main experiment, and invader presence was tracked over time by spot plating 20μL

of each culture from each timepoint onto 1X TSA plates containing 70μg/mL gentamicin. The invader isolates had been previously transformed to contain gentamicin-resistance cassettes through mini-Tn7 insertion, allowing us to track the presence of the invaders over time. In this way, we identified invader densities which produced a variety of invasion outcomes (i.e., failure to establish versus persistence at variable abundances).

Spent media assays – We performed spent media assays that relate to the "early" and "late" invasion treatments. Specifically, to assess the nutrient environment invaders encountered in the "early" treatment, we isolated spent medium from uninvaded communities after 24 hours of growth (immediately prior to passaging and addition of invaders). For the "late" treatment, we isolated spent medium from uninvaded communities after 7 days of growth (immediately prior to the 7th passage). We isolated spent medium from each community by pelleting the bacterial cells (centrifuged for 10 minutes at 3000 RCF) and filtering ~150μL of supernatant through 0.2μm polytetrafluoroethylene filtration plates (Pall Corporation). The filtrate was then pooled by community (to homogenize variation among replicates and produce a representative spent medium for a given community) and amended with M9 salts (at a final concentration of 0.3X). Prior to inoculation, invader stocks were pelleted and washed in 10mM $MgSO_4$ twice to minimize media carryover from the stocks. Each invader was subsequently resuspended in 10mM $MgSO_4$, and 5μL was inoculated into 200μL of each spent medium in triplicate and cultured statically at 28°C in 96-well clear bottom plates. Negative controls were present in each plate, containing only 10mM $MgSO_4$ buffer and M9 salts (0.3X). These negative controls were used to subtract background growth from the invaders cultured in spent medium. Growth was assessed by optical density (OD600) after 24 and 48 hours.

DNA Extraction – DNA was extracted from synthetic communities using an enzymatic digestion and bead-based purification. Cell lysis began by adding 250μL of lysozyme buffer (TE + 100mM

NaCl + 1.4U/μL lysozyme) to 300μL of thawed sample and incubating at room temperature for 30 minutes. Next, 200μL of proteinase K buffer (TE + 100mM NaCl + 2% SDS + 1mg/mL proteinase K) was added. This solution was incubated at 55°C for 4 hours and mixed by inversion every 30 minutes. After extraction, the samples were cooled to room temperature before adding 220μL of 5M NaCl to precipitate the SDS. The samples were then centrifuged at 3000 RCF for 5 minutes to pellet the SDS. A Tecan Freedom Evo liquid handler was used to remove 600μL of supernatant. The liquid handler was then used to isolate and purify the DNA using SPRI beads prepared as previously described (Rohland & Reich, 2012). Briefly, samples were incubated with 200μL of SPRI beads for 5 minutes before separation on a magnetic plate, followed by two washes of freshly prepared 70% ethanol. Samples were then resuspended in 50μL ultrapure H2O, incubated for 5 minutes, separated on a magnetic plate, and supernatant was transferred to a clean PCR plate. Purified DNA was quantified using a Picogreen assay (ThermoFisher) and diluted to 0.5ng/μL with the aid of a liquid handler.

Sequencing library preparation – Libraries were prepared using Illumina Nextera XT kits. Our protocol differed from the published protocol in two ways: 1) the tagmentation reaction was scaled down such that 1μL of purified DNA, diluted to 0.5ng/μL, was added to a solution of 1uL buffer + 0.5μL tagmentase, and 2) a KAPA HiFi PCR kit (Roche) was used to perform the amplification in place of the reagents included in the Nextera XT kit. PCR mastermix (per reaction) was composed of: 3μL 5X buffer, 0.45μL 10mM dNTPs, 1.5μL i5/i7 index adapters, respectively, 0.3μL polymerase, and 5.75μL ultrapure H2O. The PCR protocol was performed as follows: 3 minutes at 72°C; 13 cycles of 95°C for 10 seconds, 55°C for 30 seconds, 72°C for 30 seconds; 5 minutes at 72°C; hold at 10°C. Sequencing libraries were manually purified by adding 15μL of SPRI beads and following the previously described approach, eluting into 12μL of ultrapure H2O. Libraries were quantified by Picogreen assay, and a subset of libraries were run on an Agilent 4200 TapeStation

system to confirm that the fragment size distributions were of acceptable quality. The libraries were then diluted to a normalized concentration with the aid of a liquid handler and pooled. The pooled libraries were concentrated on a vacuum concentrator prior to size selection for a 300-600bp range on a Blue Pippin (Sage Science). The distribution of size-selected fragments was measured by TapeStation. Size-selected pool libraries were quantified by Picogreen assay and qPCR (KAPA Library Quantification Kit).

Sequencing – We characterized the compositions of our synthetic communities with a shallow metagenomics approach. Samples were sequenced on a NovaSeq 6000 platform. Reads were quality filtered and adapter/phiX sequences were removed using BBDuk from the BBTools suite. Reads were mapped to reference genomes using Seal (BBTools) twice, once with the "ambig" flag set to "toss" (where ambiguously mapped reads were left out) and once with the "ambig" flag set to "random" (where ambiguously mapped reads were randomly distributed to equally likely references). By comparing the results between these two strategies, we identified sets of reference genomes which resulted in high numbers of ambiguous reads (due to similarity) and corrected for such ambiguity by reallocating "tossed" reads based on proportions of unambiguous reads mapped in each sample containing a given set. To avoid mischaracterizing the composition of our synthetic communities due to contamination or non-specific mapping, for a given sample, isolates with less than 1% of total mapped reads were ignored.

PERMANOVA analysis – Single-factor PERMANOVA tests were used to determine if the shifts in community composition resulting from the invasion timing treatments were distinct from the uninvaded control communities. Tests were performed using the "adonis2" function from the R package "vegan" (v2.6-4, Oksanen et al., 2013). Bray-Curtis dissimilarity was used to measure the

compositional effect of a given treatment. All tests were performed with 999 permutations and the permutations were blocked by community identity.

Statistical analysis and data visualization – Statistical analysis and figure generation was performed in R v4.0.2 with aid from the following packages: tidyverse (Wickham et al., 2019), reshape2 (Wickham, 2007), car (Fox and Weisberg, 2018), vcd (Meyer et al., 2020), and vegan (Oksanen et al., 2013). All scripts are provided in the supplementary materials. Figures were made using Microsoft Paint 3D and PowerPoint.

3.7 Figures and Tables

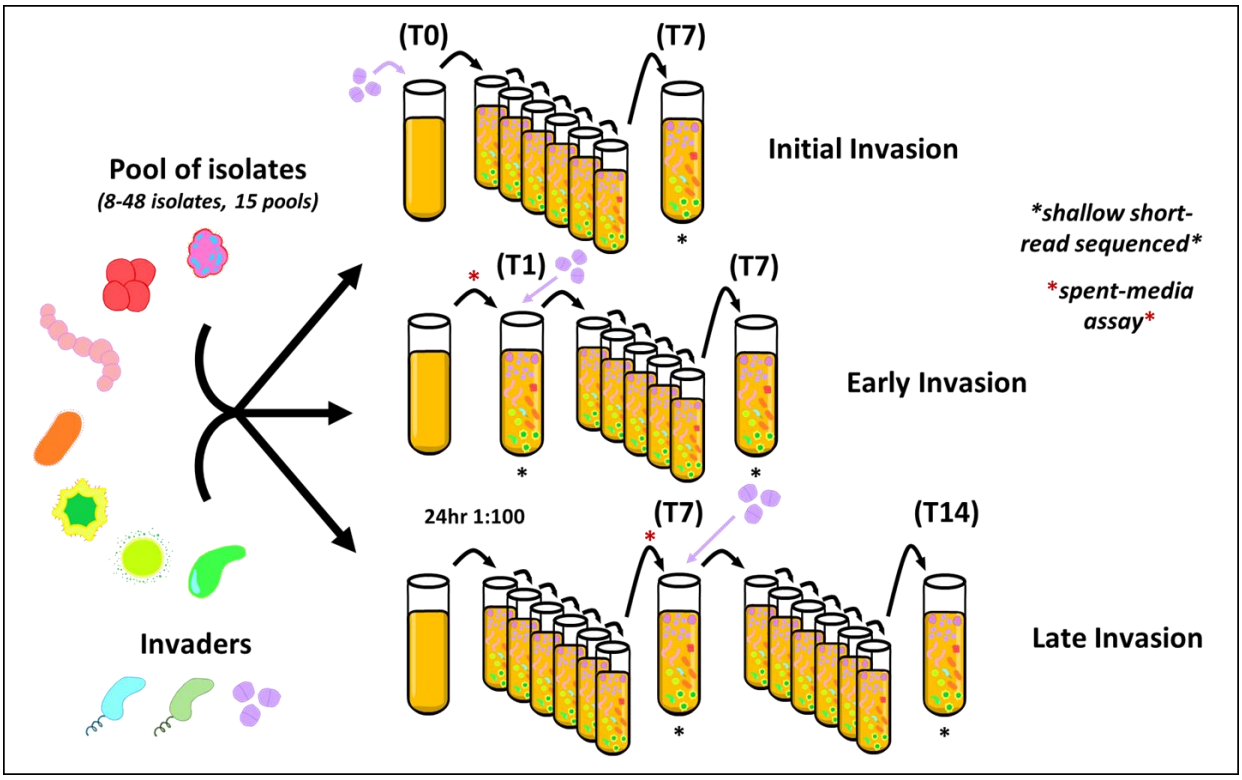


Figure 3.1 – Experimental design: 15 pools of bacterial isolates were separately subjected to invasion by three bacterial isolates across three timing treatments. Across all treatments, passaging always occurred after 24 hours and involved a 1:100 dilution into fresh media. In the "initial invasion" treatment, a given invader was added alongside the other community members at the time of community initiation (T0). In the "early invasion" treatment, a given community was assembled and passaged after 24 hours (T1), before an invader was added immediately after passaging. And in the "late invasion" treatment, a given community was assembled and passaged for 7 days (T7) before an invader was added. Post-invasion, communities were passaged for seven days and characterized using shallow short-read sequencing. Sequenced samples are indicated by (*) while samples filtered for the spent-media assays (methods) are indicated by (*).

Invader	*P. poae*	*P. viridiflava*	*X. campestris*
Successful invasions	37	8	49
Unsuccessful invasions	98	127	78
Success rate	0.27	0.06	0.39
Mean relative abundance	0.16	0.27	0.03

Community	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Success rate	0.31	0.12	0.54	0.54	0.38	0.54	0.12	0.12	0.22	0.19	0.07	0.22	0	0.22	0

Table 3.1 – Summary of invasion outcomes by invader and invaded community: Success rate is calculated as the number of successful invasions out of the total number of invasions.

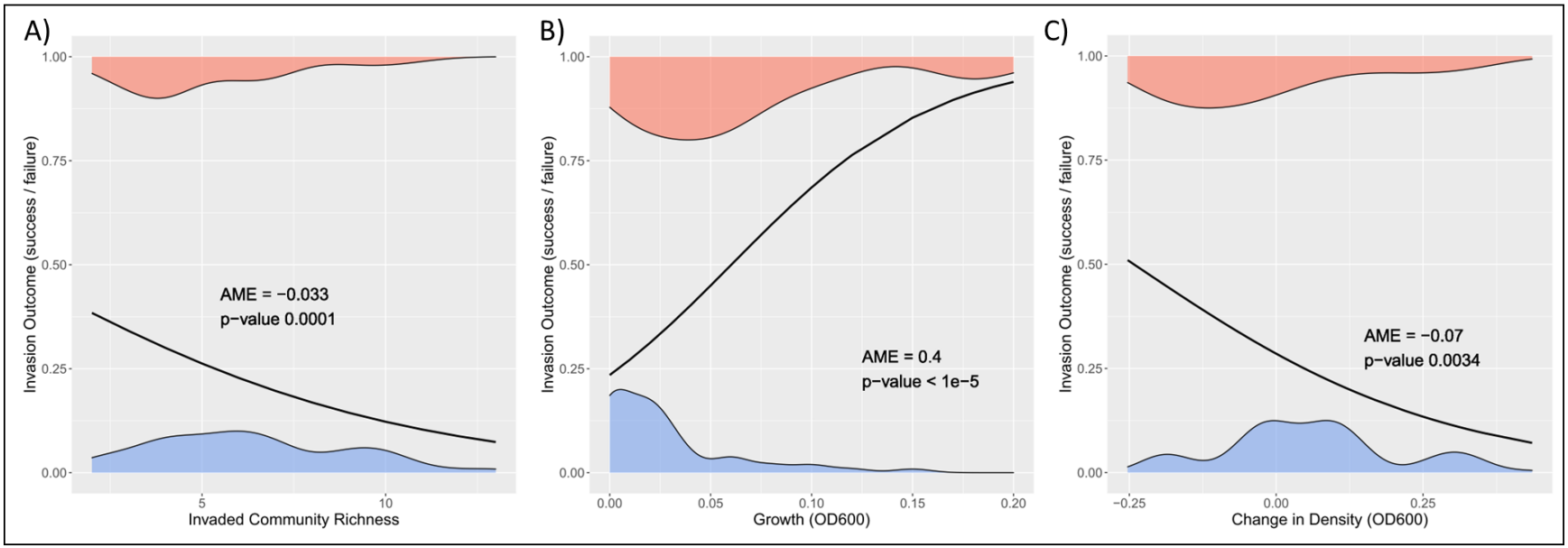


Figure 3.2 – Invaded community richness and estimates of community resource use efficiency were predictive of invasion outcome: Plots depicting the results of logistic regressions analyzing the relationship between invasion outcome (success/failure) with A) the richness of an invaded community B) the growth (optical density) of the invader *Pseudomonas poae* (MEJ082) on spent-media, and C) the change in density between day 1 and day 6 of the "late" invasion communities prior to invasion. Average marginal effects (AME) describe the change in probability of a successful invasion associated with an increase in richness of 1 or a change in density of 0.1. Density plots at the top and bottom of each plot represent the observed distributions of invaded community richness, density, or change in density, partitioned by invasion outcome (blue = failed invasion, red = successful invasion).

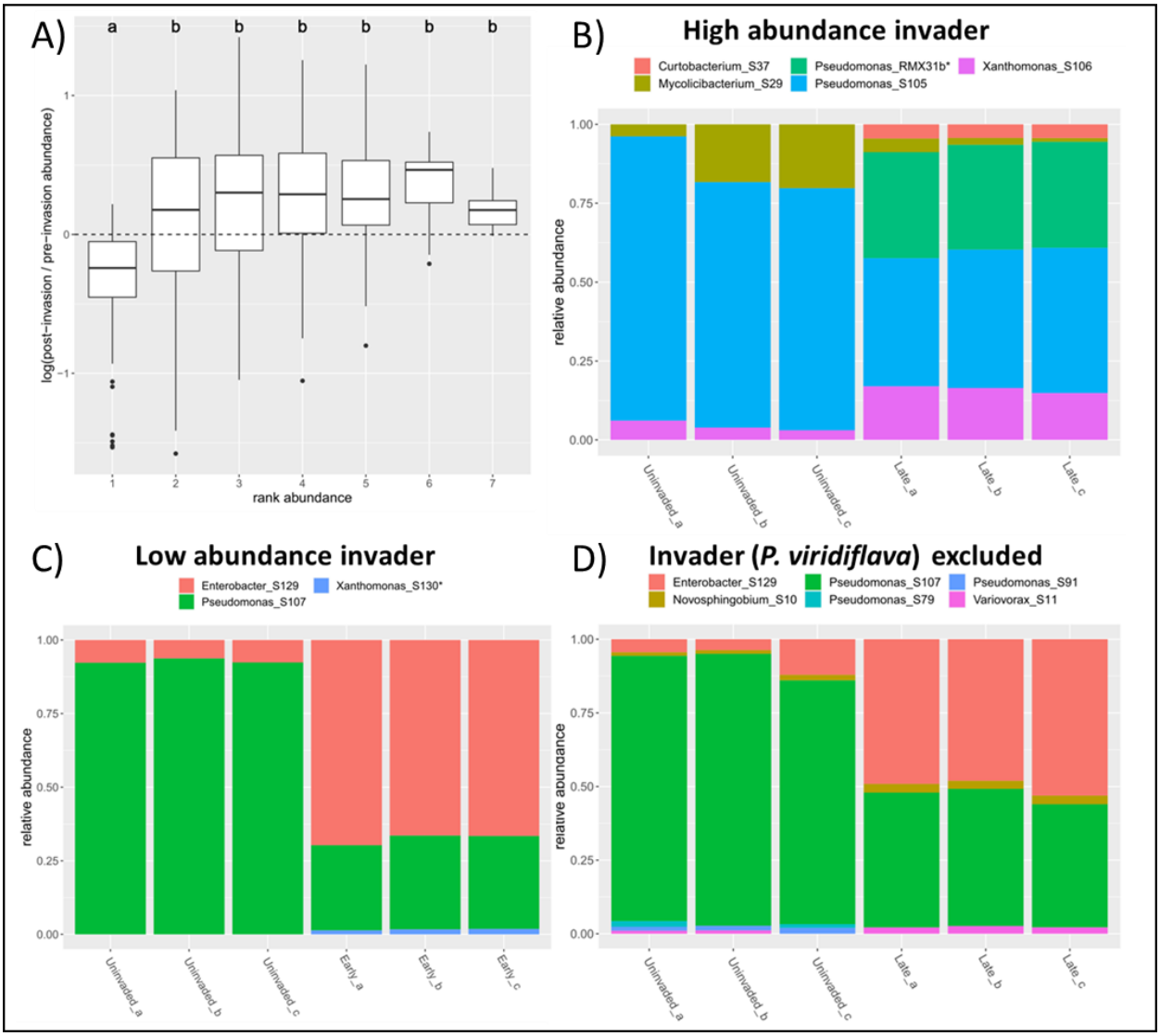


Figure 3.3 – Highest rank abundance community members were most affected by invasion: A) Bar and whisker plots display the distribution of log-ratios between post- and pre-invasion relative abundances for all community members across all communities, grouped by rank abundance in a respective uninvaded context. Significant differences between groups determined through Tukey's honest significance test. B-D) Examples demonstrating how invasion affected community composition. In B), the invader (Pseudomonas_RMX3.1b) had a strong effect on the most-abundant member of the invaded community (Pseudomonas_S105) by supplanting it. In C), the invader (Xanthomonas_S130) had a strong effect on the most-abundant member of the invaded community (Pseudomonas_S107), despite itself occupying a very low relative abundance. And in D), the invader (Pseudomonas_RMX3.1b) had a strong effect on the most-abundant member of the invaded community (Pseudomonas_S107), despite not persisting in the community.

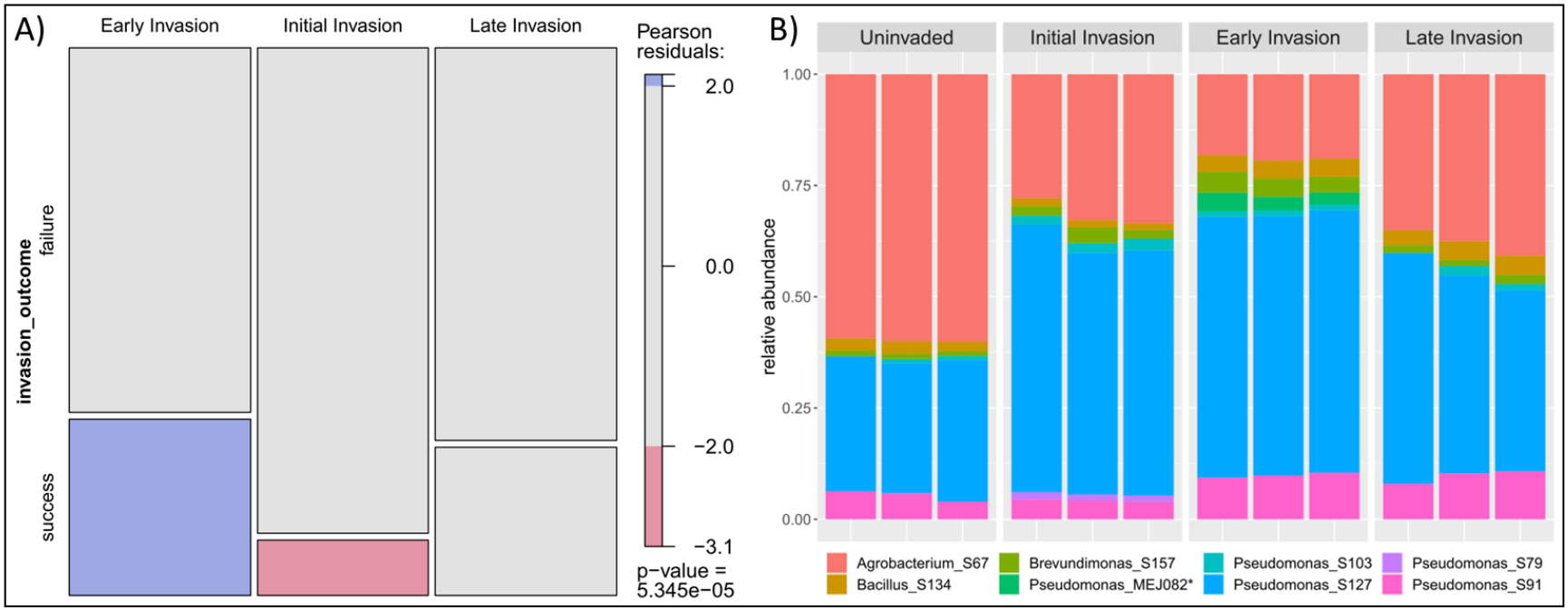
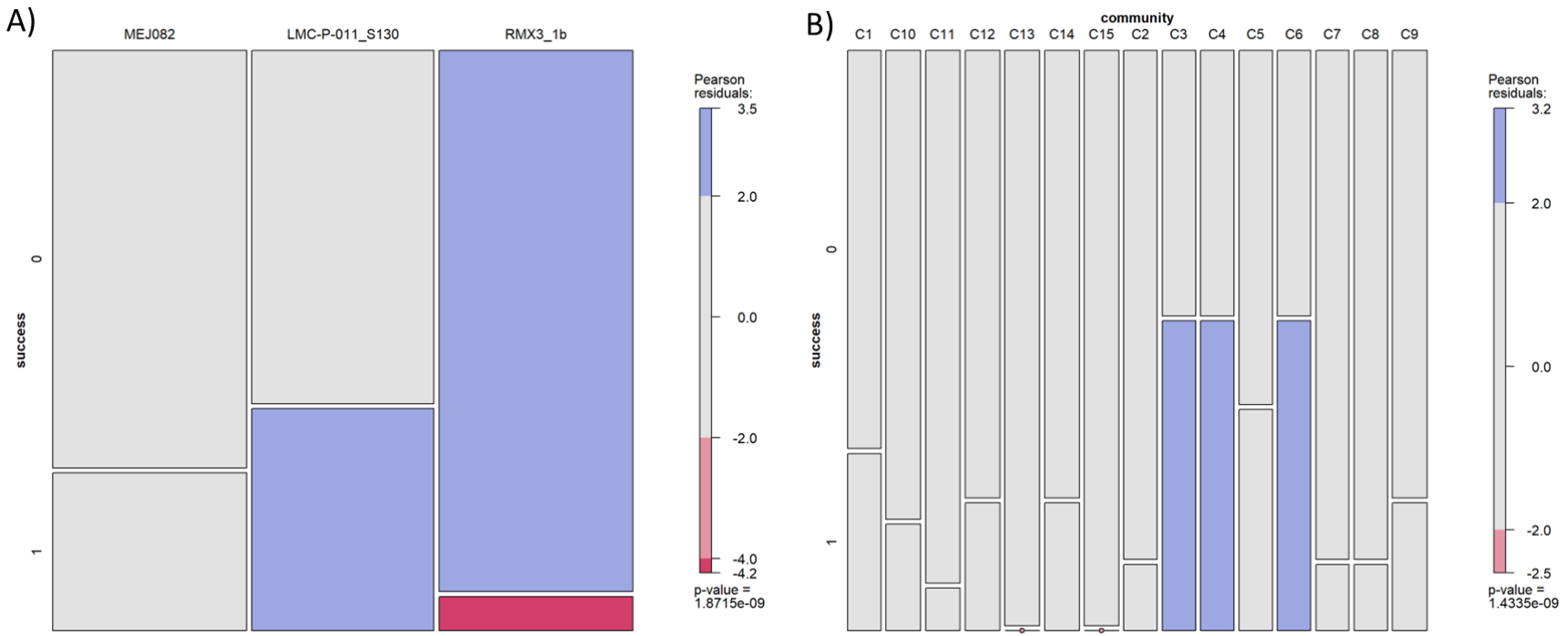


Figure 3.4 – Invasion timing affected invasion outcome: A) Mosaic plot representing the relative frequencies of invasion outcome across the three invasion timing treatments. Invasion outcome was defined as a "success" if the invader persisted over time, or a "failure" if the invader was ultimately excluded. Chi-square test of independence demonstrates a significant association between invasion outcome and invasion timing, with the Pearson residuals indicating invasion success was over-represented in the early-invasion treatment but under-represented in the initial-invasion treatment. B) An example displaying community composition (as stacked bar plots representing biological triplicates) across the different invasion treatments. Here, the same community was never invaded ("uninvaded") or invaded by Pseudomonas_MEJ082 (*P. poae*) at community initiation ("initial invasion"), after 24 hours and one passage ("early invasion"), after seven days of passaging ("late invasion"). In this example, the invader only persisted in the "early" treatment. Bray-Curtis dissimilarity described in text was calculated between a given replicate of an invasion treatment and each replicate of the uninvaded control community and then averaged.

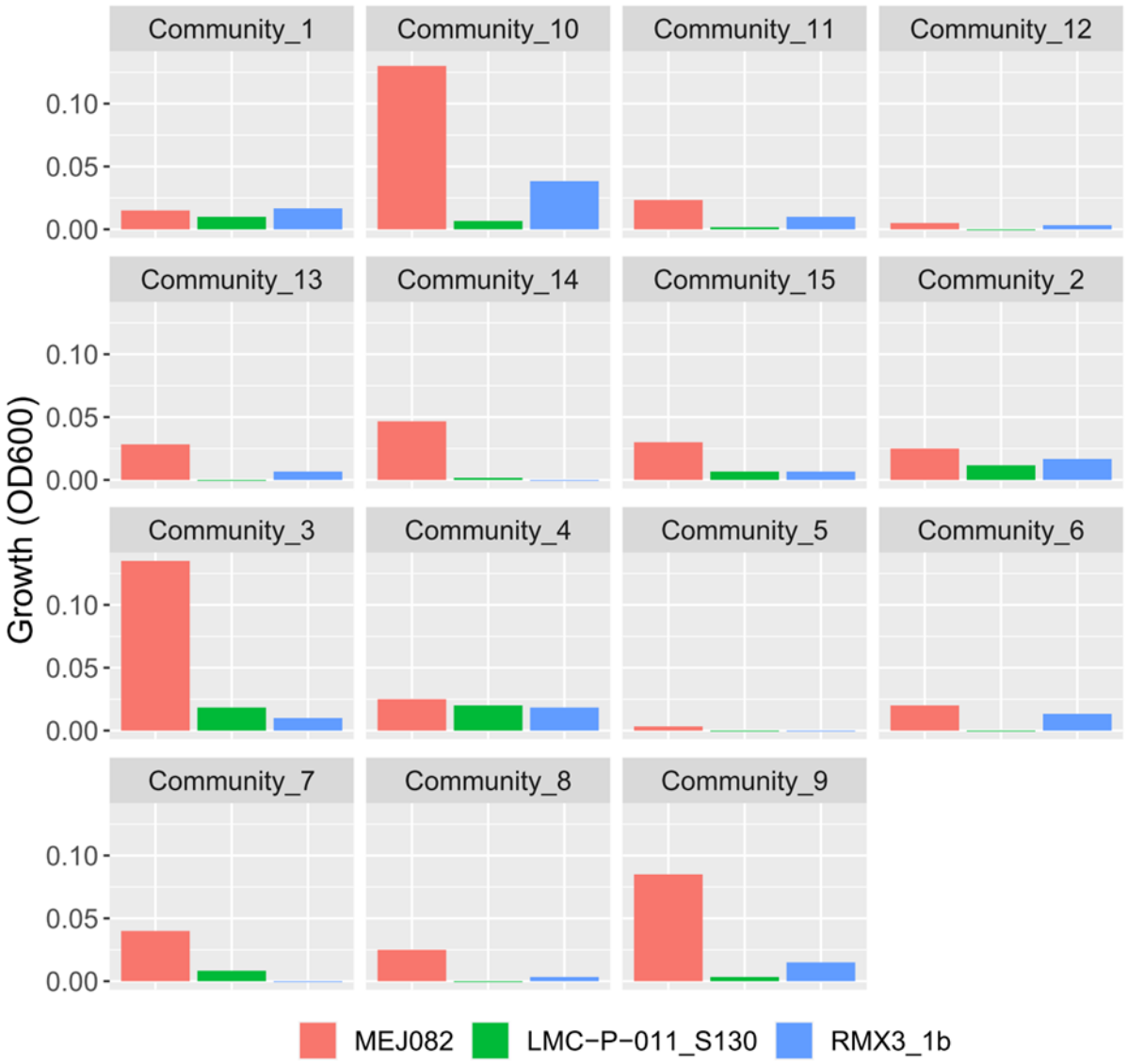
3.8 Supplementary Figures and Tables

name	genome length (bp)	number contigs	N50	GC content	percent completion	percent redundancy	genus
MEB111_S57	6.23E+06	35	3.12E+05	72.5	98.6	5.6	Promicromonospora
fls2-241-R2A-57_S113	5.28E+06	229	3.52E+04	36.4	100.0	1.4	Lysinibacillus
efr-133-R2A-89_S7	4.55E+06	18	5.27E+05	66.2	98.6	1.4	Pseudarthrobacter
efr-133-TYG-23_S107	5.68E+06	58	1.60E+05	60.7	100.0	0.0	Pseudomonas
LMC-P-059a_S33	3.89E+06	11	5.98E+05	69.7	100.0	0.0	Plantibacter
ME-Dv--P-122a_S32	3.77E+06	20	2.86E+05	71.1	100.0	0.0	Curtobacterium
efr-133-R2A-84_S152	4.45E+06	23	4.00E+05	52.9	100.0	0.0	Unknown_genera
MEB024_S129	4.77E+06	18	4.65E+05	54.6	100.0	0.0	Unknown_genera
FR229a_S127	5.93E+06	46	2.32E+05	60.2	98.6	1.4	Pseudomonas
efr-133-TYG-118_S22	5.28E+06	50	1.77E+05	63.6	98.6	5.6	Arthrobacter
lyk4-R2A-8_S103	5.48E+06	44	2.11E+05	61.1	97.2	0.0	Pseudomonas
lyk4-40-TSB-66_S53	3.57E+06	18	4.77E+05	70.3	100.0	0.0	Microbacterium
fls2-241-R2A-200_S19	3.81E+06	27	2.55E+05	62.6	100.0	0.0	Unknown_genera
MEB011_S37	3.66E+06	20	3.53E+05	70.9	100.0	0.0	Curtobacterium
CFBP2214_S67	5.58E+06	33	6.62E+05	59.1	100.0	0.0	Agrobacterium
MEB006b_S157	2.94E+06	11	4.56E+05	66.3	100.0	0.0	Brevundimonas
lyk4-R2A-10_S69	6.33E+06	30	4.94E+05	60.6	98.6	1.4	Pseudomonas
lyk4-R2A-2_S134	3.64E+06	19	7.64E+05	42.0	100.0	0.0	Bacillus
efr-133-TYG-120_S21	4.97E+06	55	1.60E+05	63.7	98.6	1.4	Arthrobacter
fls2-241-R2A-168_S45	4.88E+06	64	1.79E+05	64.4	98.6	1.4	Pseudarthrobacter
fls2-241-TYG-188a_S156	5.35E+06	16	1.49E+06	57.1	100.0	0.0	Agrobacterium
ME-Dv--P-095_S58	4.01E+06	21	4.58E+05	69.6	100.0	0.0	Plantibacter
lyk4-R2A-23_S48	7.71E+06	60	2.02E+05	66.6	100.0	2.8	Burkholderia
ME-Dv--P-043b_S39	4.41E+06	50	1.72E+05	66.2	100.0	1.4	Luteibacter
CFBP2511_S1	6.19E+06	48	2.16E+05	58.6	100.0	1.4	Pseudomonas
fls2-241-TYG-175_S105	5.98E+06	64	1.53E+05	59.3	98.6	1.4	Pseudomonas
MEB032_S102	7.15E+06	67	2.02E+05	62.3	98.6	2.8	Rhodococcus
MEB041_S88	4.57E+06	15	1.09E+06	68.5	98.6	1.4	Rhodococcus
lyk4-40-TYG-92_S29	6.17E+06	34	3.42E+05	66.8	98.6	1.4	Mycolicibacterium
efr-133-TYG-104_S8	3.85E+06	38	1.78E+05	62.6	100.0	0.0	Micrococcaceae
LMC-A-07_S149	4.94E+06	31	2.90E+05	65.3	100.0	1.4	Xanthomonas
fls2-241-TYG-148_S106	4.97E+06	32	2.74E+05	65.4	100.0	0.0	Xanthomonas
MEJ086_S119	5.85E+06	64	1.64E+05	62.3	100.0	0.0	Pseudomonas
efr-133-TYG-103a_S91	6.12E+06	51	2.07E+05	60.3	98.6	0.0	Pseudomonas
efr-133-TYG-130_S11	7.08E+06	64	2.46E+05	67.2	100.0	2.8	Variovorax
LMC-P-041_S56	3.70E+06	14	5.44E+05	67.8	100.0	0.0	Microbacterium
ME-P-080_S24	3.28E+06	11	5.88E+05	71.5	100.0	2.8	Frigoribacterium
lyk4-TYG-107_S51	6.15E+06	50	2.60E+05	60.1	97.2	1.4	Pseudomonas
MEB105_S97	6.09E+06	24	3.28E+05	60.3	100.0	0.0	Pseudomonas
lyk4-40-TSB-59a_S125	6.36E+06	26	3.93E+05	60.5	98.6	1.4	Pseudomonas
ME-P-057_S98	5.70E+06	59	1.60E+05	60.1	98.6	0.0	Pseudomonas
fls2-241-R2A-127_S5	4.70E+06	59	1.70E+05	65.6	98.6	2.8	Pseudarthrobacter
fls2-241-R2A-110_S136	6.84E+06	53	2.19E+05	59.0	98.6	0.0	Pseudomonas
fls2-241-R2A-195_S10	5.74E+06	61	1.88E+05	65.2	100.0	2.8	Novosphingobium
efr-133-R2A-120_S25	5.12E+06	81	1.24E+05	63.7	98.6	1.4	Arthrobacter
efr-133-TYG-5_S79	6.12E+06	13	7.94E+05	62.1	94.4	1.4	Pseudomonas
lyk4-40-TYG-31_S151	5.14E+06	25	4.66E+05	57.5	100.0	1.4	Agrobacterium
lyk4-40-TYG-27_S50	4.56E+06	24	5.42E+05	66.2	98.6	1.4	Pseudarthrobacter

Supplementary Table 3.1: isolate details



Supplementary figure 3.1 – Invasion success varied between invader and community: A) Mosaic plot representing successful and failed invasions for each invader. Chi square test of independence shows invader identity was statistically associated with invasion success (p-value $1.87e^{-9}$). Successful invasions were observed more often than expected for Xanthomonas_S130. Pseudomonas_RMX3.1b demonstrated more failures than expected, and fewer successes. B) Mosaic plot representing successful and failed invasions in each community. Chi square test of independence shows community identity was statistically associated with invasion success (p-value $1.43e^{-9}$). Successful invasions were more common than expected for communities 3, 4, and 6, and less common than expected for communities 13 and 15.



Supplementary figure 3.2 – Growth on spent media varied by community and invader: Bar plots display the average optical density, as a proxy for growth, of each invader across spent-medium from each community. Averages were taken across both the initial and early invasion treatment spent media.

Logistic regression:	invasion outcome ~ richness			
Variable	**Estimate**	**Std. Error**	**z value**	**Pr > |z|**
(intercept)	-0.098	0.301	-0.325	0.745
Richness	-0.187	0.051	-3.66	0.0003
Null Deviance	433.5	on 394 df		
Residual Deviance	418.74	on 393 df		
invasion outcome ~ density on spent media (*P. poae*)				
(intercept)	-1.18	0.33	-3.58	0.0003
Density	1.96	0.6	3.25	0.001
Null Deviance	121.91	on 89 df		
Residual Deviance	107.46	on 88 df		
invasion outcome ~ change in density (day 1 vs day 6)				
(intercept)	-0.918	0.199	-4.61	$< 4e^{-6}$
Change in Density	-0.379	0.141	-2.69	0.007
Null Deviance	158.56	on 134 df		
Residual Deviance	150.48	on 133 df		

Average Marginal Effects					*95% CI*	
Variable	**AME**	**SE**	**z**	**p-value**	**lower**	**upper**
Richness	-0.033	0.009	-3.82	0.0001	-0.05	-0.016
Density (*P. poae*)	0.402	0.095	4.22	0.0000	0.216	0.589
Change in Density	-0.071	0.024	-2.93	0.0034	-0.118	-0.024

Supplementary Table 3.2 – Invaded community richness and resource use efficiency were associated with invasion success: Details of logistic regressions analyzing the relationships between invasion outcome with the richness of an invaded community ("Richness"), the density of a specific invader (*P. poae MEJ082*) on spent-media ("Density"), or the change in community density between day 1 and day 6 (all density measurements represent optical density at 600nm). The average marginal effects of: an increase of 1 in richness, an increase of 0.1 in the density of *P. poae*, and a change in density of 0.1 are presented in the lower section of the table.

A) One-way ANOVA: Bray-Curtis dissimilarity ~ invasion treatment				
variable	***df***	***sum of squares***	***F***	***p-value***
invasion treatment	2	0.456	5.23	0.006
residuals	381	16.6		
B) Tukey's Honest Significance Test: invasion treatment				
invasion treatment	**mean difference**	**95% CI lower**	**95% CI upper**	**adjusted p-value**
initial – late	-0.026	-0.088	0.036	0.596
early – late	0.056	-0.005	0.117	0.081
early – initial	0.082	0.021	0.143	0.005

Supplementary Table 3.3 – The initial invasion treatment was the most dissimilar treatment relative to the uninvaded communities: A) One-way ANOVA (Type III) results analyzing the relationship between Bray-Curtis dissimilarity (relative to the uninvaded communities) and invasion timing treatment. B) Tukey's Honest Significance Test results comparing the differences in Bray-Curtis dissimilarities between the invasion timing treatments.

A) One-way ANOVA: Bray-Curtis dissimilarity ~ invasion treatment				
variable	***df***	***sum of squares***	***F***	***p-value***
invasion treatment	2	0.456	5.23	0.006
residuals	381	16.6		
B) One-way ANCOVA: Bray-Curtis dissimilarity ~ invasion treatment + richness				
invasion treatment	2	0.438	5.02	0.007
richness	1	0.015	0.336	0.562
residuals	380	16.6		
C) Two-way ANOVA: Bray-Curtis dissimilarity ~ invasion treatment + invader				
invasion treatment	2	0.455	5.22	0.006
invader	2	0.09	1.03	0.358
residuals	379	16.5		
D) One-way ANCOVA: Bray-Curtis dissimilarity ~ invasion treatment + spent media growth				
invasion treatment	1	0.17	4.32	0.0386
spent media growth	1	0.324	8.24	0.004
residuals	254	9.99		

Supplementary Table 3.4 – Invader growth on spent media is a significant covariate in the relationship between invasion timing and outcome: A) One-way ANOVA results from supplementary table 3.3, for reference. B) One-way ANCOVA including pre-invasion community richness as a covariate. C) Two-way ANOVA including invader identity as a blocking effect. D) One-way ANCOVA including invader growth on spent media as a covariate (limited to "early" and "late" invasion treatments).

3.9 References

Amor, Daniel R., Christoph Ratzke, and Jeff Gore. "Transient invaders can induce shifts between alternative stable states of microbial communities." *Science Advances* 6.8 (2020): eaay8676.

Bergelson, Joy, et al. "Functional biology in its natural context: A search for emergent simplicity." *Elife* 10 (2021): e67646.

Bittleston, Leonora S., et al. "Context-dependent dynamics lead to the assembly of functionally distinct microbial communities." Nature communications 11.1 (2020): 1440.

Britton, Robert A., and Vincent B. Young. "Role of the intestinal microbiota in resistance to colonization by Clostridium difficile." *Gastroenterology* 146.6 (2014): 1547-1553.

Burke, Mike JW, and J. P. Grime. "An experimental study of plant community invasibility." *Ecology* 77.3 (1996): 776-790.

Burns, Adam R., et al. "Interhost dispersal alters microbiome assembly and can overwhelm host innate immunity in an experimental zebrafish model." *Proceedings of the National Academy of Sciences* 114.42 (2017): 11181-11186.

Carlström, Charlotte I., et al. "Synthetic microbiota reveal priority effects and keystone strains in the Arabidopsis phyllosphere." *Nature Ecology & Evolution* 3.10 (2019): 1445-1454.

Clark, Graeme F., and Emma L. Johnston. "Temporal change in the diversity–invasibility relationship in the presence of a disturbance regime." *Ecology Letters* 14.1 (2011): 52-57.

Clark, Ryan L., et al. "Design of synthetic human gut microbiome assembly and butyrate production." *Nature communications* 12.1 (2021): 1-16.

Dillon, R. J., et al. "Diversity of locust gut bacteria protects against pathogen invasion." *Ecology Letters* 8.12 (2005): 1291-1298.

Eisenhauer, Nico, Stefan Scheu, and Alexandre Jousset. "Bacterial diversity stabilizes community productivity." *PloS one* 7.3 (2012): e34517.

Fiegna, Francesca, et al. "Evolution of species interactions determines microbial community productivity in new environments." *The ISME journal* 9.5 (2015): 1235-1245.

Fox, John, and Sanford Weisberg. *An R companion to applied regression*. Sage publications, 2018.

Hodgson, David J., Paul B. Rainey, and Angus Buckling. "Mechanisms linking diversity, productivity and invasibility in experimental bacterial communities." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269.1506 (2002): 2277-2283.

Innerebner, Gerd, Claudia Knief, and Julia A. Vorholt. "Protection of Arabidopsis thaliana against leaf-pathogenic Pseudomonas syringae by Sphingomonas strains in a controlled model system." *Applied and environmental microbiology* 77.10 (2011): 3202-3210.

Jones, Matt L., et al. "Biotic resistance shapes the influence of propagule pressure on invasion success in bacterial communities." Ecology (2017): 1743-1749.

Kennedy, Theodore A., et al. "Biodiversity as a barrier to ecological invasion." *Nature* 417.6889 (2002): 636-638.

Lawrence, Diane, et al. "Species interactions alter evolutionary responses to a novel environment." *PLoS biology* 10.5 (2012): e1001330.

Li, Wei, and M. Henry H. Stevens. "Fluctuating resource availability increases invasibility in microbial microcosms." *Oikos* 121.3 (2012): 435-441.

Liu, Yong-Xin, Yuan Qin, and Yang Bai. "Reductionist synthetic community approaches in root microbiome research." *Current Opinion in Microbiology* 49 (2019): 97-102.

Loreau, Michel, and Andy Hector. "Partitioning selection and complementarity in biodiversity experiments." *Nature* 412.6842 (2001): 72-76.

Mallon, Cyrus A., et al. "Resource pulses can alleviate the biodiversity–invasion relationship in soil microbial communities." *Ecology* 96.4 (2015): 915-926.

Martínez, Inés, et al. "Experimental evaluation of the importance of colonization history in early-life gut microbiota assembly." *Elife* 7 (2018): e36521.

Meyer, D. Zeileis, and A. Zeileis. "A. & Hornik K. 2011. vcd: Visualizing Categorical Data." *R package version* (2020): 1-2.

Morella, Norma M., et al. "Successive passaging of a plant-associated microbiome reveals robust habitat and host genotype-dependent selection." *Proceedings of the National Academy of Sciences* 117.2 (2020): 1148-1159.

Nguyen, Jen, Juanita Lara-Gutiérrez, and Roman Stocker. "Environmental fluctuations and their effects on microbial communities, populations and individuals." *FEMS microbiology reviews* 45.4 (2021): fuaa068.

Oksanen, Jari, et al. "Package 'vegan'." *Community ecology package, version* 2.9 (2013): 1-295.

Peart, David R., and Theodore C. Foin. "Analysis and prediction of population and community change: a grassland case study." *The population structure of vegetation* (1985): 313-339.

Tilman, David. "The ecological consequences of changes in biodiversity: a search for general principles." *Ecology* 80.5 (1999): 1455-1474.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rivett, Damian W., et al. "Resource-dependent attenuation of species interactions during bacterial succession." *The ISME journal* 10.9 (2016): 2259-2268.

Rivett, Damian W., et al. "Elevated success of multispecies bacterial invasions impacts community composition during ecological succession." *Ecology Letters* 21.4 (2018): 516-524.

Robinson, Courtney J., Brendan JM Bohannan, and Vincent B. Young. "From structure to function: the ecology of host-associated microbial communities." *Microbiology and Molecular Biology Reviews* 74.3 (2010): 453-476.

Rohland, Nadin, and David Reich. "Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture." *Genome research* 22.5 (2012): 939-946.

Stachowicz, John J., Robert B. Whitlatch, and Richard W. Osman. "Species diversity and invasion resistance in a marine ecosystem." *Science* 286.5444 (1999): 1577-1579.

Santos, A., M. San Mauro, and D. Marquina Díaz. "Prebiotics and their long-term influence on the microbial populations of the mouse bowel." *Food Microbiology* 23.5 (2006): 498-503.

Symons, Celia C., and Shelley E. Arnott. "Timing is everything: priority effects alter community invasibility after disturbance." *Ecology and evolution* 4.4 (2014): 397-407.

Wei, Zhong, et al. "Trophic network architecture of root-associated bacterial communities determines pathogen invasion and plant health." *Nature communications* 6.1 (2015): 8413.

Wickham, Hadley. "Reshaping data with the reshape package." *Journal of statistical software* 21 (2007): 1-20.

Wickham, Hadley, et al. "Welcome to the Tidyverse." *Journal of open source software* 4.43 (2019): 1686.

Yang, Tianjie, et al. "Resource availability modulates biodiversity-invasion relationships by altering competitive interactions." *Environmental microbiology* 19.8 (2017): 2984-2991.

Zhou, Kang, et al. "Distributing a metabolic pathway among a microbial consortium enhances production of natural products." *Nature biotechnology* 33.4 (2015): 377-383.

Conclusion

In this dissertation, I used synthetic bacterial communities to investigate interspecific interactions, coexistence, and ecological invasion. In the first chapter, I evaluated the assumption that the interaction between two community members is unaffected by the surrounding community context and found that changes in community richness and density were robust predictors of variation in interaction effects. In the second chapter, I deconstructed a synthetic community into all pairwise and *n-2* communities to compare coexistence between "bottom-up" and "top-down" contexts and concluded that pairwise observations of coexistence and exclusion were useful but imperfect predictors of the structure of more complex assemblages. And lastly, in the third chapter, I investigated how the timing of an ecological invasion affected the outcome of that invasion and found evidence indicating that the effect of timing on invasion outcome was linked to changes in resource use efficiency over the community assembly process.

Though each chapter focuses on separate aspects of microbial community ecology, the findings of each are not wholly unrelated. The relationships between richness, community density, and interactions observed in chapter 1 are related to the relationships between richness, productivity and invasibility observed in chapter 3. Namely, in interpreting the results of chapter 1, I proposed that the attenuation of competitive interactions associated with an increase in richness and total density may have arisen due to increased community metabolic activity which spilled over into positive effects for the less fit species present. This is akin to the "sampling effect" mentioned in chapter 3, where increased richness increases the chances of an especially fit species being present in a community and thus increasing community productivity; except in this case, the fitness of that species additionally may have further increased community productivity by positively affecting other community members. More generally, the sampling effect is thought to reduce the invasibility of

ecological communities, as indeed appeared to be the case in chapter 3, where reduced invasibility was associated with higher community resource use efficiency.

Further, the context dependence of interactions observed in chapter 1 is related to the observations of context-dependent coexistence I observed in chapter 2. Namely, the emergent (and potentially higher order) effects that led to the attenuation of interactions in chapter 1 may have been present in the complex communities from chapter 3 and led to the context-dependent coexistence we observed there. And finally, the emergent coexistence we observed in the complex communities of chapter 2 indicates another mechanism by which increased richness might lead to decreased invasibility, namely, that starting from a richer initial pool results in a diversity-begets-diversity effect which in turn increases the resistance of a community to invaders. Thus, although the three chapters of this dissertation were designed to ask different questions, they all overlap in some of the fundamental ecological principles they investigate.

Though the experimental system I developed was effective for investigating the community ecology of simple microbial systems, there are limitations that I feel it would be fruitful to discuss here. The major limitation of this system is its reliance on sequencing to assay the composition of each community. This is a limitation in at least two important ways. First, my strategy of shallow metagenomic sequencing was imperfect due to genomic similarity between the isolates I worked with. Despite my best effort to select isolates that were sufficiently divergent in genomic content, I ultimately had to apply a correction to my sequencing data as some samples had high numbers of reads that were thrown out due to ambiguous read mapping. This addition of technical uncertainty was better than the alternative (dropping affected samples) but could have been avoided. One potential solution would have been to genetically barcode the isolates I was working with and use amplicon sequencing to assay community composition. This would have removed the challenge of

ambiguous read mapping with the additional benefit of reducing the per library price of sequencing. Unfortunately, I chose not to pursue this option, concerned by the effort it would have taken to barcode the diverse set of bacterial isolates I was intent on working with. And second, although sequencing is an excellent way to generate large amounts of data, it is comparatively slow and expensive unless performed in large batches. This made it challenging to perform iterative cycles of experimentation. Indeed, the data for chapters 1 and 2 were each collected from two large sequencing runs, and chapter 3 from only a single large sequencing run. However, all chapters likely would have benefited from smaller scale follow-up experiments based on earlier results. This is partly an issue of my own limitations in how quickly I could produce and process these results, but also a limitation of the system's reliance on sequencing. What would have been a better approach, you ask? Well, dear reader, please permit me to dream for a moment and conclude my dissertation with a description of what I feel would be an exciting and powerful improved experimental system.

My ideal experimental system is one that allows for the rapid, inexpensive, and unambiguous characterization of microbial communities. Others have recently developed approaches that come close to this ideal, specifically using microfluidics to randomly assemble large quantities of simple microbial communities, which can then be characterized through sequencing or microscopy (Hsu et al., 2019; Kehe et al., 2019). Although I feel these approaches are powerful and useful, to me, they represent limited tools for the study of microbial communities for a couple of reasons. First, one does not have precise control over community composition. And second, such systems offer little opportunity to study the dynamics of microbial communities, as communities are encapsulated within microdroplets which quickly limit growth. I would ideally take a different approach that permits a greater level of experimental control and the opportunity for community dynamics to unfold, thus allowing the study of more complex and realistic communities.

I envision a system that uses flow cytometry and multiplexed fluorescent markers to allow the rapid and direct tracking of microbial community dynamics. This system would rely on a recently developed approach for transforming microbes to express multiple fluorescent proteins at different expression levels (Anzalone et al., 2021). By discriminating between the unique fluorescence profiles resulting from the unique expression profiles, this approach was able to distinguish between 20 yeast strains using only two fluorescent proteins. Such a system would certainly require a good deal of technical and molecular effort to establish, which would likely necessitate focusing on a relatively limited set of bacterial isolates one could work with. However, with even 20 uniquely marked strains, one would have no shortage of communities to assemble and study. And given sufficient methods development, I think it would be feasible to use an automated flow cytometer (which could autonomously measure preserved samples from a 96 or 384-well plate) to efficiently measure and analyze many samples. With such a system, one could run experiments like mine but obtain results in almost real time. This would bypass the challenges of working with indirectly measured relative abundance sequencing data, facilitate measurements at a finer timescale, and enable rapid iterative experiments. Individual experiments would also be comparatively cheaper to perform, although the upfront and recurring cost of a flow cytometer capable of performing this work would likely require a considerable (but worthwhile) investment. I think such a system is technically feasible and would produce fruitful experimental data, especially when guided by and partnered with ecological theory.

References

Hsu, Ryan H., et al. "Microbial interaction network inference in microfluidic droplets." *Cell systems* 9.3 (2019): 229-242.

Kehe, Jared, et al. "Massively parallel screening of synthetic microbial communities." *Proceedings of the National Academy of Sciences* 116.26 (2019): 12804-12809.

Anzalone, Andrew V., Miguel Jimenez, and Virginia W. Cornish. "FRAME-tags: genetically encoded fluorescent markers for multiplexed barcoding and time-resolved tracking of live cells." *Biorxiv* (2021): 2021-04.