Evolutionary Consequences of Metabolic Competition and Cooperation in Microbial Communities

Colin McNally

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Elhanan Borenstein, Chair

Jesse Bloom

Maitreya Dunham

Program Authorized to Offer Degree:

Genome Sciences

University of Washington

Abstract

Evolutionary Consequences of Metabolic Competition and Cooperation in Microbial Communities

Colin McNally

Chair of the Supervisory Committee:
Elhanan Borenstein, Associate Professor
Department of Genome Sciences

Microbial communities are ubiquitous and perform functions fundamental to life on earth. The high density of organisms within them and the many ways in which these species interact with each other result in limitless opportunities for competition and cooperation to affect the fitness of individual species. Gaining a better understanding of the evolutionary process within such dense communities has important implications for both clarifying genomic variation observed in the present as well as predicting future changes in microbial communities of particular interest to human health and industry. Coevolution is challenging to study due to the high dimensionality of mutations interacting with ecological dynamics which further affect the fitness of future mutations arising in the same and other species. Previous studies of coevolution tend to use in vivo systems with limited scope or abstracted interactions and adaptations in

theoretical or in silico systems to achieve broader scope. In this study I focus on one type of bacterial interaction, competition and cooperation over metabolic resources, and use a computational model that captures mechanistic detail of the genomic and metabolic basis of such interactions and yet is amenable to studying large scales of replicates and conditions. In chapter 1, I describe the behaviors and mechanisms through which bacteria interact, the current state of knowledge of coevolution, and the computational methods that have been applied to microbial ecology and evolution. In chapter 2, I consider the evolutionary dynamics that occur within the tightly coupled, reductive evolution regime of insect endosymbionts. Using a modeling approach, I simulate thousands of replicate evolutionary trajectories and identify relationships between changes to genome content and cooperative metabolic phenotypes. In chapter 3, I apply a similar modeling approach to study how the fitness landscape of a focal species changes depending on what other species live in the same community. By simulating over one thousand mutations and hundreds of community compositions I identify complex and prevalent relationships between mutations and community contexts. Finally, in chapter 4 I reflect on commonalities in the findings of these studies, future directions for similar research, and potential applications to synthetic biology.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGEMENTS

I would like to thank my friends and family for their support and encouragement and providing needed distractions from work. I would like to thank my labmates for our camaraderie, thought provoking discussions, and fun collaborations on side projects. And most importantly I would like to thank my mentor Elhanan for his mentorship, optimism, and support of both my research interests and personal growth.

Chapter 1. INTRODUCTION

1.1 MICROBIAL COMMUNITIES

Although unseen to the naked eye, microbial communities cover nearly the entire earth. They are found in the oceans, in the soil, on every terrestrial surface and even on and inside the digestive systems of animals. Formed from viruses, archaea, bacteria, protists, and fungi, these communities are among the most diverse, dense, and productive of all life on earth and account for a huge portion of total life [1]. They are the foundation of natural ecosystems and are involved in global scale biogeochemical systems such as the nitrogen cycle [2,3]. Although microbiology was long focused on the study of individual species, sequencing technologies have allowed measurements of entire communities to identify the constituent species and their abundances [4]. This has greatly propelled interest in understanding and utilizing these communities.

In addition to their ecological importance microbial communities also play essential roles in many processes of import to many aspects of human life. They are a powerful component of wastewater treatment in methods such as activated sludge, where sewage is aerated to allow aerobic bacteria to break down organic matter [5]. They also live on and inside humans, playing a key role in host health [6]. The human microbiome is important for the development of the immune system [7] and the gut microbiome has been associated with various diseases [8]. Thus, gaining a better understanding of these communities and the activities of individual species within them will have broad impacts.

1.2 Interspecific Interactions Between Bacteria

These communities aren't just mixtures of independently acting species, but instead bacteria interact with each other in many ways [9]. On a basic level bacteria compete for limited environmental resources like carbon and nitrogen sources [10]. They can also cooperate through metabolic interaction with one species leaking or excreting a metabolite that can be beneficially consumed by another. Many such examples have been found to occur in natural conditions, sometimes essential for growth [11–13]. Experimental studies have identified many more metabolites that can potentially be beneficially cross-fed [14–16]. However in general most forms of interaction are competitive, not cooperative [17]. Bacteria may actively harm or kill each other through various mechanisms of secreting or injecting toxins, such as the type VI secretion system [18,19]. They may send signals to members of their own species or others through quorum sensing to coordinate behavior of individual cells in response to environmental conditions [20]. And finally they may construct biofilms, physical structures of linked cells that act to create a more favorable environment for the cells and species included in the film while physically excluding competitors [21]. This large array of potential interactions creates a daunting level of complexity in microbial communities for ecological interactions to play out.

There is plenty of evidence that such interactions are important for determining the composition of microbial communities and the fitness of individual species. Significant evidence for widespread co-dependence between microbes comes from the large number of species that cannot be cultured in isolation in laboratory conditions. Many of these "unculturable" species can be grown in devices that separate cells but allow small molecules to pass between chambers, showing the importance of cooperative interactions [22,23]. Evidence of the importance of competition comes from the reproducibility of community composition when similar species are

grown in co-culture, where primary and secondary resource preference can play a large role in determining species abundance [24]. Similar reproducibility of composition was found in experimental evolution of a co-culture of six species [25]. Lastly analysis of many data sets has found evidence that metabolic cooperation results in consistent co-occurrence of species across disparate environments [26]. All these lines of evidence paint a picture of a tangled web of interspecies interactions that is vitally important for the involved bacteria.

1.3 Bacterial Evolution

Bacteria evolution occurs rather differently than most plants and animals due to their asexual reproduction. This makes it difficult to define what a bacterial species, due to the large set of strains that may deviate from each other but retain many similarities. Despite the potential for clonal sweeps to reduce diversity it has been found that wild bacterial populations harbor a large amount of genotypic diversity [27,28]. Without sexual reproduction bacteria lack the ability for recombination to separate beneficial from harmful mutations during reproduction, but they make up for this limitation through various mechanisms to acquire new external genetic material. Through several different mechanisms of horizontal gene transfer bacteria can gain new genes or mutations that can give them a fitness advantage. Overall these features of bacterial evolution create many differences to the evolution of plants and animals and demand their own care and thought.

Bacterial evolution is further complicated by the vast order of interspecific interactions occurring in microbial communities, which create countless opportunities for species to influence the fitness of other species, and thus to coevolve. Coevolution between species has been a topic of study for as long as evolution has been studied, with Darwin's image of the tangled bank summarizing the high potential complexity of evolution when many species influence each other.

Several frameworks have been proposed to view the mechanisms of coevolution. The red queen hypothesis states that direct competition between two species will accelerate the evolution of both species as they gain adaptations to continue getting an advantage over the other [29]. This has been demonstrated in a parasitic relationship between a phage and bacteria [30]. With cooperation, on the other hand, the focus of study is more on the evolutionary origins. Several mechanisms have been suggested for how cooperation can evolve even in the face of natural selection, such as direct reciprocity and group selection [31]. One bacteria specific hypothesis for the evolution of cooperative interactions is the black queen hypothesis, which states that when a beneficial function is leaky there is a selective advantage for some members of the community to lose the ability to produce it themselves [32]. This has been experimentally demonstrated with hydrogen peroxide resistance and found to stabilize species co-existence [33]. Interspecies interactions result in complicated coevolutionary dynamics but are important to study to fully understand bacterial evolution in the context of communities.

Some bacterial systems that show the strongest evidence of coevolution are the obligate endosymbionts of sap feeding insects. These endosymbionts live inside specialized cells called bacteriocytes and complement the host's metabolism and nutrient poor diet by producing amino acids and vitamins [34]. The endosymbionts have greatly reduced genomes compared to their nearest free-living relatives and, like their host, lack the ability to synthesize many amino acids and vitamins and thus reciprocally rely on metabolites produced by their host [35]. These tight mutualistic associations between host and symbiont have lasted for millions of years and have convergently evolved repeatedly. Some insects have a single endosymbiont, such as aphids and their symbiont *Buchnera*, while others have two or more endosymbionts, such as the glassy winged sharpshooter with its dual symbionts *Baumannia cicadellinicola* and *Sulcia muelleri*

[36]. When multiple endosymbionts are present they also metabolically complement each other. Even within this diversity endosymbionts have been found to further diversify and subspecialize within a single host. In one case *Candidatus* Hodgkinia cicadicola, an endosymbiont of cicada of the genus *Tettigades*, was found to have diverged into two lineages which each lacked at least twenty genes that the other had retained [37]. More recently another species of *Tettigades* has been found to contain six distinct Hodgkinia strains [38]. This phenomenon of mutualistic relationships arising within such small communities of endosymbionts is a consequence of the general mode of reductive evolution that these bacteria experience. Due to their small population sizes, vertical transmission between hosts, and asexual reproduction, endosymbionts experience weak selective forces and instead primarily evolve through genetic drift and heavy gene loss [39]. The lack of selection allows their genomes to dramatically degrade and results in endosymbionts having the smallest known genomes of all bacteria [40]. This coevolutionary process is theorized to be extremely similar to the origin of the symbiotic relationships between organelles and their eukaryote hosts [41]. Thus endosymbionts provide a fascinating example of the evolutionary origin and maintenance of cooperation, albeit one that occurs under very particular circumstances.

1.4 COMPUTATIONAL METHODS TO STUDY ECOLOGY AND EVOLUTION

Experimental studies have been foundational in the study of bacterial ecology and evolution, but many advances into coevolutionary processes have also been made with theoretical and computational approaches. Computational approaches are a powerful tool for studying complicated phenomena such as ecological and evolutionary processes due to the ability to consider scales and levels of generality not possible in experiments. Both theoretical and simulation-based approaches have been widely used to understand the fundamental

processes and behaviors that underlie many features of natural systems. The dimensionality and complexity of coevolution makes it a particularly appealing topic to apply theoretical and computational methods.

In particular theoretical approaches have been often applied to study the emergence of cooperation. A major tool has been the use of simple games such as the prisoner's dilemma. The prisoner’s dilemma involves two players each making a single binary choice, to cooperate or to defect. The best global outcome occurs when all players cooperate, but cheaters have an advantage over naïve cooperators. Despite this simplicity complex dynamics arise when considering which strategies are superior, and under what conditions cooperation is encouraged. Simulations have been used to identify the best strategies for playing the game in a mixed population of competitors and examine which conditions allow cooperation to succeed [42,43]. The Prisoner’s Dilemma and Snowdrift, a similar game, have been used to study the conditions that facilitate the emergence of cooperation such as spatial structure and iterative encounters [44,45].

Games like the prisoner’s dilemma are powerful in their simplicity, but are so abstract that it can be hard to generalize findings from them to evolution in specific real communities. Other studies have studied bacterial ecology and evolution using models that account for the specific lifestyles and behaviors of bacteria. Systems of differential equations for modeling bacterial growth have been used to study the evolutionary dynamics of cooperative cross-feeding between bacteria [46,47]. Agent-based modeling methods have similarly been used to simulate the growth and interactions of communities of bacteria [48,49]. Such methods are powerful in their flexibility, but limited in their mechanistic detail to what is believed to be important for interaction.

1.5 Constraint Based Modeling with Flux Balance Analysis

Constraint based metabolic modeling is a powerful approach that incorporates genomic content and the metabolic network without requiring the detailed kinetic information used in some dynamic models. Flux balance analysis (FBA) is a popular method that can simulate the growth of a cell using a genome scale metabolic model of the organism [50]. FBA finds a solution of fluxes through a metabolic network that optimized some objective function, which is often biomass production, equivalent to growth rate. The method works on the assumption that the internal abundances of metabolites are at a steady state, creating the constraint that fluxes through all reactions must be balanced so that there is no change in internal metabolites, only exchange of metabolites with the environment (both consumption and excretion) and production of biomass. In this method a model of a cell consists of all the metabolic reactions that can be catalyzed in the cell and the stoichiometry of each reaction, with upper and lower bounds on the flux through each of those reactions. It is assumed that enzyme abundance is high enough to not be limiting for any possible flux between the bounds. Reaction stoichiometries and bounds are formulated into a matrix and vectors, respectively, that are used as inputs for a linear programming solver to find the optimal growth solution. FBA is a flexible tool that has been extended to variations on objective functions [51], approaches that account for and describe the uncertainty in predictions [52], and extensions that incorporate gene regulation [53].

FBA has previously been used to study issues in ecology and evolution of microbes. Sometimes FBA is used to mechanistically study growth in a small community of interest, for example FBA has been used to model mutualistic growth between two species [54]. Other studies have exploited the power of FBA to quickly assay growth phenotype in order to examine large scales. For example, one study characterized the prevalence at which co-culturing two

species results in production of a metabolite that neither species would produce on their own, and looked for such instances across all pairs of six species across one hundred different media [55]. Another study examined thousands of media compositions to find ones that induced synthetic obligate cross-feeding relationships between pairs of bacteria [56]. FBA was used to assay a large scale of pair-wise growth between different bacterial species to identify competition and cooperation [57]. FBA was used to simulate in detail the divergence of genotypes within a single co-culture at a fine temporal scale [58]. Another study used FBA to model the long term evolution of an endosymbiont. Starting with a model of a close extant relative (*Escherichia coli*), they randomly and continually removed genes that were not vital for growth. The resulting minimal genomes compared favorably to actual endosymbionts in terms of genome content.

1.6 METABOLIC MODELING AS A TOOL TO STUDY BACTERIAL COEVOLUTION

Constraint based modeling has already proven useful for studying bacterial ecology and evolution, but there is more that can still be done. In the chapters that follow I have taken advantage of the power of FBA and applied it to new coevolutionary scenarios to study large scale questions in bacterial coevolution. The evolution of endosymbionts has been previously studied, but more could be done to study how specifically metabolic cooperation between diverging strains arises. In Chapter 2 I model the coevolutionary process of two bacteria evolving reductively, such as the conditions experienced by endosymbionts. Using the scale afforded by FBA I simulate thousands of evolutionary trajectories and study the frequency at which obligate cross-feeding interactions emerges between the two species, and how genomic and ecological factors influence the emergence. Most evolution does not occur in such constrained circumstances with such relaxed selection, and so in Chapter 3 I conversely examine the opposite situation of a more complex community at a short time scale and use simulations to

explore how the fitness consequences of all possible mutations in a focal species vary depending on what other species are present in the same community.

.

Chapter 2. METABOLIC MODEL-BASED ANALYSIS OF THE EMERGENCE OF BACTERIAL CROSS-FEEDING VIA EXTENSIVE GENE LOSS

This chapter is based on the manuscript:

McNally CP, Borenstein E. Metabolic model-based analysis of the emergence of bacterial cross-feeding via extensive gene loss. BMC systems biology. 2018 Dec;12(1):69.

2.1 SUMMARY

Metabolic dependencies between microbial species have a significant impact on the assembly and activity of microbial communities. However, the evolutionary origins of such dependencies and the impact of metabolic and genomic architecture on their emergence are not clear. To address these questions, we developed a novel framework, coupling a reductive evolution model with a multi-species genome-scale metabolic model to simulate the evolution of two-species microbial communities. Simulating thousands of independent evolutionary trajectories, we surprisingly found that under certain environmental and evolutionary settings metabolic dependencies emerged frequently even though our model does not include explicit selection for cooperation. Evolved dependencies involved cross-feeding of a diverse set of metabolites, reflecting constraints imposed by metabolic network architecture. We additionally found metabolic 'missed opportunities', wherein species failed to capitalize on metabolites made available by their partners. Examining the genes deleted in each evolutionary trajectory and the deletion timing further revealed both genome-wide properties and specific metabolic mechanisms associated with species interaction. Our findings provide insight into the evolution of cooperative interaction among microbial species and a unique view into the way such relationships emerge.

2.2 INTRODUCTION

Most microorganisms in nature do not live in isolation but are rather part of complex communities [59]. The various species that form these communities not only share a common environment, but rather interact with other community members in various ways including competition for extracellular nutrients, cooperation through metabolite cross-feeding, signaling, biofilm formation, and antimicrobial secretion [9,60]. Such interactions allow community members to impact each other's behavior and thus play an important role in shaping community structure and activity. A better understanding of how these interactions emerge through ecological and evolutionary dynamics, how they are maintained or lost, and how they impact community-level behavior is therefore crucial for both elucidating the forces that have shaped current natural communities and for designing synthetic communities or targeted modulation of natural communities [15].

Perhaps the most intriguing form of microbial interaction is interspecies cooperation. The prevalence of cooperative interaction is evident from the large number of microbes that cannot be individually cultured, suggesting that they are reliant on symbiotic interactions with other members of their communities [22]. In the context of metabolism, cooperation often takes the form of cross-feeding, where one species secretes metabolites that other species uptake and utilize. Indeed, metabolic cross-feeding has been found to occur in a wide variety of environments and between diverse species [61], often benefiting both partners [62]. For example, *Bifidobacterium* species in the gut microbiota regularly cross-feed fermentation products and partial digestion byproducts of polysaccharides to butyrate-forming bacteria [63]. Furthermore, evidence suggests that metabolic cooperation drives species co-occurrence in diverse microbial communities [26].

Importantly, however, metabolic cooperation is not limited to complex communities and has also been demonstrated in small two- or three-species communities, such as those occupying various insect hosts. For example, it was shown that the two endosymbionts that occupy a sharpshooter insect (and the insect host) each lack necessary steps of several amino acid synthesis pathways, and consequently only when the three organisms grow together can they synthesize the entire complement of amino-acids [64,65]. Similarly, it was shown that two endosymbiotic bacteria that inhabit a *Cicadoidea* host had recently diverged into 3 species, with metabolic complementarity between the two recently split lineages [37]. In such small communities, cooperation likely emerges not through selection (e.g., via the Black Queen hypothesis [66], where loss of metabolic capabilities and development of dependence has a selective advantage [16,67,68]), but rather by chance as a consequence of extreme genome reduction [35]. Indeed, most insect endosymbionts have extremely small genomes (some of which are the smallest bacterial genomes known) that are majorly reduced compared to their closest free-living relatives [40]. Moreover, such tightly coupled minimal metabolic systems, where two or three species strongly depend on each other for survival, can be viewed as an idealized model of microbial cooperation and provide insight into the evolution of cooperative interactions [69].

Notably, however, despite the prevalence and diversity of cooperative endosymbiont systems, the process through which extensive long term genome reduction leads to metabolic cross-feeding and the mechanisms involved in such evolution are not clear and are challenging to study. Experimental evolution studies, for example, have demonstrated the emergence of cooperative interactions between divergent polymorphic sub populations [70,71], but are generally limited in both the duration of evolution and the number of replicates. These limitations hinder a systematic and comprehensive study of long term genomic reduction or

identification of general principles in the emergence of species interaction. On the other hand, theoretical and computational models have been broadly useful for studying microbial communities [46,72], and indeed several recent studies have used computational models to specifically address the evolution of cooperation [73,74]. Such studies have allowed long time scales to be easily modeled and have produced useful insights into genetic and environmental determinates of cooperation, but tend to explicitly model interaction in a non-mechanistic manner. Such models may therefore fail to capture the mechanisms underlying metabolic cross-feeding and the processes through which genome reduction can give rise to such mechanisms.

To address this gap, in this study, we utilized a model of microbial evolution over a *long time scale* coupled with a *mechanistic model* of multi-species microbial metabolism and growth. Our model is inspired by a previous study that modeled reductive evolution of a single endosymbiont species and investigated how historical contingency and timing of gene deletions affects future genome reduction [75,76]. In the work we present here, we extended this evolutionary framework to a co-culture model of two species, using a mechanistic model of microbial growth in co-culture based on a multi-species genome-scale metabolic modeling approach [55]. Such a framework allows us to simulate long evolutionary trajectories, to investigate metabolic mechanisms on a genome-level, and to generate a large number of simulated trajectories for inferring general principles in the evolution of metabolic interaction.

We specifically aim to examine whether species interaction can emerge in a simple multi-species community without explicit selection for it, which mechanisms can drive a selfishly evolving species to support a dependent species, and how the architecture of the metabolic and genetic networks affects the evolution of such interactions. Notably, we do not model the process by which an evolving population bifurcates into multiple subpopulations, but rather explicitly

assume the community harbors two evolutionary isolated species (e.g., following an initial split), each undergoing an extreme reductive evolution process (such as the one experienced by insect endosymbionts). This assumption allows us to examine evolutionary trajectories and species interactions between two well-defined linages in a fixed community context (and see also Discussion). We further assume that these two species co-exist over a long time scale, without one outcompeting the other. It is also important to note that we do not necessarily aim to model the evolution of any specific species or community, nor the evolution of any specific metabolic pathway, but rather to examine general principles and patterns that may be observed when the evolution of species' metabolic networks are driven by such a reductive evolutionary scheme. Using this framework, we simulated thousands of independent evolutionary trajectories, tracked the emergence of metabolic cross-feeding, and carefully analyzed the evolving species. Our findings shed light on the evolution of species interactions and could inform future effort to construct stable microbial communities for medical, agricultural, and industrial applications.

2.3 RESULTS

2.3.1 *A Framework for Modeling the Evolution of Species Interactions*

To study the emergence of metabolic species interaction in bacteria we developed a computational framework that integrates models of microbial coevolution, metabolic activity, and ecological interaction (Figure 2.1). Briefly, in this framework, we model a community comprised of two generalist species growing in a shared environment (and that can therefore exchange metabolites) that go through a reductive evolution process. In our model, evolution is an iterative process (as in [75]) in which a gene is first chosen at random from either of the two species for deletion (Figure 2.1A). The fitness effect (measured as the change in growth rate) of losing that gene in the context of the community is calculated using a co-culture metabolic model

(described below). If the decrease in fitness to the species losing this gene does not exceed a predefined threshold the deletion is assumed to fix; otherwise the deletion is assumed to be selected against and is reverted. Importantly, during the course of this coevolutionary process, the presence of each of the two species in the community can markedly impact the evolution of the other (and specifically, the set of genes that can be deleted). This process repeats until no more genes can be deleted from either species.

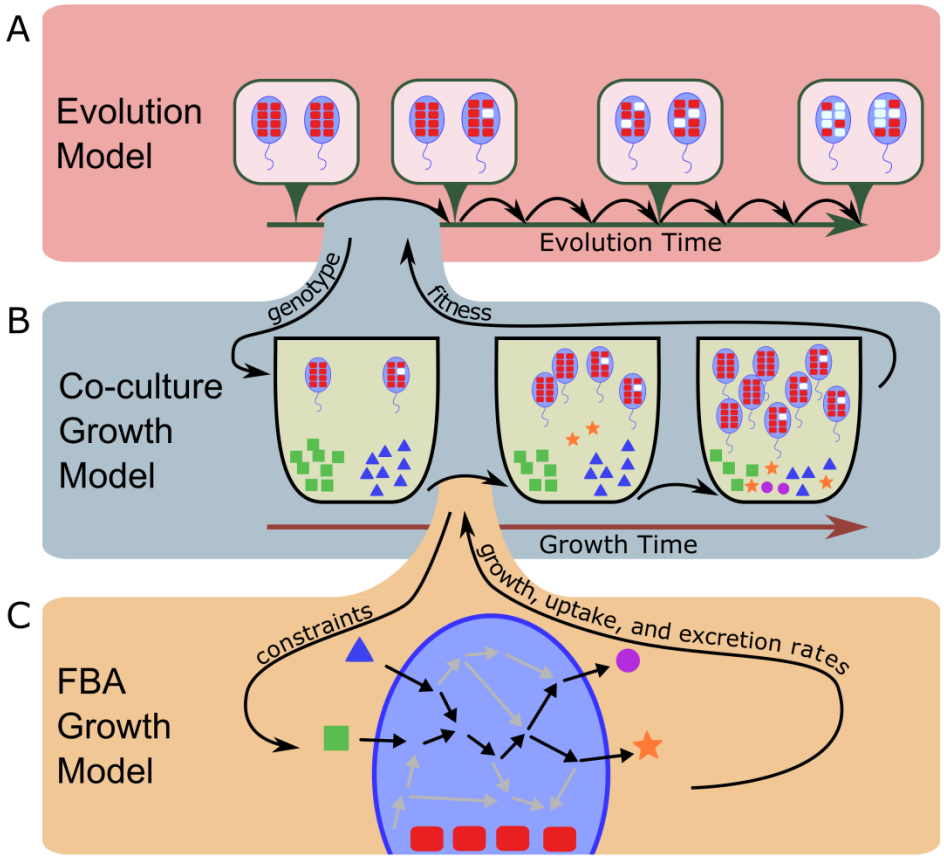

Figure 2.1. A framework for modeling the evolution of species interaction.
(A) To model reductive evolution, genes are iteratively chosen at random as candidates for deletion, the fitness effect of their deletion is evaluated (using a co-culture growth model; panel B), and if the fitness effect is relatively small, these genes are deleted. (B) The co-culture growth model simulates the growth of the two species in a shared environment, and is based on a previously introduced dynamic multi-species model (28). This model iteratively infers the behavior of each species in the shared environment based on an FBA approach (panel C). The predicted growth of each species and the predicted rates at which it uptakes and excretes various metabolites are used to update the abundances of species in the co-culture and the concentration of metabolites in the shared environment over time. (C) An FBA model is used to predict the growth of each species in a given environment based on the set of metabolic reactions and constraints encoded by the species and the concentration of metabolites in its environment.

To model growth in co-culture and to determine the fitness consequence of gene deletions while accounting for the way the presence of one species in the community may impact the fitness of the other, we used a co-culture metabolic modeling framework [55]. This framework employs dynamic Flux Balance Analysis (FBA) [77] to predict the metabolic activity and growth of two species in a shared environment over discrete time points (Figure 2.1B). At each time point and for each species in the co-culture, this framework uses a genome-scale metabolic model of the species (based on its metabolic capacity as determined by the set of genes present in its genome), the current concentration of metabolites in the environment, and a flux balance analysis to predict the species' behavior, including its growth rate and the rate at which it imports and excretes various metabolites [50] (Figure 2.1C). The estimated growth rates of the two species are then used to update the abundances of the species in the community and the predicted uptake and excretion fluxes are used to update the concentration of metabolites in the shared environment. Growth is simulated over several time points and the growth rates of each species at the last time point are used as proxies for their fitness. This co-culture model accounts for gene loss only through the resulting loss of metabolic capacities, and therefore includes only metabolic genes and ignores the potential consequences of loss of other genes.

To classify the interaction between the two species in each community and at each evolutionary step, we also simulated and evaluated the growth of each of the two species in isolation (i.e., in mono-culture). We define a species as being dependent on its partner if it can grow in co-culture but not in mono-culture. We further distinguish between three possible types of interaction a given community can exhibit: (i) *Independent* – neither species is dependent on the other, (ii) *commensal* – one species (*'dependent'*) is dependent on the other species but the other (*'provider'*) is not, and (iii) *mutualistic* – both species are dependent on each other. Note

that we use the terms ‘commensal’ and ‘mutualistic’ to describe the presence/absence of dependency and whether dependency is unidirectional or mutual, ignoring the more subtle distinction (and pertaining ecological definitions) of whether species harm each other or not. The observed interaction type at the end of the simulation run (i.e., when both species reach minimal genomes) was used to label each evolutionary trajectory (Figure 2.2A). Notably, in some cases, one of the two species can go through a catastrophic drop of fitness (>50%) even in co-culture (e.g., due to a change in the *other* species’ behavior that limits the availability of a metabolite it requires). In such cases, that species was considered to have gone extinct and the simulation was labeled as a collapsed community. A detailed description of the framework is provided in Methods.

2.3.2 *The Emergence and Prevalence of Metabolic Species Interaction*

We used the framework described above to simulate 16,317 independent evolutionary trajectories of a simple community comprising two generalist species that go through a reductive evolution process (see Methods). As noted above, we assumed that the community composition is fixed as a two genotype community, with no new species migrating into the community and no standing genetic diversity. In each simulation, we initialized the community with two identical *E. coli* strains (as a generalist model species [75]), representing the evolution of two obligate symbionts that may have diverged from a common ancestor [37].

Surprisingly, although our framework does not impose an explicit pressure toward species interaction, we found that a substantial fraction of simulations resulted in a community with some sort of metabolic dependency between the two species. Specifically, 35.3% of the simulations ended with a commensal community, and 3% of the simulations ended with a

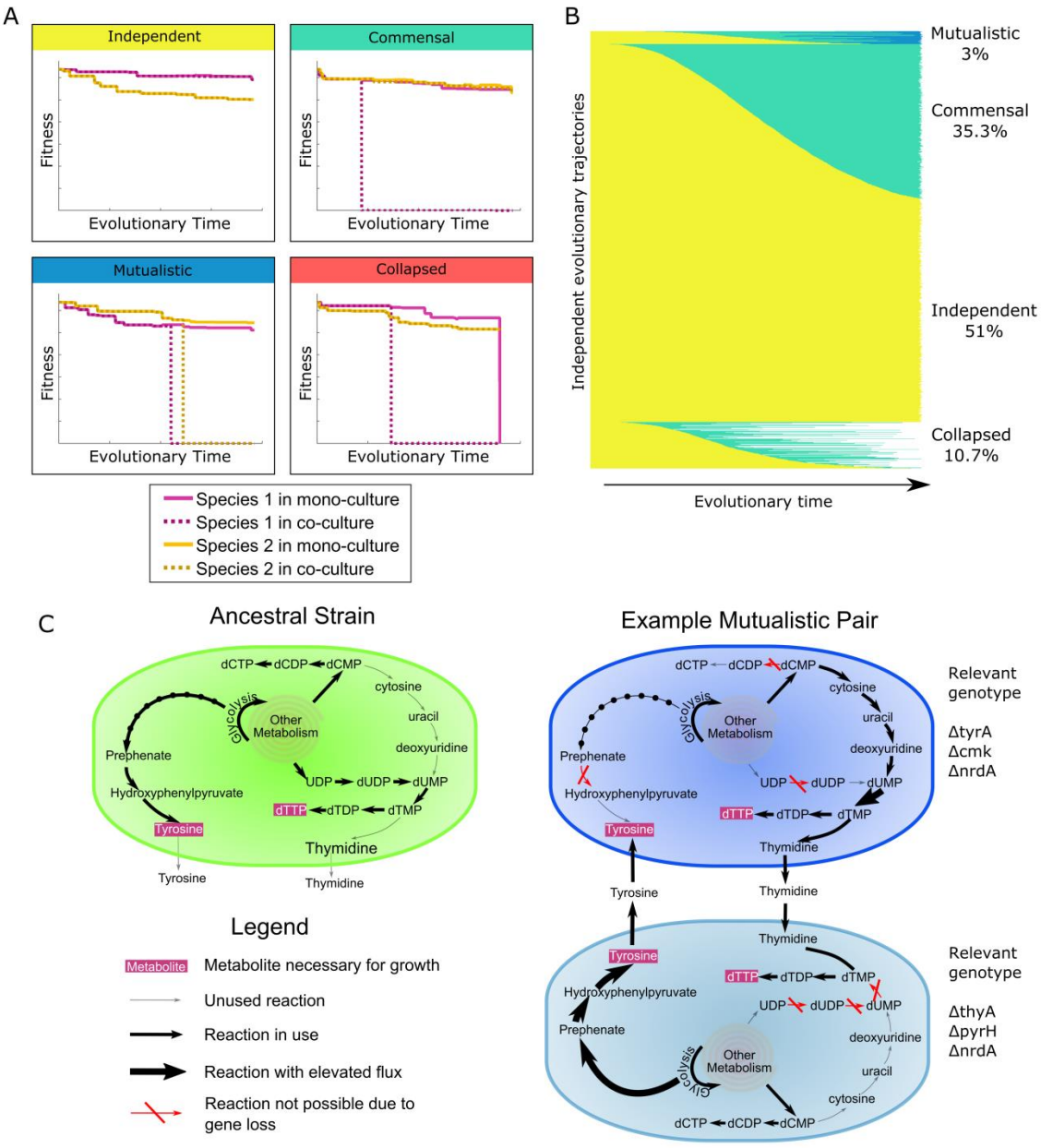

Figure 2.2. Metabolic interactions and their emergence over time.
(A) Evolutionary simulations could result in one of four unique outcomes, determined by the ability of evolved species to grow in mono-culture and co-culture. Plotted are examples of each of these four outcomes, illustrating the fitness of each of the two species in mono-culture and in co-culture over evolutionary time. (B) The changes in interaction type over time for all 16,317 simulation runs. Each horizontal bar represents a single simulation run, and the color corresponds to the interaction type using the same colors as the titles in panel A. (C) Example of an evolved mutualistic community. In the ancestral species tyrosine is produced through the shikimate pathway and dTTP is produced from UDP. In this example evolved mutualistic community, deletions in both species have led to obligate cross-feeding of tyrosine and thymidine. The relevant gene deletions and their impact on metabolic fluxes in each species are highlighted.

mutualistic community (Figure 2.2B). In 10.7% of the simulation, the community collapsed as described above. The remaining 51% of simulations ended with independent communities.

We additionally examined the impact of various simulation settings on the prevalence of different interactions types. We first tested the effect of using different fitness cutoff values. In natural communities the strength of selection against deleterious gene deletion reflects multiple factors, ranging from population size to environment stability, which therefore indirectly affects the likelihood of emergent cooperation. Indeed, we found that using different fitness cutoffs for allowing deleterious gene deletions to fix affected the ratio of the different interaction types, with a more stringent cutoff resulting in more independent communities and a less stringent cutoff resulting in more commensal and mutualistic communities (see Appendix A: Supporting Text). We next examined the impact of evolving communities in rich rather than minimal media. A richer media may hinder the emergence of species interactions since useful metabolites could be obtained from the environment rather than via cross-feeding. Exploring the impact of several types of richer media, including media with additional carbon sources or with various sets of amino acids, we found that indeed such media generally increased the prevalence of independent communities, though the specific effect was different for different media types, with varying balance between commensal, mutualistic, and collapsed communities (see Appendix A: Supporting Text). Finally, we examined the impact of different gene deletion strategies, including using different gene loss rates by the two species or allowing deletion of more than one gene at a time. We found that these strategies generally had little impact on the prevalence of various interaction types, although deleting multiple genes at each iteration did increase the prevalence of collapsed communities (see Appendix A: Supporting Text).

2.3.3 *An Example of an Emergent Cross-Feeding Interaction*

Before exploring large-scale patterns concerning emerged mechanisms involved in species interaction, we set out to characterize in detail one evolved mutualistic community as an example of the kind of metabolic interaction that could emerge and the gene deletions that underlie such an interaction. In this community, the two species (arbitrarily referred to below as species A and species B) had retained only 306 and 304 genes respectively, compared to 1260 genes in the ancestor species. Per our definition above, these two minimal species could still grow in co-culture (albeit at only 78% and 73% of the ancestor's growth rate, respectively), but neither could grow in mono-culture. Analyzing the evolved metabolic dependency of these species (Methods), we found that species A became dependent on tyrosine (and could grow on the initial medium once tyrosine was added) and that species B became dependent on thymidine (and similarly could grow on the initial medium once thymidine was added).

We further examined the fluxes through the metabolic models of the evolved species and compared them to the fluxes observed in the ancestor species, to identify the specific gene deletions that gave rise to these dependencies (Figure 2.2C). We found that species A became dependent on external tyrosine due to a loss of the gene *tyrA*, which is necessary for tyrosine synthesis [78]. Indeed, species A's loss of *tyrA* occurred at the exact same point in the evolutionary trajectory as its loss of the ability to grow in mono-culture. Similarly, Species B became dependent on external thymidine due to a loss of the gene *thyA*, which is necessary for dTMP synthesis [79].We were also able to identify the evolved mechanisms that allowed each of the two species to excrete the metabolite necessary for growth of the other species. Specifically, species A started excreting thymidine due to a loss of the gene *cmk*, which is necessary to phosphorylate CMP to CDP [80]. The loss of several other reactions prevented species A from

converting CMP to cytidine, uridine, uridine monophosphate, excreted uracil, or thymine, which resulted in species A only being able to eliminate excess CMP by converting it to thymidine and excreting it. Notably, a *cmk* deletion in *E. coli* has been shown experimentally to result in 30-fold elevated CMP and dCMP pools relative to wild-type [80]. Species B similarly excreted tyrosine due to an overproduction of this metabolite following a complex combination of gene losses that resulted in elevated activation of the pentose phosphate pathway and converting excess erythrose-4-phosphate into tyrosine. This example highlights the complex mechanisms that may be involved in the evolution of metabolic species interaction and how the architecture of the metabolic network could facilitate such interactions.

2.3.4 *Metabolite Cross-Feeding and Dependency in Evolved Pairs*

After characterizing one cooperating pair in detail, we set out to examine the complete set of communities evolved by our model, focusing initially on identifying the metabolites underlying emergent species interactions (Methods). We found that the majority of dependent species (94.3%) required only a single essential metabolite to be cross fed from their partner, with only a small fraction of dependent species requiring two or three such metabolites (5.6% and 0.2% respectively), and no species requiring more than three. Formate was the most common essential metabolite (68.7% in commensal dependent species; Figure 2.3), followed by Tyrosine (18.8%) and Phenylalanine (6.9%). Notably, the dependence on a single (or very few) metabolites reported above contrasts observations made in several insect symbionts systems where cooperating symbionts exchange multiple essential compounds (and see Discussion below), yet the exchange of aromatic amino-acids is in agreement with cross-fed metabolites often observed in such systems [36,64].

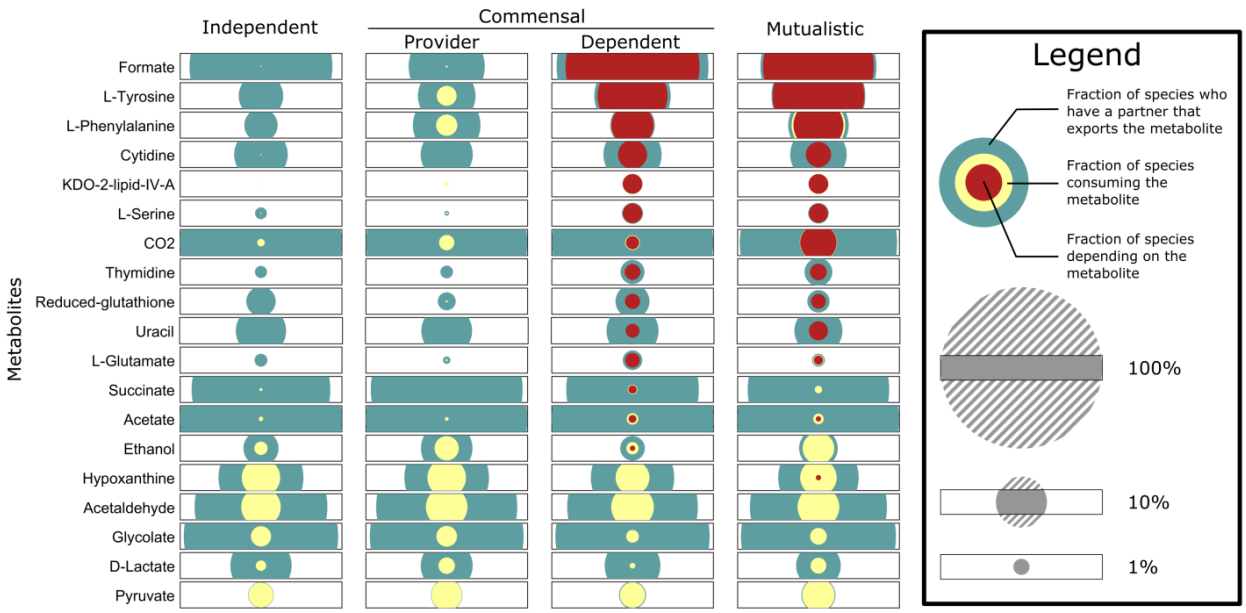

Figure 2.3. Frequencies of metabolites' availability, cross-feeding, and dependence.
The frequencies of metabolites' availability, cross-feeding, and dependence are shown for species of each interaction type and for each metabolite. Each set of nested circles shows the frequency at which the given metabolite is produced by their partner species and hence available for uptake (blue), the frequency at which this metabolite is utilized by the species through cross-feeding (yellow), and the frequency at which this metabolites is depended on (red). The area of the circle scales with the frequency, but for visualization purposes the portions of the circle extending beyond the rectangular box are not shown. Only metabolites that are depended upon at least 10 times or consumed at least 30 times are shown.

Complete dependence on cross-fed metabolites (such as those identified above) is the most defining feature of species interactions in these communities, but may represent an extreme form of interaction. Clearly, cross-feeding can be beneficial to a species even when it is not essential for growth, and in fact this form of cross-feeding may be a common precursor state of complete dependence. Examining secretion and uptake fluxes (Methods), we indeed identified multiple metabolites that are being cross-fed but are non-essential to growth (Figure 2.3, yellow circles). Many of these non-essential cross-fed metabolites were rarely if ever depended on (such as acetaldehyde and pyruvate) and were observed at similar frequencies across all interaction types. Interestingly, however, we also detected non-essential cross-feeding of metabolites that were commonly depended upon (such as tyrosine and phenylalanine; Figure 2.3), but these were

rare in independent communities and occurred surprisingly often in commensal communities where providers were cross-fed such metabolites by the dependent partner. This finding suggests that species cooperation may involve two species that evolve a similar metabolic strategy (and therefore have the potential to both excrete and utilize a similar set of metabolites). In such cases, cross-feeding is likely to emerge, first as a non-essential process, which may later evolve into species commensal or mutual dependence. To confirm this hypothesis, we specifically examined, for each dependent species, the time that elapsed from when this species started consuming a metabolite via cross-feeding to when it became dependent on that metabolite. In most cases dependence does not immediately follow cross-feeding, and there is often a substantial delay between cross-feeding and dependency (Figure A.1).

Clearly, uptaking a metabolite is only possible if the partner species is producing that metabolite and excreting it to the shared environment, thereby providing an opportunity for cross-feeding. We additionally quantified the frequency and time at which such opportunities arose, regardless of whether the metabolite was consumed or not (Figure 2.3, blue circles). We found that metabolites vary greatly in the frequency at which they are excreted, and in a way that is not fully correlated with the frequency at which they are cross-fed or dependent on. For example, various metabolites, including cytidine, succinate, and acetate, are excreted at relatively similar frequencies in all interaction types, suggesting that dependency on these metabolites is not limited by their availability. Conversely, other metabolites such as serine and thymidine are rarely excreted in independent communities, suggesting that the availability of these metabolites often leads to cross-feeding and dependency on them. Most importantly, while cross feeding often started almost immediately after the metabolite was available (in cases in which it occurred; see Figure A.1), in many cases evolving species failed to utilize available metabolites,

thus completely missing cross-feeding (both essential and non-essential) opportunities (Figure 2.3). This finding implies an intriguing dichotomy where available opportunities are either utilized immediately or are not utilized at all, potentially due to evolutionary constraints.

2.3.5 *The Genomic Basis of Evolved Species Interactions*

Our mechanistic model of microbial metabolism allows us to move beyond a phenotype-level description of evolved communities and to directly investigate patterns of genome evolution and identify genomic mechanisms involved in species interactions. We first examined the number of genes that were retained or lost in different simulations to explore the relationship between genome size (in terms of the number of genes retained) and species interaction. Surprisingly, with the exception of collapsed communities, evolutionary trajectories exhibited a markedly low variation in the total number of genes retained, with an average of 297.6±4.4 genes retained in each species. Yet, we found that both dependent and mutualistic species had slightly but significantly smaller genomes compared to independent species ($P < 10\text{-}30$ and $P < 10\text{-}9$ respectively; two sample t-test; Figure 2.4A), while provider species had slightly but significantly larger genomes than independent species ($P < 0.001$). Similarly, within commensal communities, the genomes of dependent species were slightly but significantly smaller than the genomes of provider species ($P < 10\text{-}30$). Moreover, while these differences in average genome size were generally very small (often less than a single gene; see Discussion), the differences between the smallest genomes observed in the dependent or mutualistic species and the smallest genomes observed in provider or independent species was much larger (Figure 2.4A). These results are consistent with the idea that cross-feeding allows dependent species to lose genes they would not be able to lose otherwise [65,81]. Interestingly, provider species had on average

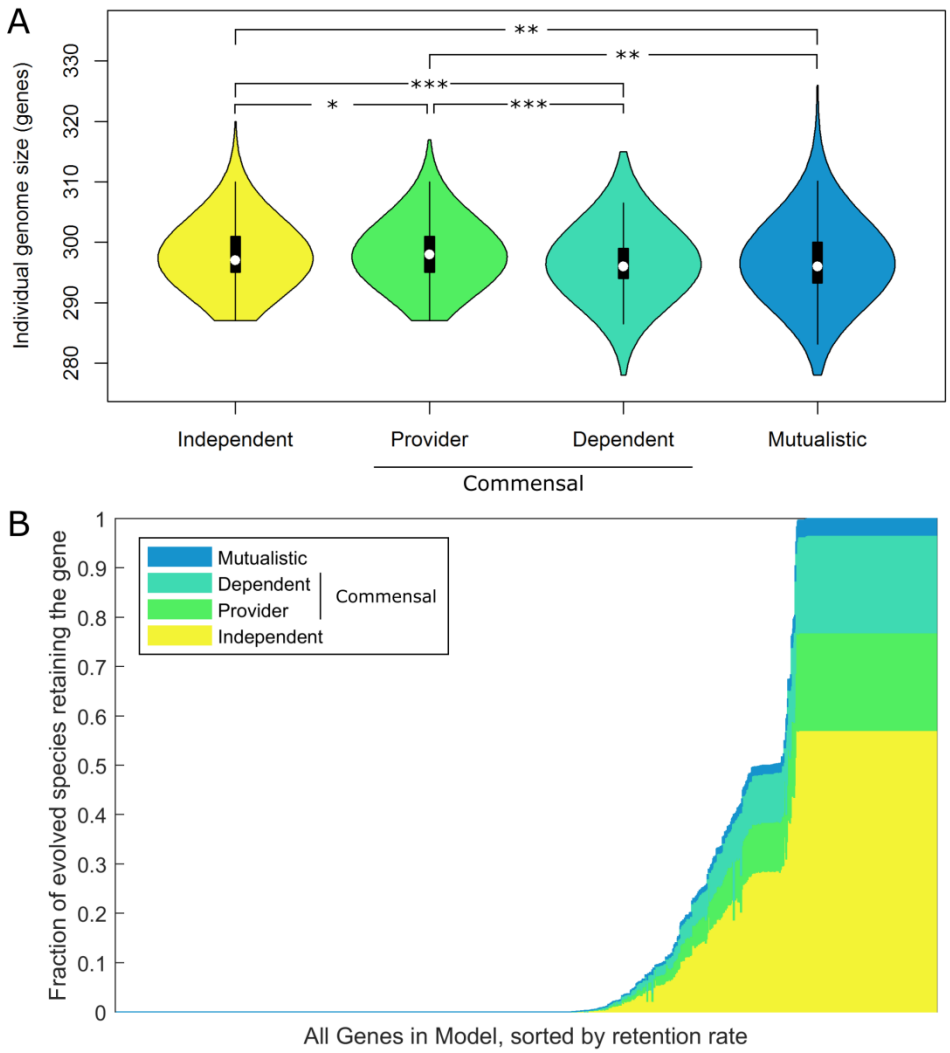

Figure 2.4. Genome size and gene retention frequency in evolved genomes.
(A) Distributions of the genome size of evolved species from each interaction type. (*: P < 10-3 ; **: P < 10-9 ; ***: P < 10-30). (B) The distribution of retention rates of different genes in the model. Each gene is plotted as a vertical bar with height equal to the fraction of species in non-collapsed simulations that retained it, and the genes are sorted in ascending order of this overall retention rate. Each bar is colored by the fraction of species retaining that gene that are in each of the different interaction types.

slightly larger genomes than independent species (P < 10-3). This could suggest that provider species retain additional genes (i.e., that are generally lost in independent species), likely as a result of constraints associated with earlier gene deletions. Such genes and the extended metabolic capacities with which they endow provider species may in turn give rise to metabolite overproduction. Alternatively, additional genes may be retained by provider species to allow

overproduced metabolites to be excreted outside the cell. Additional simulations further demonstrated that the genome size of independent species was similar to the genome size of species evolving in mono-culture conditions (see Appendix A: Supporting Text).

Notably, examining the genes retained across the evolved species, we found that of the initial 1260 genes present in the ancestral species, 560 were always lost and 149 were always retained, with only 551 genes being retained at intermediate frequencies (Figure 2.4B). The specific subset of these 551 genes that were retained in each evolved species therefore determines the types of interaction that emerged, and indeed a statistical analysis was able to detect specific genes whose retention or loss was associated with specific types of interaction (Figure 2.5A, Appendix A: Supporting Text). Further examining how similar, on average, are the sets of genes retained between the two partners in each community, we also found that cooperating partners (i.e., from mutualistic or commensal communities) were less similar to each other than independent partners, suggesting that the evolution of metabolic dependency is associated with a process of functional diversification (Appendix A: Supporting Text).

To better understand the diversification between partners in commensal communities (in which the two species can be labeled clearly as dependent and provider and therefore the direction of dependency is clear), we compared the set of genes retained in providers vs. those retained in dependents, identifying 80 genes that are more frequently retained in the provider and 41 that are more frequently retained in the dependent species. For example, pflA, a pyruvate formate lyase, was retained in 55.5% of providers but only 16.8% of dependents, whereas aceE and aceF, both components of the pyruvate dehydrogenase complex, were retained in 81.1% in dependent species and only 19.0% in provider species (these genes were also those with the greatest differential retention rate). We additionally identified a set of 263 gene pairs that are

significantly exclusively-retained in commensal communities (i.e., the dependent species retained the first gene when the provider lost the second gene or vice-versa more often than expected by chance; Methods). Interestingly, this set was enriched for gene pairs that shared a pathway annotation ($P < 10{-}4$; permutation-based test), suggesting complementation at the pathway level. To finally examine the dynamics of gene deletion events in these commensal communities and to reveal key evolutionary steps on the route to cross-feeding, we further used a permutation-based analysis to identify instances where a gene in one species tended to be deleted only after another gene was deleted in the partner species (Methods). Our analysis discovered several such ordered events, where a deletion leading to dependency on a given metabolite in one species could occur only after deletions that promoted over-production and excretion of that metabolites occurred in the second species (Appendix A: Supporting Text and Figure A.2).

2.3.6 *Linking Genome Evolution to Metabolite Cross-Feeding*

Having identified both the metabolites involved in cross-feeding and the genes involved in the emergence of species interaction, we finally turned to examine the association between specific gene retention or loss events and the cross-feeding of specific metabolites. Towards this end we again considered the set of all commensal communities and, for each of the 13 cross-fed metabolites that were depended upon in at least 10 communities, identified genes whose retention or deletion are significantly correlated with the excretion of this metabolite in the provider or with the dependency on this metabolite in the dependent (Methods). In total we identified 310 gene-metabolite associations (including 204 retentions and 106 deletions) in provider species associated with essential metabolite excretion, and 315 gene-metabolite associations (114 retentions and 201 deletions) in dependent species associated with metabolite dependency (Figure 2.5B; $\chi 2$ test, 1% FDR). In total, retention or loss of 194 of the 551 variable

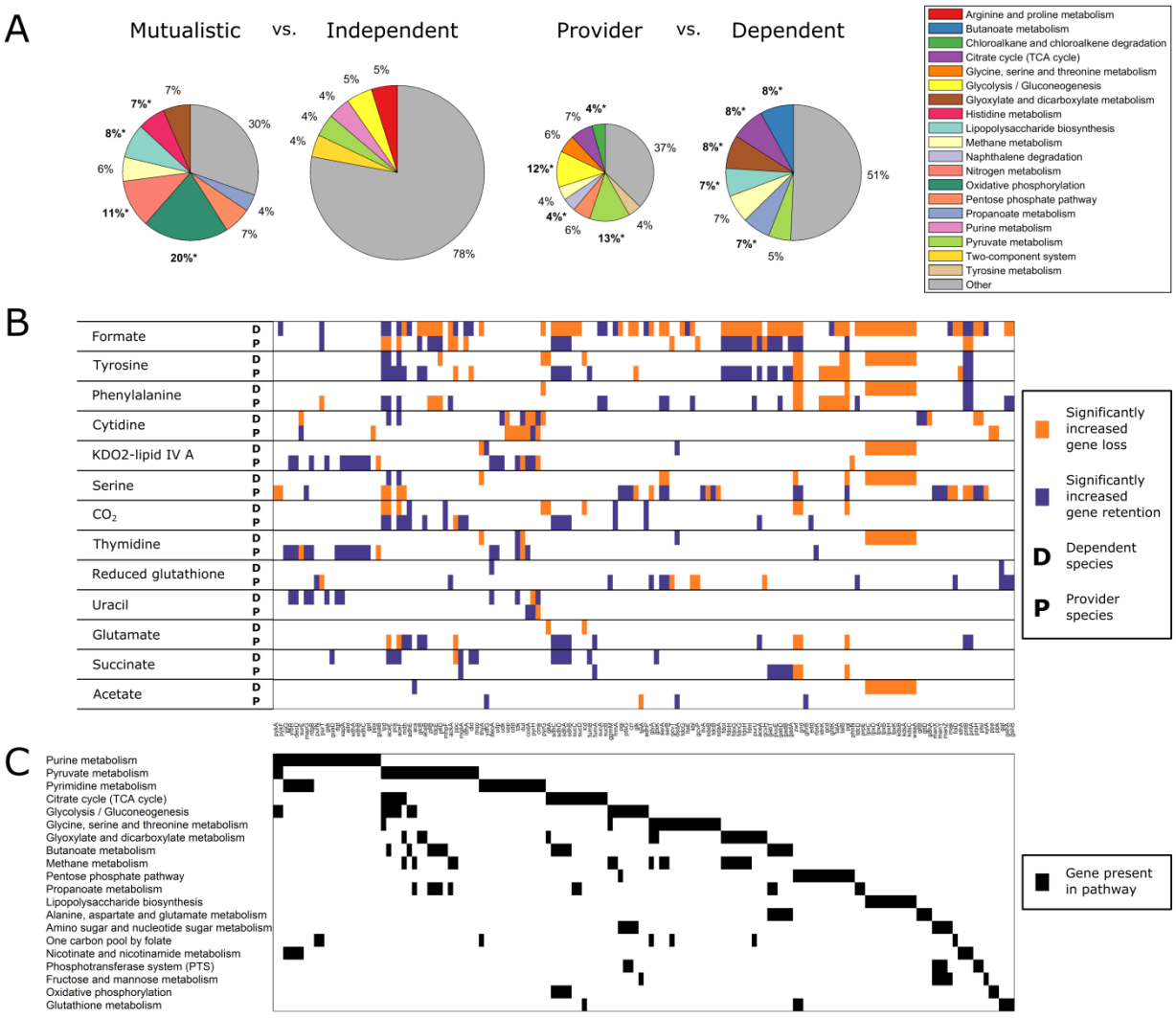

Figure 2.5. Genes and pathways associated with species interaction and cross-feeding phenotypes. (A) The fraction of genes whose deletion is associated with each interaction type that are assigned to each pathway. The overall size of the pie indicates the number of genes associated with that interaction type. Pathways that are significantly enriched among this set of genes are highlight (bold font and asterisk). Only pathways that contain >4% of the genes are listed (the rest are pooled into the 'Other' category) (B) Associations between providing or depending on specific metabolites and deletion or retention of specific genes. Orange (purple) bars indicate genes that are lost (retained) significantly more often in species that provide or depend on a specific metabolites compared to independent species. Genes are sorted by membership to pathways (see (C)), and only genes present in at least one pathway are shown. (C) Pathway membership of genes. Black bars indicate that a gene belongs to that pathway. Pathways are sorted by the number of genes assigned to them, and only pathways with five or more assigned genes are shown.

genes was significantly associated with excretion of or dependence on at least one metabolite, with many genes associated with the excretion of or dependence on multiple metabolites. This finding demonstrates how the interconnectedness of the metabolic network impacts the evolution of species cooperation, wherein the loss of a key gene could give rise to multiple metabolic phenotypes.

Interestingly, certain genes were associated with both the excretion of a metabolite by the provider and with the dependence on that same metabolite in the dependent species (Figure 2.5B). For example, retention of tyrP and aroP (aromatic amino acid transporters) were associated with both providing tyrosine and dependence on tyrosine, as cross-feeding of that metabolite required that both species in the pair could exchange it with the shared environment. In other cases the same gene was associated with a metabolite being excreted by the provider and utilized by the dependent but in a different direction. For example *serA* and *serB* – genes involved in serine biosynthesis – tended to be lost in species dependent on serine, but retained in species producing it. In rare cases there were genes whose loss was associated with both dependence and providing of a specific metabolite. For example, the loss of the gene *codA* was associated with both excretion of cytidine by providers and dependency on cytidine in dependent species (and see our analysis of that gene above). Combined, these associations suggest that metabolite cross-feeding can evolve via multiple mechanisms and that such mechanisms are often metabolite- and gene-specific.

Finally, to further explore metabolic mechanisms that may be involved in cross-feeding, we examined the pathways to which genes associated with various metabolic phenotypes are assigned (Figure 2.5C). Many pathways reflect expected associations, such as the pyrimidine metabolism pathway (that includes many genes whose loss or retention is associated with

cytidine cross-feeding) and the pyruvate metabolism pathway (that includes many genes associated with formate cross-feeding). In other cases the link between deletion of genes from a specific pathway and metabolic cross feeding is less obvious. For example, cross-feeding of thymidine and uracil, both pyrimidines, is associated with the deletion/retention of several genes in the pyrimidine metabolism pathway but also with a number of genes in the purine metabolism pathway. This could suggest that cross feeding of these metabolites is driven by a metabolic overflow that originates in the purine metabolism pathway but results in excess nucleotides being converted to pyrimidines before being excreted as waste. More generally, the association of genes from multiple pathways with cross-feeding of a single metabolite further highlights how the interconnectedness of the metabolic network gives rise to non-trivial links between gene deletion and cross feeding.

2.4 DISCUSSION

In this study we investigated the potential for metabolic interactions to emerge between two species inhabiting a shared, constant, nutritionally limited, and isolated environment and undergoing extensive gene loss. We found that cross-feeding interactions emerged frequently in these settings, and that such evolved communities exhibited diverse, multifaceted, and non-trivial metabolic interactions that were not necessarily optimized at the community level; useful metabolites were often excreted by one species but not utilized by its partner and other metabolites were cross-fed without evolving complete dependency. Such “suboptimal” interactions and missed metabolic opportunities are a reasonable outcome of selfish species evolution in the absence of explicit selection for interaction, and could also occur in natural communities. Another potential contributor to such suboptimal interactions is the interconnectedness of different metabolic phenotypes. This interconnectedness may also account,

for example, for the relatively frequent occurrence of provider species utilizing metabolites excreted by their dependent partners, where the gene retention and loss events that cause a dependent relationship in one direction may also facilitate emergence of a reciprocal cross-feeding relationship.

Indeed, our analysis has demonstrated that genes were often associated with the excretion and/or production of multiple different metabolites and that the excretion and/or production of each cross-fed metabolite generally involved combinations of multiple deletions. This 'many-to-many' mapping between gene loss events and emerged interactions is perhaps not surprising given the coupling between different pathways induced by the architecture of the metabolic network. With each gene deletion metabolic fluxes are redistributed across the network, ultimately rendering the link between gene deletions and cross-feeding of specific metabolites non-trivial and challenging to understand. This observation can also be viewed as an instance of genetic epistasis, where the deletion of two (or more) genes results in an unexpected and non-additive behavior that cannot be easily explained by the cumulative effect of the deletion of each gene in isolation. Moreover, in this context, our framework highlights an exciting extension of genetic epistasis, wherein unexpected and non-trivial interactions between gene deletions can be observed when the two genes are encoded by two different community members (and see [82]). Future work can further delineate and explore this form of multi-species, community-level epistasis.

Interestingly, in our simulations, metabolic *dependency* usually involved a single metabolite, while real world mutualistic endosymbionts often exchange and are dependent on multiple metabolites [64]. One potential explanation is that in our model bacterial growth is optimized (given the metabolic capacities encoded by their reduced genome), whereas in reality

extreme genome reduction likely impacts cell regulation and control of growth. Put differently, while generally a reduction in genome size is likely associated with a reduced set of metabolites the cell can synthesize, such disrupted regulation (e.g., via extreme loss of non-metabolic genes) may potentially cause cells to excrete a larger variety of the metabolite they still synthetize and these could be beneficial to their partners. This growth optimization may also account for the relatively small difference in genome size observed in our simulation between dependent and independent species. Our analysis also suggests that the likelihood of missed metabolic opportunities may vary across metabolites, with some metabolites (e.g., cytidine, succinate, and acetate) being excreted at relatively similar frequencies in all interaction types and others (e.g., serine and thymidine) being rarely excreted in independent communities.

Our findings additionally demonstrated how functional diversification leads to metabolic cooperation, where each species retains certain metabolic capacities that the other species has lost. Given a diversification process, it is interesting to speculate about what causes one community to evolve a commensal interaction and another to evolve a mutualistic interaction. We found, for example, that provider species had on average a slightly larger genome than independent species, suggesting that a provider state is the outcome of more constrained evolutionary trajectories that end with larger minimal genomes (*e.g.,* due to early gene deletion events that render other genes essential for growth). This could be the result of providers being forced to eliminate flux overflows through longer pathways that results in more useful waste metabolites being excreted. Moreover, examining the total number of metabolites being excreted by each species, we found that dependent species in fact tend to excrete more metabolites at the beginning of the evolutionary process (see Appendix A: Supporting Text and Figure A.3), potentially suggesting that early ‘wasteful’ metabolic strategies may contribute to the evolution

of dependence. Another interesting outcome of our results was the dichotomy observed when a new metabolite became available, with species either starting to consume it immediately and later becoming dependent on it or never consuming it at all. These missed opportunities seem to be examples where the evolutionary events that occurred before the availability of the metabolite precluded utilization of that metabolite by potentially losing the necessary transporter or other reactions necessary for uptake. With this in mind, the non-essential cross feeding observed in commensal communities may simply represent communities that were on the path toward mutualism, but where cross-feeding emerged too late in the evolutionary process when the providers have already lost genes that would be necessary for dependence. This suggests a role for historical contingency in the emergence of cross-feeding, though its importance compared to chance and the extent to which it limits accessible states in coevolution settings will need to be further examined [75,83].

Despite these exciting results, there are clearly some caveats in the framework used in this work. For example, our framework assumes that bacteria grow selfishly, and accordingly cross-feeding often requires extensive gene deletions to force excretion of cross-fed metabolites. In reality bacteria can be leaky and release metabolites into their environment even without mutations [15,66]. Alternative community-modeling methods assume that species evolve to optimize total community growth [54] or to simultaneously optimize their own growth and the community growth [84]. Such assumptions may not be evolutionarily reasonable in various settings but will likely result in markedly more prevalent cooperative behavior. Moreover, our modeling framework can only account for the function of metabolic genes, whereas in reality, a large variety of non-metabolic regulatory mechanisms could potentially impact the evolution of cross-feeding. Another drawback stems from the fact that FBA does not take into account factors

such as entropy or pH. For example, the emergence of formate cross-feeding that occurred in our simulations might be less biologically feasible because excess formate accumulation inhibits *E. coli* growth and acidifies the local environment [85]. Another key simplification underlying our work is the assumption of exactly two species that co-exist over a long time scale. This simplification is based on the small population size and drift-dominated evolution in insect endosymbiont systems. More importantly, however, this simplification facilitates many of the analyses reported above, as it enables a clear, rigorous, and well-defined quantification of the impact of one community member on the other. With only two genotypes, the causality and direction of symbiotic relationships is clear, evolutionary emergence of cross-feeding can be compared across trajectories in a fixed context, and the simulation of many replicates is computationally feasible. Finally, following the study reported in [75], our simulations all used *E. coli* as the evolutionary starting point and accordingly some gene- or metabolite-specific results identified in our study may be restricted to this species. We believe, however, that many key patterns observed in the study during the emergence of species interaction would generalize to other bacterial systems.

Future work can further extend our framework or explore the impact of various environmental and evolutionary settings on the emergence of species interactions. For example, most of our analyses above are based on simulating genome reduction as occurring one gene at a time [86], but do not account for the possibility of simultaneous loss of larger genomic regions [87]. Such a process could give rise to different patterns, and indeed our limited analysis of the impact of deleting pairs of genes in each iteration (Appendix A: Supporting Text) has demonstrated an increase in the prevalence of collapsed communities. Notably, however, a study of a single reductively evolving species that examined both evolutionary regimes did not observe

qualitative differences [75]. Similarly, we have demonstrated that using richer media (as opposed to the minimal media used in our primary simulation set to promote cross-feeding) or a different fitness cutoff could have a marked impact on the prevalence of different interactions (see Support Text). Future work can further explore the effects of different media compositions or of different limiting concentrations on the emergence of species interaction and on the evolving underlying mechanisms in a more systematic and comprehensive manner [56]. A potential future extension of this study could also aim to identify minimal sets of gene deletions that could still give rise to the obligate cross-feeding phenotypes observed in our study after extreme gene loss. Such reduced sets could be, in principle, explored experimentally to validate specific cross-feeding behaviors or more general trends observed in our study. It would also be interesting to expand our framework and to model the evolution of more complex communities (which we did not do here in consideration of simulation time) or to account for spatial heterogeneity [58,88].

Using our model, we identified frequent emergence of metabolic dependencies (under specific environmental and evolutionary settings) despite selfish evolution of each species, and we further identified genes and metabolites involved in such evolved cross-feeding interactions. Looking forward, the framework presented in this study could be broadly relevant for improving our understanding of how mutualistic relationships can naturally emerge between bacterial species. This, in turn, would facilitate a deeper understanding of both simple communities, as in the case of insect endosymbionts, and significantly more complex communities, such as those inhabiting the human gut. Moreover, translationally, our approach could be useful to aid and inform the design of dependencies between bacterial species in order to increase the stability and reliability of synthetically constructed bacterial communities or interventions.

2.5 METHODS

2.5.1 *Evolution and Growth Simulations*

Every evolution simulation was initiated with two identical copies of the iAF1260 *E. coli* model [89]. During each step in the evolutionary process, a gene (and all the metabolic reactions that depend on this gene) from one of the two species was selected randomly for deletion. If the fitness effect of this deletion in the context of the community (using the co-culture growth model described below) was smaller than the chosen cutoff (5%, as in [75]), the deletion became permanent and the process repeated with the reduced model. Otherwise, the deletion was considered too harmful to occur and the process repeated until a gene that could be deleted was found. The evolutionary process continued until no additional genes could be deleted. The co-culture growth simulation was based on a previously introduced dynamic flux balance analysis framework and is described in detail in Ref [55]. Briefly, given a multi-species community inhabiting a shared medium, the framework assumed that at each time step, each species grew optimally given the current concentration of metabolites in the medium, and then updated the abundance of each species and the concentration of metabolites in the medium based on the predicted growth and activity of each species. For the purpose of this study, both species started at a biomass of 0.01 grams dry mass in 1L volume for mono-culture or 2L for co-culture, resulting in the same cell density for both (which is equal to about $4x10^7$cells per liter for *E. coli*). The species were grown on a medium based on M9 minimal media [90]. A low concentration (0.0001 mM) of 'jumpstart' metabolites were also included to allow growth of obligate mutualistic pairs. Each co-culture simulation consisted of 8 steps of 0.125 hours followed by 4 steps of 0.5 hours. Additional co-culture growth simulations were performed on the resulting minimal models using a finer time resolution and until the medium was exhausted

to confirm that evolved interactions were stable and consistent. For a more detailed description of the evolutionary simulation, the co-culture growth model, and the media see Supporting Methods.

2.5.2 *Determining Interaction Types and Metabolic Dependencies*

Interaction type was determined by comparing the fitness of each species when grown in co-culture with its fitness when grown in mono-culture, to assess whether the species was dependent or independent. Communities were labeled as independent, commensal, or mutualistic based on the relationships between the two species. The metabolites (if any) a species depends on were determined by first identifying metabolites that were exchanged between the two species at the final co-culture growth time, and then assaying the growth of each dependent species on minimal medium supplemented with all possible combinations of these exchanged metabolites. If no combination of supplement metabolites allowed such growth the search was expanded to include all combinations of metabolites present in the medium at the end of the co-culture simulation and that were not part of the minimal media (to account for metabolites excreted by the provider at previous time steps). Simulations with ambiguous metabolite dependencies were excluded from metabolite analyses. Additional details can be found in the Supporting Methods (Appendix A).

2.5.3 *Analyzing Evolved genomes and Gene Retention/Deletion*

The Jaccard similarity coefficient was used to measure the similarity of two genomes (e.g., in an evolved community). A hypergeometric test was used for determining whether a pair of genes has been co-retrained significantly more or less often than expected by chance (at 1%FDR). This analysis was limited to gene pairs that were both retained and deleted at least

three times and genes were grouped into sets of perfectly co-varying genes to efficient calculation. Test for enrichment of shared pathways among significant gene pairs was done by permuting the links between pairs. To identify significantly common ordered pairs of gene deletions (at 1% FDR), the number of times gene A in the dependent was deleted before gene B in the provider was recorded and compared to the number observed once the times of deletion (i.e., the positions in the ordering of all gene deletions in that simulation) were permuted.

2.5.4 *Identifying Gene-Metabolite Associations*

To identify associations between retention or deletion of specific genes and metabolic phenotypes, the frequency of deletion of a given gene in commensal species that are dependent on a given metabolite was compared to the frequency of deletion of that gene in independent species. This was repeated for commensal species that provided the metabolite their partner depends upon. In both cases, genes being deleted more often or retained more often in species with that metabolic phenotype were identified (at 1% FDR). Only genes that were both deleted and retained at least 10 times and metabolites that were depended upon at least 10 times were considered. These cutoffs were chosen to restrict analysis to genes and metabolites for which sample size would provide statistical power to confidently identify significant correlations.

Chapter 3. EFFECT OF COMMUNITY COMPOSITION ON THE EVOLUTIONARY POTENTIAL OF MICROBIAL SPECIES

3.1 SUMMARY

The patterns and mechanisms by which the evolution of individual species depend on their ecological context is an area of interest, but past studies have not considered how the full fitness landscape changes across contexts. Here we use a computational model of microbial co-culture growth to explore how the adaptive landscape of an individual species changes as a function of other species inhabiting the same environment. We used dynamic flux balance analysis to model the growth of a community of representative human gut dwelling bacterial species in a shared environment. Specifically, using this framework, we measured the fitness of all possible single reaction deletion mutants of *Escherichia coli* in co-culture with all possible subsets of nine other species. We found that indeed, community composition can have a marked impact on the fitness effect of a given mutation. In some cases, for example, a beneficial mutation in one community composition was deleterious in another. Out of 1568 mutants tested, 135 showed variance in fitness effect across community backgrounds. We found that community context similarity was a strong determinant of fitness landscape similarity, but different community members had different relative importance. Increasing community diversity increased the evolutionary potential for *E. coli* to gain beneficial mutations, in accordance with several prior studies. Finally we investigated the mechanisms underlying this community-dependent fitness for each mutation. Gaining a better understanding of how evolutionary forces

act in microbial communities could inform future efforts to design robust targeted interventions in communities of clinical, environmental, and industrial interest.

3.2 INTRODUCTION

The evolution of a species can be influenced by both its abiotic and biotic environments. Adaptation of an individual species to its biotic environment has the potential to then alter community dynamics and the evolution of other species, resulting in complex feedback between ecology and evolution [91,92]. Despite this daunting complexity, a tractable starting point is the study of the adaptation of individual species to their specific biotic environment on short time scales. Here we will refer to the set of other species existing in the same environment as the community context. Adaptation to a community context is the result of selective pressures exerted by interspecies interactions. Many such types of interactions can lead to selective pressures, including competition, predation, symbiosis, and niche construction. In diverse communities the evolutionary dynamics may be dominated by pairwise interactions or by complex, non-additive selective pressures, called diffuse coevolution [93]. Previous research into coevolution has investigated systems such as the dependence of a sexually selected trait in an amphipod on the level of predation [94], the dependence of swimming speed in damselfly larvae on the level of predation [95], and the dependence of adaptive diversification of a bacteria in a new environment on the presence of a native bacterial community [96]. Such prior investigations have primarily examined adaptation of single traits to different ecological contexts.

A fuller understanding of the community-context dependence of the evolutionary potential of a focal species could be gained by considering the full fitness landscape in different community contexts. Such a fitness landscape, or adaptive landscape, would encompass the fitness effect of all possible mutations in a particular community context [97]. Studying how

fitness landscapes change as community context changes could provide more generalizable insight into patterns of coevolution. However, the throughput needed to measure the fitness of many variants in many community contexts is not feasible in the macro-organism systems that have been used for most previous research on coevolution. Simulation based approaches, on the other hand, allow easy manipulation of ecology and genotype across vast numbers of conditions. In particular, models of microbial metabolism can be used to simulate growth of multiple species in co-culture while perturbing gene content and community composition [98]. Such a system provides the scale and control needed to measure a fitness landscape across different community contexts.

Bacteria have importance far beyond just being a useful model system. Microbial communities are ubiquitous and play key roles in many important natural and artificial processes [2,6]. Evolution occurs within these communities on time scales relevant to human life and has the potential to alter or disrupt the functioning of such communities [99,100]. The high density and diversity within bacterial communities creates a massive scale of potential interactions between species and thus great potential for coevolution [9,60]. These interactions come in many forms, including secretions of toxins, quorum sensing, biofilm construction, and competition for metabolic resources, all of which may result in significant selective pressures [101]. Many of these mechanisms are not purely competitive but can also result in cooperation. For example, bacterial species antagonistically compete for limited metabolic resources in the environment but may also cooperate with each other when metabolic products excreted or leaked from one species can be beneficially consumed by another [26,61,62]. Such competitive and cooperative interactions have been shown to affect the evolution of bacterial species in lab conditions. One study found that forcing two species into a cooperative metabolic interaction results in extensive

adaptive mutations that increased the stability and productivity of this interaction while reducing independence [102,103]. Another study found that coevolving four wild species resulted in less antagonistic interactions and higher diversity of resource usage than when the species were evolved in mono-culture and later combined [104]. Finally coevolution has been found to result in reproducible ecological dynamics that were largely driven by dominant species that gained early advantageous mutations [25]. Further contributions to understanding mechanisms and patterns of coevolution in bacteria will aid understanding of evolution in complex microbial communities.

Many different computational frameworks have been used to study microbial growth and interaction, but one flexible model is the use of genome scale models paired with dynamic flux balance analysis (DFBA) to simulate growth of multiple bacterial species in a co-culture [50,77]. FBA has previously been shown to accurately predict bacterial growth and has been widely used [105,106]. Growth can be simulated quickly, and such models have proven useful for generating large scales of environmental and evolutionary perturbations [56,107]. These metabolic models are useful for studying evolution because they can explicitly link genotype (presence of genes and reactions in the species) to phenotype (growth pattern, which is optimized to the current environment). In fact this type of model has been previously used in a two-species interaction to examine changes in a species' adaptive landscape when it is in a competitive or cooperative relationship with a second species, and found greater robustness under cooperative conditions [108].

Here we take a similar basic approach but consider a more complex community of ten species in order to ask how the fitness landscape of a focal species varies across many different community contexts. Using a set of bacterial species commonly found in the human gut

microbiome, with *E. coli* as the focal species, and a previously developed dynamic FBA framework [55] we simulated the growth of all *E. coli* deletion mutants in co-culture with all possible subsets of the other species. With these results we sought to explore the patterns of landscape change and the mechanisms by which the community context causes this change. Specifically, we investigated which reactions show community dependent fitness and what those reactions have in common. We tested whether more similar community contexts cause more similar fitness landscapes and identified which community member species had the largest influence on the fitness landscape. We explored the structure of the fitness landscape's dependence on the community and tested whether mutations alter the nature of *E. coli*'s interactions with other species or just the dynamics of how the interactions play out. Along the way we used our findings to address questions of interest in the coevolution field, such as whether coevolution primarily acts through pairwise or diffuse mechanisms, and whether diversity in the community stifles potential adaptation or creates new opportunities [109–111].

3.3 RESULTS

3.3.1 *Quantifying mutant fitness across community contexts*

We investigated community-dependent fitness on a broad scale by utilizing a DFBA model of microbial co-culture growth. In DFBA the growth rate of each species is individually optimized based on both the metabolic reactions encoded in their genome (compiled in a genome scale metabolic model) and the metabolites available to them in the media. At each time point this optimization finds a new growth solution for each species. The rates of metabolite uptake and excretion in these solutions are then used to update the abundances of metabolites in the media and the growth rates from these solutions are used to update the species abundances. Interspecies interactions are not modeled explicitly but instead arise spontaneously as the past

growth of other species alters the available metabolites. Competition arises when multiple species consume the same metabolite until it becomes limiting, and cooperation arises when species excrete waste metabolites into the media that are then consumed by other species. For this study we used a previously developed multi-species implementation of DFBA [55].

Our community of study was modeled on a simple gut microbiome community. It consisted of the ten representative members of the human gut microbiome that had previously been used in an experimental mouse gut microbiome model [112]. We used high quality, semi-automatically constructed genome-scale metabolic models of the ten bacterial species (Table 3.1) [113] and further modeled the co-culture conditions and media after the mouse gut and diet in which this community was experimentally grown. The system was modeled as a chemostat and growth was simulated for 96 hours with two time steps per hour. The media was taken from one of the mouse chow diets used by Faith et al., resulting in a rich media containing several carbon sources and all amino acids. Out of the ten species in this community *E. coli* was chosen as a focal species due to its long history as a model organism.

To get at the community-dependence of adaptation, we investigated how the fitness consequence of each reaction deletion in the *E. coli* model would vary across different community contexts (Figure 3.1A). The set of all reaction deletions was used as the mutational landscape to explore. Deletions in the metabolic network have a clear effect on growth in this model and deletion of reactions directly rather than genes avoids dependencies and redundancies in the genetic network. Reaction deletions were modeled by limiting flux through the reaction to zero in the *E. coli* model, and the set of community contexts used was the full 511 possible combinations of between one and nine of the remaining species. We modeled the growth of all 1568 reaction deletion mutants in co-culture with every community context, with *E. coli* forming

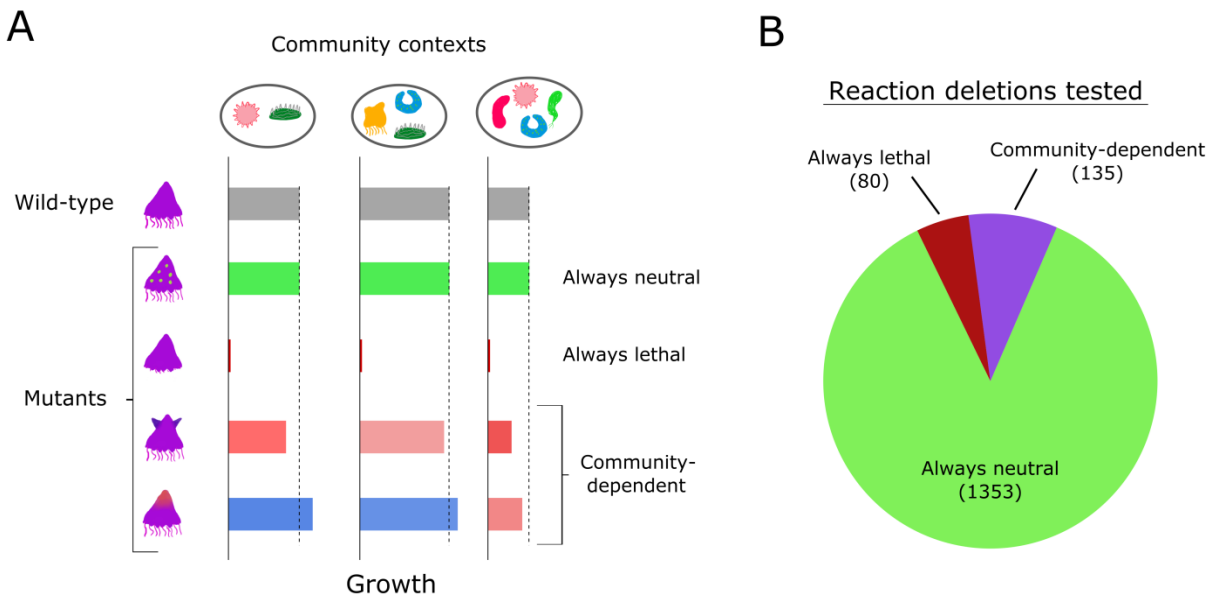

Figure 3.1. Method for measuring community-dependent fitness.
A) Each mutant is grown in co-culture with each community context. The biomass of each mutant at the end of a simulation is divided by the biomass of wild-type *E. coli* under the same conditions to get a fitness value. Mutants show community-dependent fitness if their fitness varies across different community contexts. B) The large majority of mutations tested had neutral fitness under all community contexts. A smaller set of 135 mutations showed varying fitness across community contexts, while the remaining 80 mutations were always lethal.

10% of the initial biomass and the rest of the community members forming equal portions of the remaining 90%. The mono-culture growth of each *E. coli* mutant was similarly modeled but as 100% of the same total initial biomass. In each case we generated a fitness value of a mutant in a community-context relative to wild-type in the same context by taking the final biomass of *E. coli* at the end of the simulation and dividing it by the final biomass of wild-type *E. coli* grown in the same context. Using this data set we proceeded to identify reactions with variable fitness across community contexts.

3.3.2 *Mutations with community-context dependent fitness*

As a first step toward understanding the influence of community-context on the adaptive landscape of *E. coli* we identified which of the 1568 mutants showed evidence of varying fitness

across different contexts. We considered each mutation separately and quantified the variance of its fitness effect across all communities. The large majority of mutations had zero variance, with 80 being always lethal and 1353 being always neutral (**Figure 3.1**B). The remaining 135 mutations showed variation in fitness across community contexts (**Figure 3.2**). These showed a wide range of patterns, from mutations that were nearly always neutral except for small deviations from neutral in a handful of contexts to mutations that had severe fitness defects in every context but to varying degrees. Some mutations resulted in better-than wild-type fitness, with 62 having positive fitness in at least one context. The highest fitness in any context was

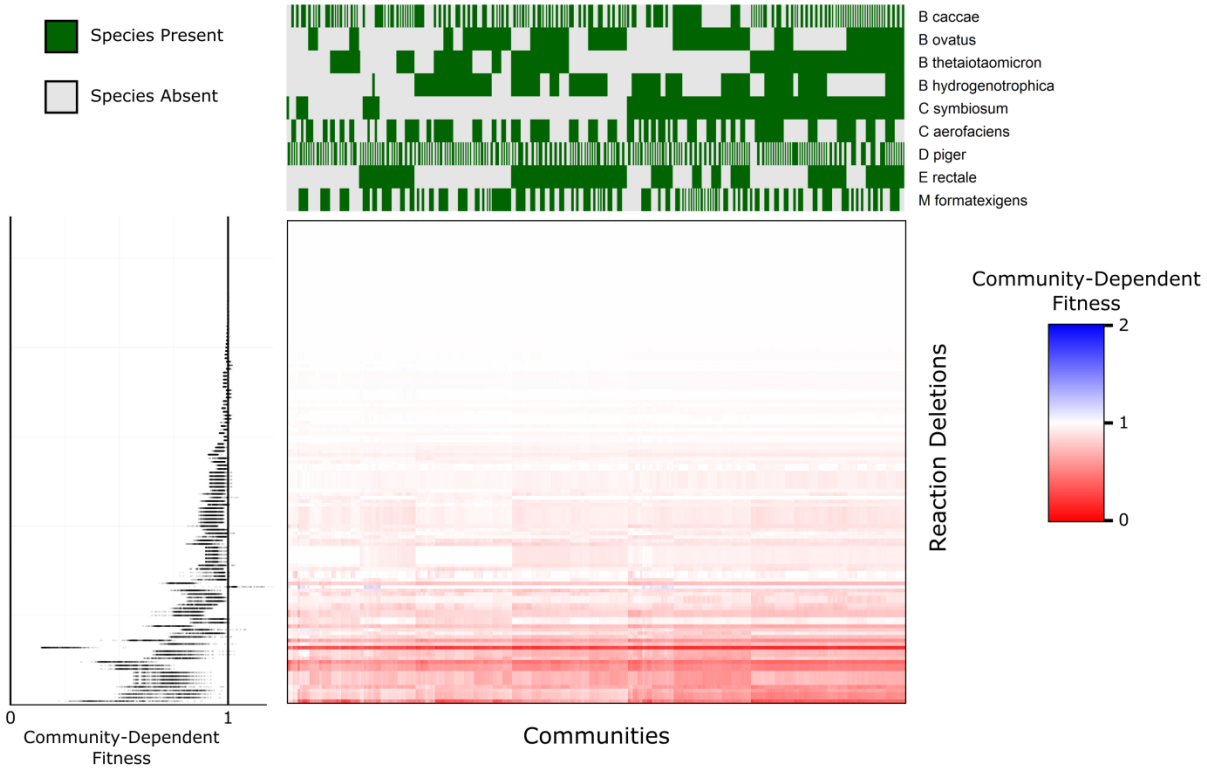

Figure 3.2. Community dependent reactions.
The central heatmap shows the fitness of all community-dependent reaction deletion mutants in each community context. Reaction deletions are sorted by mono-culture fitness and communities are hierarchically clustered. Left: the fitness of each reaction across all contexts is plotted to visualize the range of variation across communities. Top: The presence and absence of each community member in each community is shown. Any clustering of species is a consequence of clustering communities by *E. coli*'s fitness landscape.

1.29, achieved by the formate transport deletion mutant when co-cultured with *Clostridium symbiosum* and *Bacteroides caccae*. All of the mutations in this study are loss of function and the model doesn't include any benefit to having a smaller genome so each cases of greater than wild-type fitness must be caused by an indirect mechanism. For example, it is possible that a mutation alters the way *E. coli* interacts with one or more other species such that those species grow differently in a way that then benefits the *E. coli* mutant.

Of the variable fitness mutations 45 had neutral fitness in mono-culture but non-neutral fitness in one or more co-culture conditions, while the remaining 90 had reduced mono-culture fitness. Every mutant with a non-lethal growth deficit in mono-culture had varying fitness across contexts, and among these mutants the magnitude of growth defect in mono-culture was significantly correlated with the variation in fitness across community contexts (r = -0.48, $P < 10^{-50}$, Pearson). This raises the possibility that, for at least some of these mutants, the observed community-dependent fitness is not a consequence of altered interspecies interactions but rather a consequence of slow growers being "punished" to a large or smaller degree depending on the context. We investigate this further in the section "Mechanisms of how community affects fitness". However this alternative hypothesis cannot explain the 45 mutants with neutral mono-culture fitness or the instances of greater than wild-type fitness, meaning that altered interspecies interactions must underlie much of the community-dependent fitness observed.

An illustrative example of a mutation causing community-dependent fitness is the loss of alcohol dehydrogenase for ethanol. While wild-type *E. coli* generally can't consume ethanol as a carbon source, some strains over-expressing alcohol dehydrogenase have been found to be capable of doing so [114], and in our model *E. coli* can consume it. We found that the alcohol dehydrogenase (ethanol) deficient mutant had neutral fitness in mono-culture as well as many

(238/511) other community contexts, but altered fitness in the remaining contexts (Figure 3.3A). In a clear pattern, every community context where it had altered fitness had a non-zero amount of ethanol in the media by the end of the simulation, while every context where it had wild-type fitness had zero ethanol in the media. Ethanol was not present in the growth media so its presence must be caused by other species excreting it. Every community containing *Collinsella aerofaciens* had ethanol in the media, suggesting it constitutively produces ethanol, while the remaining 17 communities with ethanol but no *C. aerofaciens* all contained *Desuflovibrio piger*, suggesting that it produces ethanol conditionally (depending on the other species present). This shows a clear example of community-dependent fitness, where losing the ability to utilize a

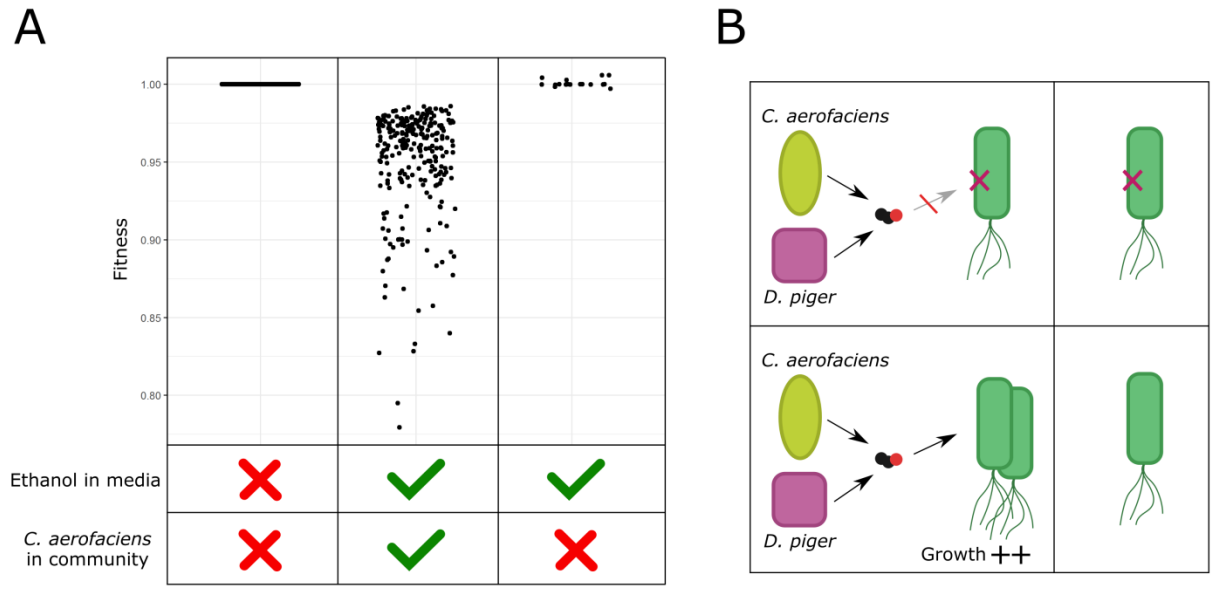

Figure 3.3. Example of community-dependent fitness in an alcohol dehydrogenase deletion mutant. A) The alcohol dehydrogenase deletion mutant has wild-type fitness in all communities where ethanol hasn't been excreted by any species. It has non-zero fitness when ethanol is present, which is usually in communities containing *C. aerofaciens*. B) Proposed mechanism for this behavior. Loss of this reaction doesn't cause a phenotype when no community members are excreting ethanol because it is not in use. However when ethanol is available there is a selective disadvantage to losing alcohol dehydrogenase because *E. coli* can't consume the ethanol.

resource is neutral when that resource isn't available, but is harmful when another species is present who can produce that resource (Figure 3.3B). Note that the ethanol transport deletion mutant also had identical fitness across all communities (several other groups of functionally related reactions also showed identical fitness to each other when deleted).

Looking more broadly at the functions of these reactions resulting in variable fitness revealed enrichment of several categories. The mutation that caused the highest overall variance in fitness between communities was loss of ammonia transport, and deletions of hydrogen sulfide transport, cysteine transport, and water transport were also in the top ten mutants with the highest variance. Overall transport reactions were significantly enriched among reactions with community-dependent fitness (47 transport reactions; Permutation test; $P < 10^{-5}$). A further 17 community-dependent reactions showed identical fitness to a transport reaction deletion, indicating that these are reactions needed for consumption or production of exported or imported metabolites. The enrichment for transport reactions is unsurprising given that in this model all interactions between species must occur through exchange of metabolites with the external environment. Looking at pathway annotations we found significant enrichment for nine pathways among mutations with community-dependent fitness (permutation test, 1% FDR). These were carbon fixation in photosynthetic organisms, butanoate metabolism, carbon fixation pathways in prokaryotes, TCA cycle, methane metabolism, glycolysis / gluconeogenesis, pyruvate metabolism, alanine, aspartate and glutamate metabolism, and arginine and proline metabolism. The number of pathways related to central carbon metabolism and amino acid metabolism suggest that these elements of metabolism play a major role in the interactions between *E. coli* and the other community members. Competition for these metabolites or cross-feeding of them may underlie some of the community-dependent fitness effects. In total, we

identified mutations with a wide range of functions that cause widely ranging patterns of community-dependence landscapes, with evidence of both direct and indirect causes through interspecies interactions.

3.3.3 *Adaptive landscape dependence on community context*

The breadth of mutations measured in this study enables going beyond analyzing them individually and considering the full adaptive landscape in each community context. This adaptive landscape is the set of fitness consequences of all mutations in a particular community context. One expectation of such fitness landscapes is that, if community member identity is important in determining community-dependent fitness, then communities with more similar species compositions should result in more similar fitness landscapes for *E. coli*. This hypothesis breaks down further into two expectations, 1) communities that have more species in common will have more similar fitness landscapes, and 2) communities with more species differing between them will have less similar fitness landscapes. We tested this hypothesis by comparing the Euclidean distance between fitness landscapes for every pair of communities, and then binning pairs by the number of species present in both communities or exclusive to one community (Figure 3.4). The results matched expectations, with a higher number of community member species in common resulting in higher average adaptive landscape similarity and higher number of differing species resulting in lower average adaptive landscape similarity, almost independent of each other and across the ranges of species shared and exclusive. The phenomena of more similar communities causing more similar fitness landscapes may act through two separate mechanisms: interactions between *E. coli* mutants and the species in common may buffer the effects of interactions with other species, but also more species in common will decrease the initial abundance of exclusive species.

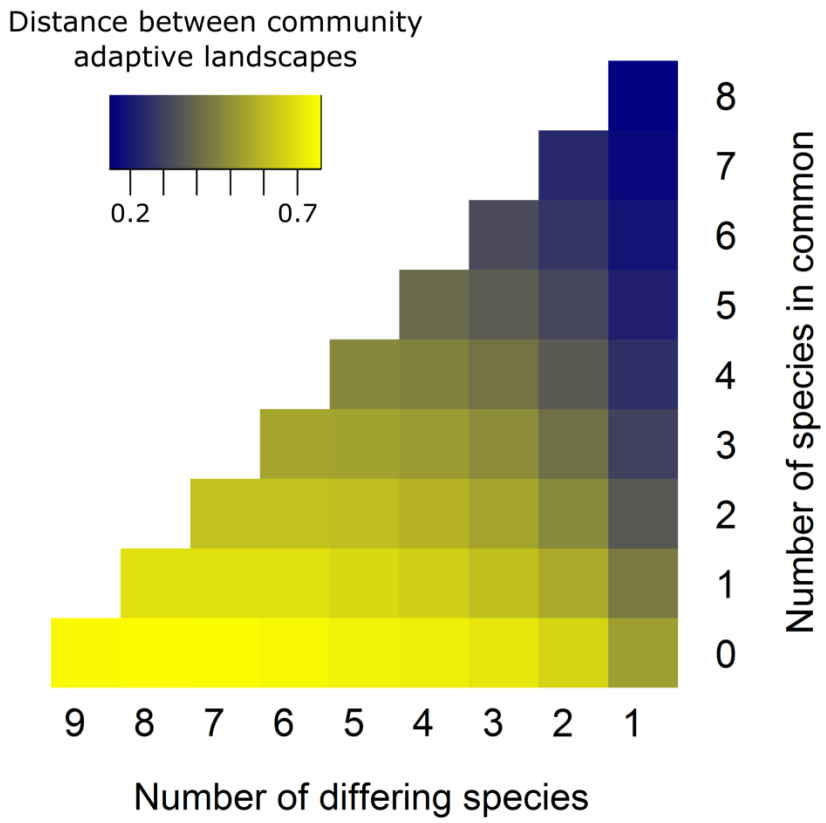

Figure 3.4. Effect of community composition similarity on the adaptive landscape similarity. The distance between *E. coli*'s fitness landscape in different contexts was measured for every pair of community contexts and binned by the number of species in common and differing between the pair. Square colors show the mean euclidean distance for all pairs of communities that have the given number of community member species in common and in different, with blue being a low difference in adaptive landscapes and yellow a high difference.

Given that community species composition is important, we next asked if community member species differ in the magnitude of their effect on the adaptive landscape of *E. coli*. To answer this we measured the Euclidean distance between adaptive landscapes for every pair of communities that differ by a single species, and grouped those comparisons by which species was added/removed (Figure 3.5). Species did indeed vary greatly in their relative importance, with *C. symbiosum*, *B. thetaiotaomicron*, *E. rectale*, and *B. ovatus* having the largest effect (in that order). *D. piger* had the smallest effect which matched its consistently low abundance. In fact, the average abundance of each species was correlated with its average effect size (Pearson;

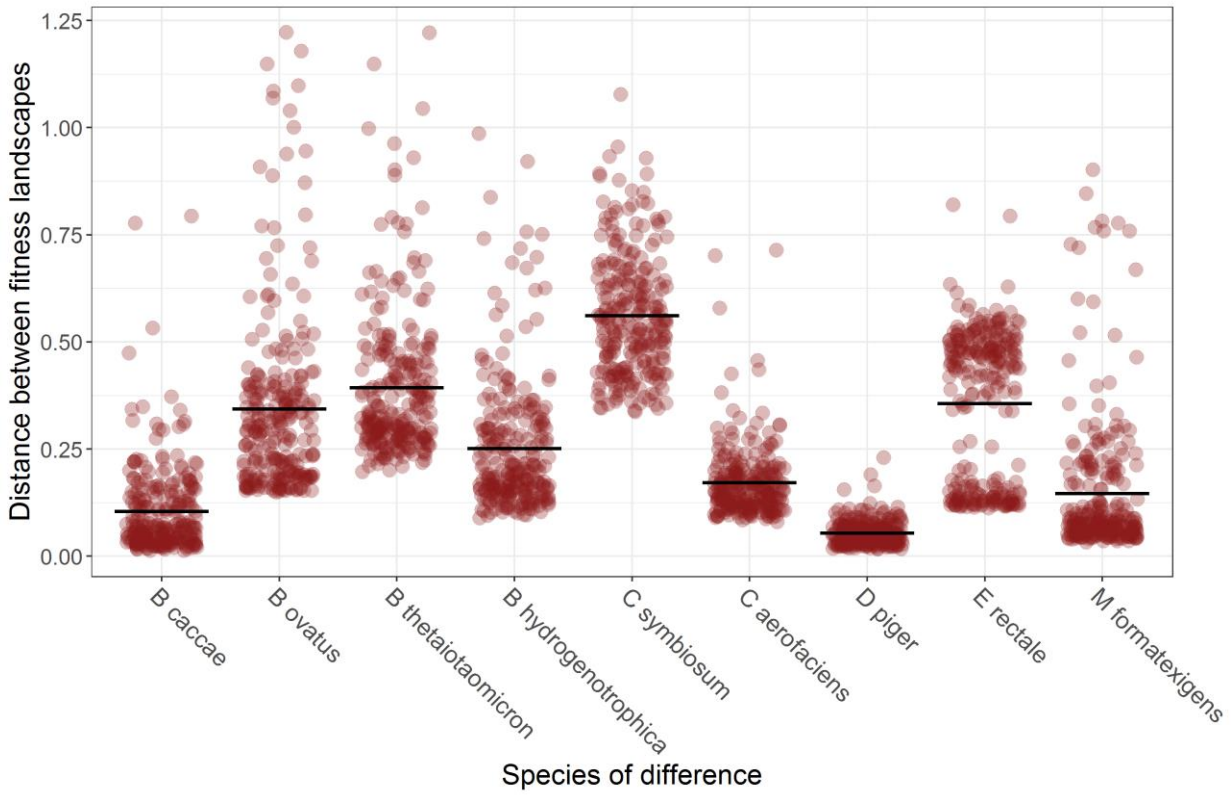

Figure 3.5. Difference in fitness landscape caused by adding/removing each species individually. Each dot represents the Euclidean distance between a pair of communities whose species are the same except for differing in the presence of the given species. The black lines show the mean distance of all community comparisons for each species.

P < 0.02), but this correlation did not fully explain variance in importance of different community members, indicating that differing types and magnitudes of competition and cooperation between *E. coli* and different community members also play a role. All species showed variance in the magnitude of their effect on the fitness landscape where in some community contexts their addition/removal had a larger effect. Although *B. thetaiotaomicron* and *B. ovatus* caused a lower mean distance than *C. symbiosum* there were some outlier contexts in which they caused a much larger distance than any communities with *C. symbiosum*. *E. rectale* showed an interesting bimodal behavior in which it affects *E. coli*'s fitness landscape differently in different contexts. This bimodality is caused by the presence or absence of *C. symbiosum*, where the impact on the landscape from adding *E. rectale* is significantly lower

when *C. symbiosum* is present than when it is absent (t-test; mean 0.23 vs. 0.49; P < 10-10). This kind of interaction between other community members, as well as the general variance of effect size across communities, paints a complicated picture of the fitness landscape's dependence on community composition.

To explore the structure of how the fitness landscape of *E. coli* differs between community contexts we utilized dimensionality reduction. We applied principal component analysis (PCA) to the community-dependent fitness, treating communities as observations and mutations as variables (Figure 3.6). The first two principle components explained 69.5% of the variance. Interestingly, the first principal component was significantly correlated with the number of species present in the community (Pearson's correlation = 0.6; P < 10-10), suggesting that it may be capturing the degree of competition which E. coli faces in each community and the variation in the fitness landscape caused by that competition. The second principle component was not significantly correlated with community size (P = 0.06). Marking communities by the presence or absence of the three species found to have the largest effect on the fitness landscape showed clustering of communities with similar species compositions. Communities with *C. symbiosum* and *B. thetaiotaomicron* or all three species clustered closely together, showing a relatively stable fitness landscape between these communities. In contrast communities with just one of the three species or the other two pairs are more dispersed, possibly suggesting that the fitness landscape is more stable in large communities with high competition.

We next investigated how the potential for adaptation changes with the complexity of the community. The number of mutations with positive fitness in each community context were significantly positively correlated with the number of species in the community (**Figure 3.7** Top; Pearson; P < 10-8). Interestingly, the number of mutations that are accessible at a particular

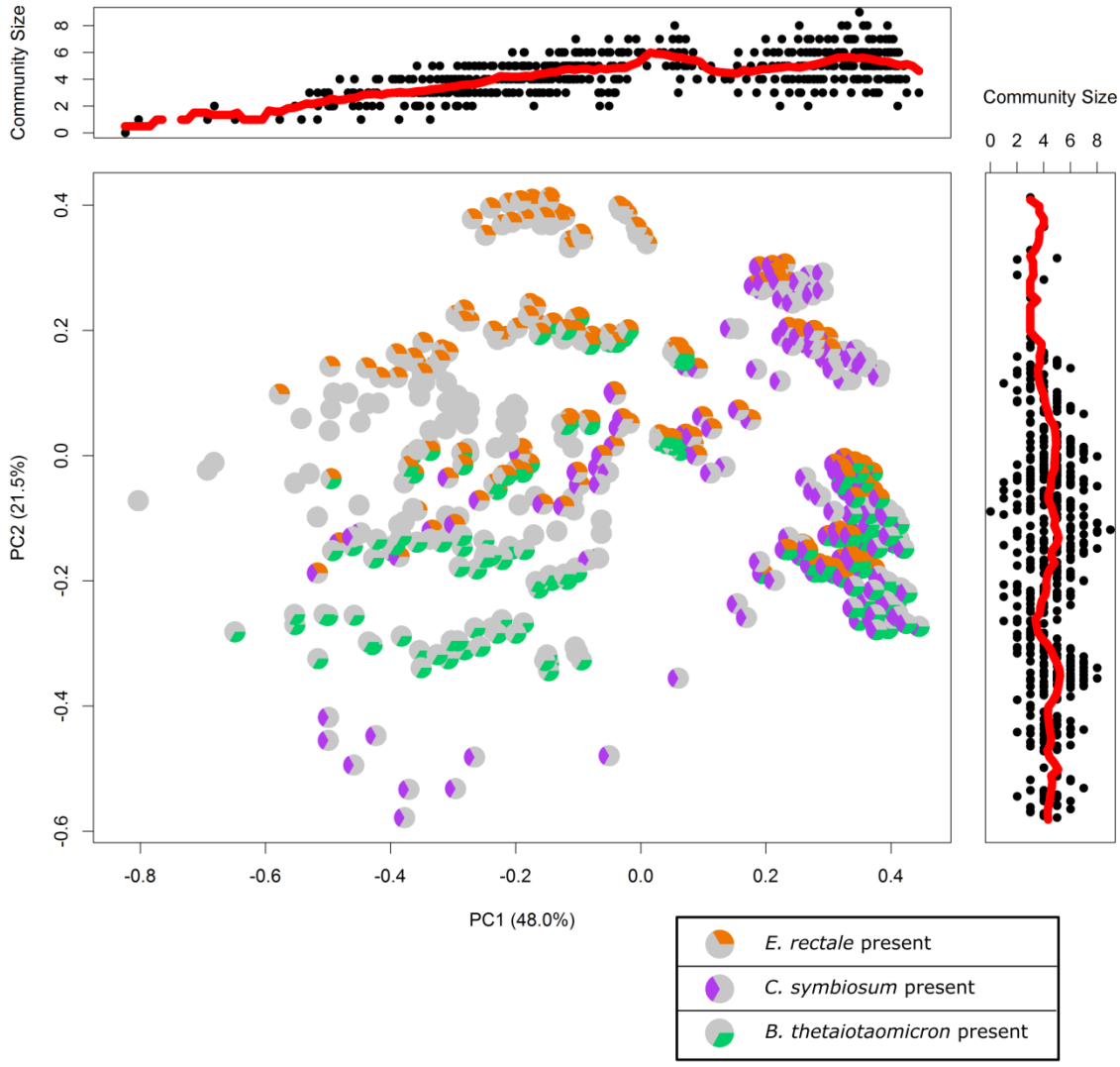

Figure 3.6. Structure of fitness landscape variation.
Each circle is a community and its position represents the location of that community on principal coordinates calculated using the fitness landscape of each community. Colored wedges indicate the presence of the three species found to have the largest impact on the fitness landscape. To the top and right are plots showing the correlation or lack of correlation between the principal components and the number of species in the community context.

fitness penalty shows an inverse trend. For example, the number of mutations accessible with a 1% fitness penalty or less is significantly negatively correlated with the number of community members (Figure 3.7 Bottom; Pearson; P < 10-15). This pattern held and was significant for all fitness penalties tested from 0.01% to 50%. This suggests that higher diversity communities have

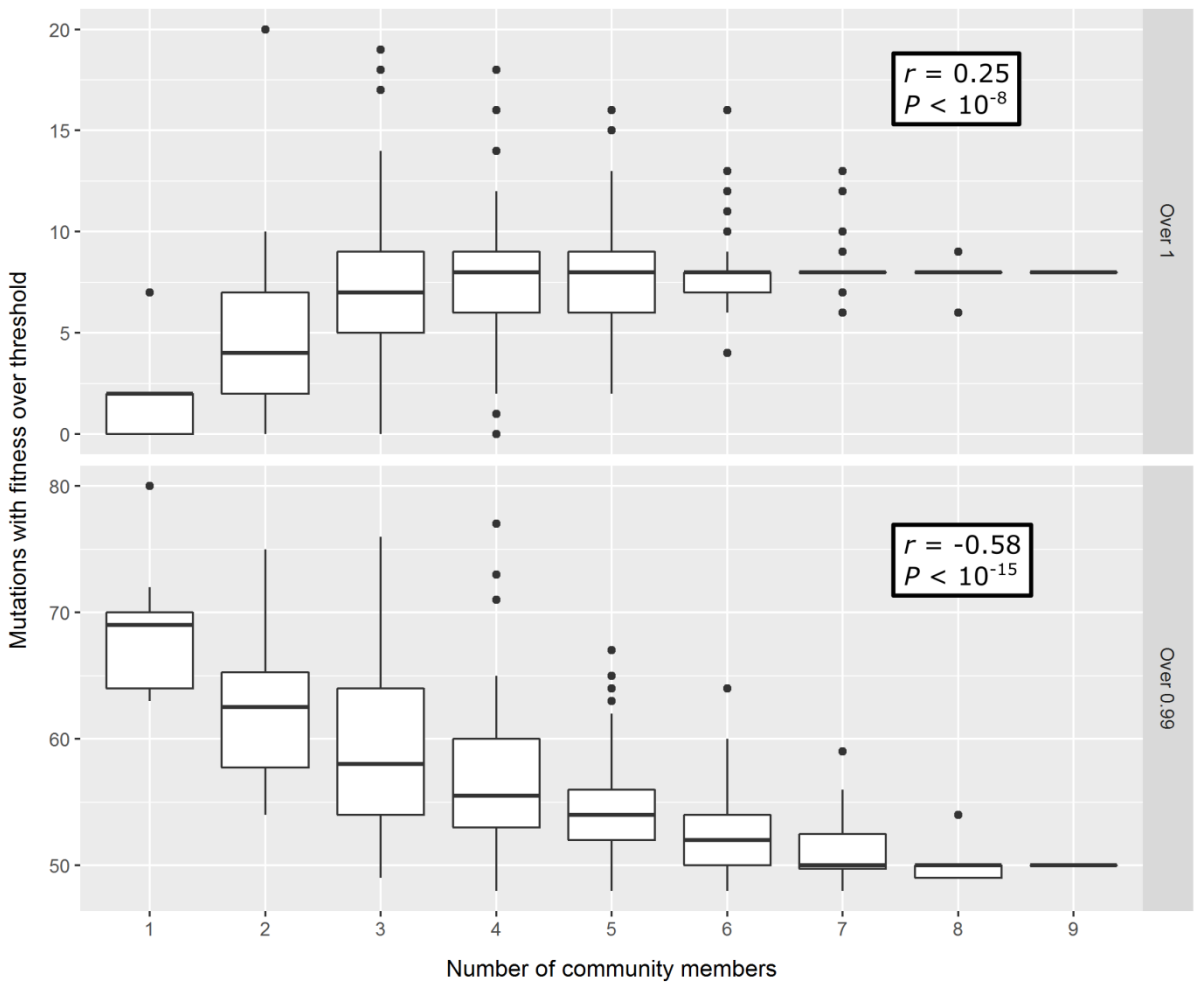

Figure 3.7. Effect of increasing community diversity on the number of mutations accessible. Box plots show the number of mutations with a fitness value over a threshold within each community of a given number of members. Top: Mutations with a fitness value over 1. Bottom: Mutations with a fitness value over 0.99. The Pearson correlation coefficient and P-value for correlation are shown for each threshold.

increased competition and fewer available niches which punish harmful mutations, but the higher potential for interactions between species create opportunities for indirect beneficial effects.

3.3.4 *Mechanisms of how community affects fitness*

To gain a fuller understanding of how the community context affects the fitness of *E. coli* mutants we sought to explore the mechanisms underlying the community-dependent fitness of

individual mutants. As described previously, the correlation between mono-culture growth defects and variance of community-dependence suggested a mechanistic null hypothesis whereby mutations do not alter the way *E. coli* interacts with other species but merely slow down its growth while it consumes and excretes metabolites the same way. Variation in fitness between communities would then arise through simple ecological dynamics as slower growing *E. coli* mutants are out-competed by other species to a more severe degree than wild-type *E. coli*, without altering the nature of the competition. The alternative mechanistic hypothesis is that community-dependent fitness results from perturbed dynamics when *E. coli* alters its consumption and/or excretion of metabolites in a way that affects its interactions with other species. To test these hypotheses we created modified versions of the *E. coli* model by multiplying all inputs and outputs to the biomass reaction by a modifier, thus slowing growth by increasing the resource demand of growth. We generated modified models with mono-culture growth matched to each reaction deletion mutant and ran co-culture simulations across all community contexts using these models (with the unaltered models for the other nine species) (see Methods). By comparing the variance in fitness across communities of each community-dependent reaction with its control we found that many mutants showed variance equal to or less than the control, but several others showed greatly elevated variance (Figure 3.8A). Those with elevated variance can be assumed to have different underlying mechanisms involving changed interspecies interactions, but the mutants with variance similar to that expected from controls could either act through simple slowed growth as the controls or altered interspecies interactions. To further differentiate between the two possibilities we measured the spearman correlation between the fitness landscapes of each mutant with a mono-culture growth defect and its matched control correlation (Figure 3.8B). Approximately half of these mutants were well

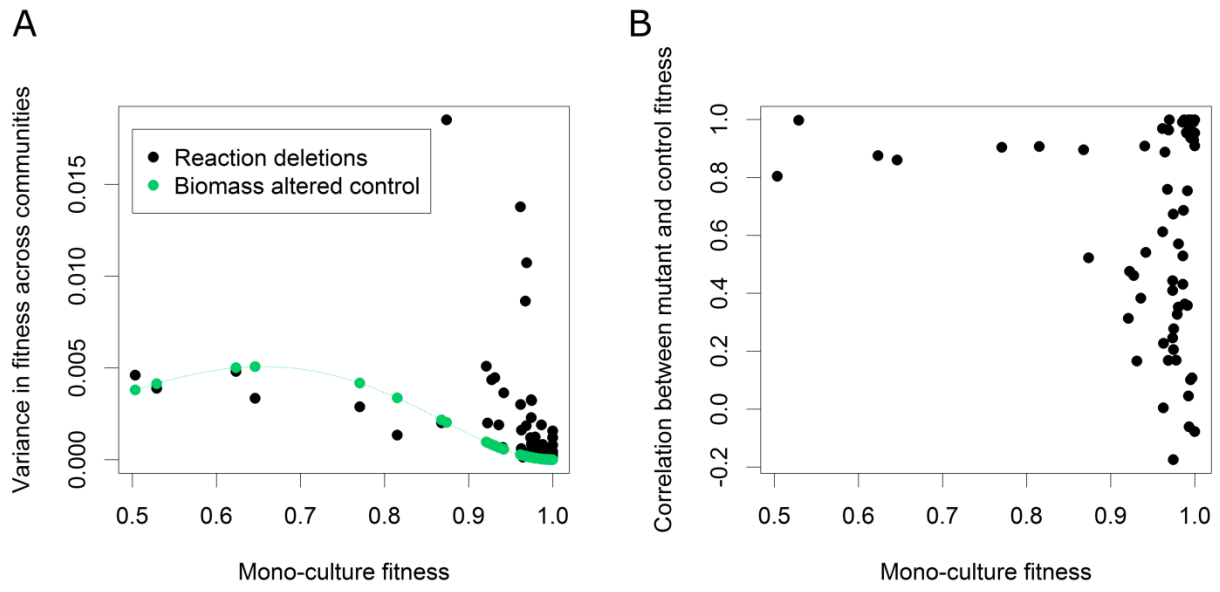

Figure 3.8. Comparison to biomass modified controls.
A) Variance of fitness across communities as measured from reaction deletion mutants (black) compared to the same measure in mono-culture growth matched controls (green). The dotted line interpolates the variance of the controls. B) Spearman correlation between the fitness across communities of the reaction deletion mutants and their matched controls.

correlated with the controls with correlation over 0.8, while the remaining half showed a large range in correlation down to below 0. The fact that nearly all mutants fell somewhere in-between 1 and 0 suggests that the mechanism behind most community-dependent fitness is a combination of dynamics due to slowed growth and altered interactions, with different ratios of contribution for each mutant.

To further identify the mechanisms of community-dependent fitness we turned to regression analysis. We sought to regress the fitness of a mutant to the species present in the community contexts to identify which species harm or benefit each mutant. Regression additionally provides the opportunity to test whether community members alter fitness independently or whether interactions between them are important. To accomplish this, we applied three different methods, basic linear regression, LASSO regression, and Logic

Regression, to the top 40 most variable reactions. Basic linear modeling does not include any interaction terms, while in our LASSO regression we included second and third order AND and OR interaction terms, but limited the total parameters to nine to be consistent with the other methods. Logic regression fits data to Boolean expressions of binary variables (in this case the presence/absence of the nine community members) using simulated annealing [115]. These Boolean expressions can potentially capture even more complicated relationships between community members. For example, if mutant Y grows much worse when either species C or D is present, the algorithm will likely create a term for (C OR D) with a negative coefficient. We performed 16-fold cross-validation using all three methods on the top 40 most variable reactions and found that on average logic regression performed significantly better than the other two (ANOVA, $P < 10\text{-}10$; Figure 3.9). Since logic regression can account for more complex interactions between variables, interactions between community members must play an important role in determining mutant fitness. This suggests that the coevolutionary process in this community is not merely the sum of pairwise interactions, but is more akin to diffuse coevolution [93]. The mechanism of action of the diffuse coevolution is likely the metabolites in the environment which may be consumed or secreted by several species.

3.4 DISCUSSION

In this study we utilized a computational model of a bacterial community to broadly study the consequences of community context on the fitness landscape of a focal species. Our computational model allowed testing the fitness of over a thousand mutations in hundreds of community contexts, producing a scale of data not previously available in similar studies of coevolution. We identified community dependent fitness in over one hundred mutations and found many functions involved, including an enrichment for transport reactions. Differences in

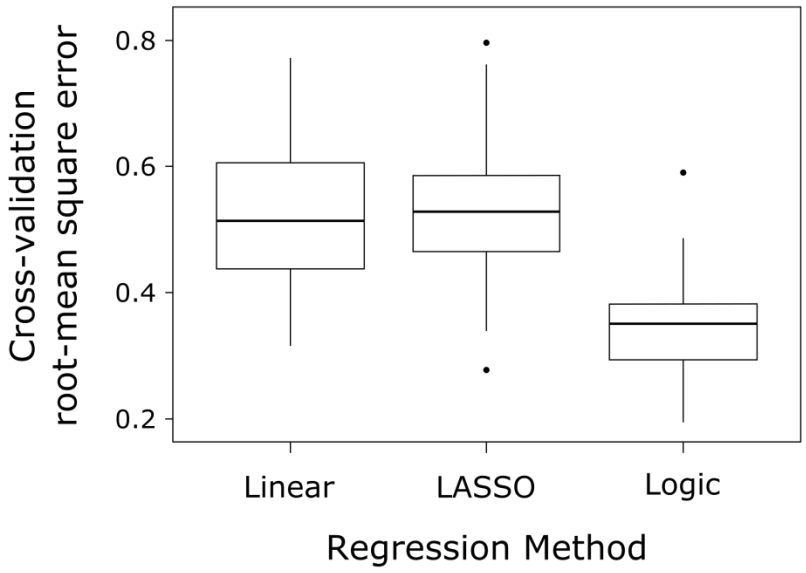

Figure 3.9. Cross-validation accuracy of fitness regression to community context.
Regressions were fit to the fitness of the 40 reactions with the most variable fitness (taking only reactions with unique profiles) using three different methods and 16-fold cross-validation. Plotted are the root-mean square error of each of the forty reactions for the three different methods. Logic Regression produced significantly lower error.

fitness landscapes between communities were closely related to differences in community members, although species varied in their relative importance to altering the fitness landscape. We rejected our null hypothesis that community-dependent fitness was arising as a mere consequence of slowed growth in all mutations and found varying relationships to community composition among these mutations. Overall, our findings suggest complex but prevalent relationships between community context and the fitness effect of mutations.

An open question of study is whether diversity in the community increases or decreases the potential for an individual species to adapt. Some previous studies have found that higher diversity cuts off opportunities for adaptation because niches are filled [96,116], while others have found that diversity in the community increases opportunities for a species to adapt to utilize underexploited resources [109,110]. In this study we found evidence to support the later

view, in that as the number of community members increased the number of beneficial mutations increased (Figure 3.7). This effect plateaued rather than increasing indefinitely, suggesting that having some diversity of community members enabled new interspecies interactions to form and thus new dynamics that could be exploited by particular mutants, but these dynamics stayed the same as even more species were added. Alternatively the plateau effect may simply be a result of the limited number of community members used in this study.

One drawback to our approach is that we are only considering loss of function mutations. Other studies have, for good reason, primarily focused on how adaptive mutations can increase the fitness of a species depending on the biological environment. How the extent of negative fitness consequences of a mutation vary across community contexts is less relevant to the course of evolution since it will be selectively disadvantageous in all contexts. Gain of function mutations could be studied using a similar model to the one used here by modeling mutations as adding metabolic reactions to a model. In order for the model to be able to utilize these new reactions it may be necessary to add entire pathways at a time, analogous to horizontal gene transfer of entire operons [117]. This approach may work better starting from a lower abundance species so that there is more potential for its growth to improve.

Another drawback of our model is the constant inflow of nutrients which is starkly different from the environments experienced by wild bacteria. A constant source of fresh media allows interspecies interactions to form and be maintained over time, thus resulting in larger fitness consequences depending on which species are present [118]. In a real gut the nutrient inflow is not constant, varying as the host eats and sleeps, and the nature of the inflowing nutrients can vary depending on the diet [119]. All of these variations in the abiotic environment

could act to weaken the effect of interspecies interactions on mutant fitness, so by excluding them in our model we may have strengthened the observed effect of community-dependence.

There are several potential future directions from these results. In this study we limited the time scale of evolutionary change to a single mutation, which is a major simplification of the potential for change. Future studies could use a similar framework to study how multiple mutations may together produce adaptations to novel ecological environments. It would also be possible to experimentally test a subset of our predicted community-dependent fitness results. Re-running our framework using other focal species in the same community could test the generality of these findings. For this study we focused on how the evolution of individual species can be influenced by the ecology around them, but future studies could expand that scope to consider how that evolution in turn feeds back to alter ecology and the evolution of other species.

These findings expand our understanding of the evolutionary potential of individual microbial species in complex communities. The dynamics of the fitness landscape across community contexts can generalize to the study of community-dependent fitness in other systems where such high-throughput interrogation of the fitness landscape is not possible. The results will additionally be of relevance to continued study of evolutionary and ecological dynamics in microbial communities and their applications to medical, environmental, and industrial processes.

3.5 METHODS

3.5.1 *Co-culture Growth Model*

Community growth was modeled using a previously developed dynamic FBA framework [55]. This model assumes a well-mixed culture and that each species grows selfishly, optimizing its own growth rate. Briefly, the cell and media concentrations are initialized and then three steps

iteratively repeat. First, the amount of each metabolite available to each cell is calculated and used to set the bounds for transport reaction fluxes. Second, given these and other constraints in the model, the optimal growth flux solution is calculated for each species. Third, using the growth rate and transport reaction fluxes calculated, the concentrations of cells and metabolites in the culture are updated. A more detailed description of this method can be found in Chiu et al.

Co-culture and media conditions were designed to simulate a mouse gut. The co-culture was modeled as a chemostat with volume (1.34 mL) and dilution rate (0.0472 per hour) similar to a mouse gut. The species used were 10 bacterial species that had been previously used as a model gut microbiome community [112], and the media used was based on media F from that same study, which contains a ratio of roughly 40% carbohydrates, 40% protein, and 20% fat. The contents of that media were translated into abundances of individual metabolites that could be input to the co-culture model. The genome scale models used for the 10 species were acquired from the AGORA collection [113](Table 3.1). Each co-culture simulation was initialized with the same total biomass of 0.5 grams dry cell weight / L, of which 10% was formed by *E. coli* and the remaining 90% was divided equally between the other species present. Co-culture growth was simulated for 96 hours with time steps of 0.5 hours.

3.5.2 *Community Dependent Fitness Framework*

E. coli mutants were made by setting the upper and lower bounds of the flux of the reaction to be deleted to 0. Exchange reactions, source/sink reactions, and the biomass reaction were excluded from the study. Community contexts were constructed by taking all 512 possible subsets (from 0 to 9 species) of the 9 non-*E. coli* species. Co-culture growth of all combinations of mutant and community context was modeled as described above. The biomass of *E. coli* at the

last time point (96 hours) was used as a proxy for its fitness. The community dependent fitness of each mutant in each community was calculated by dividing the biomass of that *E. coli* mutant in that community by the biomass of wild-type *E. coli* when grown in the same community. This resulting fitness value was found to contain a very small amount of noise for many mutations, where mono-culture growth occasionally showed very slightly negative or positive deviations from 1. It is not possible for a reaction deletion to cause an increase in growth rate in mono-culture, indicating that these deviations are indeed noise. To exclude this low level of noise from the analysis all community dependent fitness values were rounded to the nearest 10^{-7}.

3.5.3 *Growth Deficit Controls*

Control data was generated using mutations that slow growth of *E. coli* without altering resource consumption or excretion, thus capturing the ecological effect of decreased mutant growth in the absence of altered interspecies interactions. To accomplish this the stoichiometry of the biomass generation reaction in the *E. coli* model was modified by multiplying the coefficient of each metabolite other than biomass by a modifier value. The inverse of this modifier roughly corresponds to the mono-culture fitness and so modifiers ranging from 1 to 2 were used, covering the range of mono-culture fitness found in the non-lethal reactions from the main data set. Simulations were run using 32 modifiers across this range and otherwise the same conditions as the regular simulations, and then a linear regression was fit to the inverse of these modifiers using the resulting mono-culture growth of each mutant. The fit of this linear regression was extremely tight, with an R^2 of 1 and a maximum absolute residual below $1x10^{-9}$. This linear fit was used to identify modifier values corresponding to the mono-culture fitness of each reaction deletion mutant in our dataset. Simulations were then run across all community

contexts using these modifier values to generate mono-culture growth equivalent control data for all reaction deletion mutants. The fitness of each control mutant in each community context was calculated using the same method as the regular data.

3.5.4 *Analysis of Variable Reactions*

Reactions with variable community dependent fitness were identified by calculating the variance of the community dependent fitness of each reaction across all communities. Transporter categories were manually curated by finding reactions that moved metabolites between the external and cytosolic compartments. Kyoto Encyclopedia of Genes and Genomes pathway annotations for each reaction were taken from the AGORA model paper [113,120]. Pathway and transporter enrichment was tested by counting the abundance of each pathway annotation among community-dependent reactions, then taking 1,000,000 random subsets of all the reactions, with the same size as the set of community-dependent reactions, and again counting the abundance of each pathway annotation. For each pathway the P-value for enrichment was the fraction of permuted subsets with the same or more counts of that pathway. A false discovery rate of 1% was used [121].

3.5.5 *Analysis of Mechanisms of Community Effect on Fitness*

Cross-validated regression analysis was performed to determine which regression method best described the data. Cross-validation was done 16-fold and the same training/test sets were used for all regression methods, and the data were standardized. Linear regression was performed normally. For LASSO regression a λ value was chosen to keep the number of non-zero parameters at ten, to be comparable to basic linear regression. Logic Regression is a method for finding Boolean expressions of binary variables, implemented in the R package LogicReg

[115]. For logic regression simulated annealing was run with 100,000 iterations across a temperature range of 0 to -3.

Species	Strain
Eubacterium rectale	ATCC 33656
Collinsella aerofaciens	ATCC 25986
Desulfovibrio piger	ATCC 29098
Blautia hydrogenotrophica	DSM 10507
Clostridium symbiosum	ATCC 14940
Marvinbryantia formatexigens	I-52, DSM 14469
Escherichia coli	K-12 substr. MG1655
Bacteroides ovatus	ATCC 8483
Bacteroides thetaiotaomicron	VPI 5482
Bacteroides caccae	ATCC 43185

Table 3.1. Species used in the study.
All models are from the AGORA v1.01 collection. In that collection each model was constructed using the genome of a particular strain, which are listed in the right column. *E. coli* was used as the focal species in this study.

Chapter 4. CONCLUDING REMARKS AND FUTURE CONSIDERATIONS

4.1 GENOME-SCALE METABOLIC MODELING INFORMS COEVOLUTIONARY DYNAMICS

In these two works I have described novel research into the mechanisms and patterns of coevolution between bacteria. I have utilized genome-scale metabolic modeling to study how the actual genome content and metabolic networks of bacteria influence the evolutionary consequences of metabolic interaction in co-culture growth. The simulation-based nature of these works has allowed them to cover wide scales, with sufficient time resolution for co-culture dynamics to resolve and large numbers of mutations and varying ecological contexts considered. This combination of detail and scale allows identification of general patterns of how individual mutations affect ecological interactions and vice versa.

In chapter 2 I applied this model to the evolutionary regime of endosymbionts tightly co-occurring while experiencing reductive evolution. By simulating many replicate evolutionary trajectories I identified frequent emergence of metabolic dependencies even though cooperation was not explicitly defined or selected for. The numerous and tangled associations between metabolic phenotypes of providing and depending on different metabolites and loss or retention of genes revealed the importance of the interconnectedness of the metabolic network. Historical contingency resulted in opportunities for cooperation being missed when necessary genes had already been lost.

In chapter 3 I applied this model to the dependence of a focal species' fitness landscape on its ecological context. Community-dependent fitness effects were prevalent, with many different mutations affected. Community member species varied in how much their presence

impacted *E. coli's* fitness landscape, and more complex communities provided more opportunities for *E. coli* to gain beneficial mutations. Interactions between community members were important for explaining the fitness of individual mutations, suggesting that diffuse coevolution is a dominant process.

4.2 FACTORS THAT COMPLICATE THE COEVOLUTIONARY PROCESS

In the preceding chapters I have examined coevolutionary processes between bacteria under two very different regimes. However, many commonalities arose between the patterns seen in the two results. One major commonality is the complexity of dynamics. In both cases the interaction of genomic and metabolic architecture with ecological effects and the plastic growth phenotype response make it difficult to disentangle causes and effects. In chapter 2 this complexity manifested in the difficulty in identifying gene deletions that were causal to metabolic phenotypes. Few gene deletions or retentions were perfectly correlated with a metabolic phenotype of obligate receiving or providing, while most gene events were partially correlated to multiple different metabolic phenotypes. Each phenotype in turn had many different genes correlated to it. These findings and the frequency of missed opportunities for cross-feeding suggest a strong dependence on genomic and ecological context for coevolutionary events within this model. The nontrivial mapping of genes to metabolic reactions and the size of the metabolic network contribute to this complexity. Similarly, in chapter 3 I found that it was impossible to neatly deconvolute the effect of the presence or absence of each community member to the fitness consequences of each mutation. Instead the plastic phenotypes of both *E. coli* and the community members influence each other as growth progresses to determine the specific resource availability that ultimately determines the fitness of a mutant. In both of these cases my work has demonstrated the difficulty in predicting the outcomes of coevolution under different

contexts. These results suggest an importance for considering local context in any attempt to understand coevolutionary dynamics. The specific metabolism of involved species, other genes whose presence or absence limits or facilitates avenues to adaptation, and environmental metabolites will all majorly impact one species' evolutionary response to one or more other species. Additionally, the results of both of these studies demonstrate the power of mechanistic simulations.

4.3 Future Directions

These results provide exciting new insight into how genomic and metabolic mechanisms and networks can influence and complicate bacterial coevolution, but coevolution is a challengingly complex topic. The model used in this work makes many simplifications to the conditions bacteria experience, but future work could expand the model and study how these altered conditions or parameters affect the coevolutionary process. In particular spatial structure would be interesting to incorporate into the model as it has been repeatedly implicated as important or necessary for cooperative behavior to succeed in scenarios ranging from abstract games [122] to detailed mechanistic simulations of microbial growth [123] to experimental co-cultures of wild bacteria [124]. Another follow-up would be the use of a similar model with a larger range of species to identify which results are more generalizable and which depend on specific genomes. Lastly these two studies were limited in part of their scope (community complexity in one case and time in the other) to allow exhaustive examination in a different dimension, but natural coevolution will result in both diverse, complicated communities and cascading interactions lasting for generations [91]. Future modeling work could consider a larger scope to study how contingency of mutations and interactions develops in complex communities.

Despite the many advances made by these and prior theoretical and computational studies, a full understanding of evolution within bacterial communities will require more experimental work. Experimental studies of bacterial evolution in multi-species contexts are not common. More work in this area will be necessary to test the efficacy of mechanisms like the black queen hypothesis in natural contexts [32]. Experimental studies of community evolution present challenges in the length of evolution and numbers of replicates needed, but recent work in microbial ecology has demonstrated the potential of combining experimental and theoretical approaches [24].

One future horizon for studies into microbial coevolution will to consider genetic and phenotypic variability within populations. The studies described here as well as nearly all similar studies assume that each genotype only expresses a single phenotype, but stochasticity in cellular regulation, gene expression, and more result in a great deal of phenotypic heterogeneity in living cells. This heterogeneity is thought to provide an adaptive advantage [125]. Phenotypic heterogeneity, as well as standing genetic variation within populations, may further complicate coevolutionary dynamics. Future computational studies may explicitly consider variation through tools such as flux variability analysis [52], and new technological advances have resulted in ways for detailed phenotype measurements of single cells to be made in high throughput in experimental studies [126].

As our understanding of microbial ecology and evolution improves it will be increasingly possible to apply to practical applications such as synthetic microbial communities [127]. Synthetic communities have already been constructed that can successfully perform simple tasks like creating spatial structure, performing basic calculations, and generating oscillations [128–130]. As our ability to engineer synthetic capabilities in these communities improves it will

become possible to design and deploy them to complex practical applications like sensing environmental conditions or synthesizing complex molecules in a multi-step process. Obligatory cross-feeding relationships could be engineered between members of these communities to stabilize their composition, and the findings in Chapter 2 suggest metabolites and mechanisms that may be useful. Another major challenge to a designed microbial system is the potential for evolutionary events to arise that will disrupt the intended function. If such evolutionary escape routes can be anticipated, it will be possible to engineer interactions within the community that result in selective forces which oppose the spread of undesirable mutations.

Synthetic communities and other practical applications will require many further advances in our understanding of bacterial coevolutionary processes. The complexity of interactions in diverse communities and their ecological and evolutionary consequences will be challenging to fully disentangle, but these results demonstrate the necessity of using genome-level information and detailed mechanistic models to achieve that goal. This dissertation supplements the growing study of ecological and evolutionary dynamics within microbial communities.

APPENDIX A: SUPPORTING INFORMATION FOR CHAPTER 2

Supporting Text

Effect of Fitness Cutoff on Cross-Feeding Prevalence

One parameter that may have a major effect on the emergence of species interaction is the fitness drop cutoff used when allowing a mutation to fix. Presumably, using a less stringent cutoff (i.e., allowing even more deleterious mutations to fix) may allow species to explore additional regions of the fitness landscape and potentially other configurations through which the two species may rely on each other. Conversely, a more stringent cutoff (allowing only very slightly deleterious mutation to fix) may prevent the evolutionary process from reaching configurations that allow the two species to interact. To test this hypothesis, we ran additional simulations, using a cutoff of either 10% or 1% (compared to the 5% used in the main text). Indeed, we found that using a more relaxed cutoff (10%) resulted in a higher frequency of mutualistic, commensal, and collapsed communities, and less independent communities (6.3%, 46.4%, 13.4%, 33.9%, respectively, of a total of 1000 simulation runs; $P < 0.007$ for collapse and $P < 10^{-8}$ for the rest; χ^2 test). Similarly, simulations with a more strict cutoff (1%) had significantly fewer evolutionary trajectories resulting in mutualistic, commensal, or collapsed communities, and more trajectories resulting in independent communities (0%, 27.5%, 3.1%, 69.4%, respectively, of a total of 2000 simulation runs; $P < 10^{-10}$ in all cases; χ^2 test).

Notably, the fitness cutoff used in our study was chosen as an intermediate value between the two cutoffs used by [75], and as shown above has a significant effect on the frequency of evolved commensalism and mutualism. Since this fitness cutoff represents the strength of selection, a permissive fitness cutoff (as the one used in our study) allows genetic drift to play a

dominant role in determining evolutionary trajectories, in agreement with the balance between selection and genetic drift hypothesize to govern the evolution of endosymbionts [39].

Effect of Richer Media on Cross-Feeding Prevalence

Another factor that may likely affect the prevalence at which obligate cross-feeding emerges is the media on which the community grows. Media that is richer than the minimal media used in our primary simulations may inhibit the establishment of cross-feeding since the evolving species will be able to get required resources from the environment without interacting with their partner. Accordingly, we expect that evolution on richer media would increase the prevalence of independent communities at the expense of commensal, mutualistic, and collapsed communities.

To investigate this potential impact of richer media, we first ran 500 additional simulations using a carbon rich media (note that the number of simulations used in this section is substantially smaller than that used in the primary text due to time constraints). This media was similar to the minimal media used in our primary set of simulations but contained five different monosaccharides: glucose, fructose, galactose, mannose, and ribose (in contrast to the minimal media that contain glucose as the sole carbon source). We found that simulations using this carbon rich media in fact resulted in a similar prevalence of independent and mutualistic communities as that observed in simulation on minimal media (52.6% vs. 51% and 2.2% vs. 3%, respectively; $P > 0.05$; $\chi 2$ test). This media did impact however the prevalence of collapsed and commensal communities, resulting in a significant depletion of collapsed communities compared to simulations using the minimal media (5% vs. 10.7% on minimal media; $P < 10\text{-}4$) and more commensal communities (40.2% vs. 35.3%; $P < 0.04$; $\chi 2$ test). This finding suggests that the availability of diverse carbon sources may allow communities that would otherwise collapse to

survive owning to these additional nutrients in the environment but does not on its own give rise to a significant increase in independent communities.

To explore the effect of other rich media, we further ran simulations using three types of amino-acid rich media. The first contained (in addition to the content of the minimal media) the nine amino-acids that are essential in humans (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine). The second similarly contained in addition to the minimal media the eleven conditionally or non-essential amino-acids in humans (alanine, arginine, aspartate, asparagine, cysteine, glutamate, glutamine, glycine, proline, serine, and tyrosine). The third contained all twenty amino-acids. We first ran 300 simulations using either the essential amino-acids or the non-essential amino-acids media. We found that these media indeed resulted in a decrease in obligate cross-feeding and an increase in independent communities relative to minimal media. Specifically, simulations using the essential amino-acids media resulted in more independent communities (64.7% vs. 51%; $P < 10\text{-}5$) and fewer collapsed communities (1.7% vs. 10.7%; $P < 10\text{-}6$). The prevalence of commensal and mutual communities was generally similar to that observed in minimal media (31.7% vs. 35.3 and 2% vs. 3%, respectively). Simulations using the non-essential amino-acids media similarly resulted in more independent communities than minimal media (66.7% vs. 51%; $P < 10\text{-}7$), fewer commensal communities (26.7% vs. 35.3%; $P < 0.01$), fewer mutualistic communities (0.3% vs. 3%; $P < 0.01$) and fewer collapsed communities (6.3% vs. 10.7%; $P < 0.05$). Finally, we ran 300 simulations using the media that contained all amino-acids. As expected, this very rich media resulted in dramatically different prevalence of the various interaction types, with markedly more independent communities compared to simulations on minimal media (94% vs. 51%, $P < 10\text{-}10$), significantly less commensal communities (6% vs. 35.3%; $P < 10\text{-}10$), and no mutualistic or

collapsed communities (0% vs. 3%; $P < 0.01$ and 0% vs. 10.7%; $P < 10\text{-}8$, respectively). This striking result suggests that the presence of all amino-acids in the media was sufficient to drastically reduce the frequency of dependence evolving.

Effect of Genomic Evolution Pattern on Cross-Feeding Prevalence

In addition to the media, it is also possible that the pattern by which gene deletions occur could affect the likelihood with which obligate cross-feeding evolves in our simulations. Specifically, we wished to examine whether a non-symmetric deletion rate (i.e., when one species is more likely to have genes deleted than the other) or a multi-gene deletion strategy (i.e., having more than one gene deleted at a time) would impact our findings. To explore this, we ran additional simulation sets (again, using a more limited number of simulation rans) that employ different strategies for choosing which genes to delete. In the first set, we added a bias to the process of selecting which gene to delete, whereby genes in one species were twice as likely to be deleted as the genes in the other species. Running 300 simulations with this strategy we found that the prevalence of the various interaction types was similar to that observed in our main simulation set (independent: 51%, commensal: 34.7%, mutualistic: 2.7%, collapsed: 11.7%; $P > 0.5$ for all compared to unbiased deletions). We additionally did not find significant differences in the average genome size of the deletion-prone species compared to the other ($P > 0.5$, two sample t-test). In the second simulation set, we modeled deletion of larger portions of the genome by deleting two adjacent genes at a time (i.e., pairs of genes that are next to each other based on genomic position). Once no two-gene deletions are possible (i.e., due to large fitness effects) the simulation switches back to deleting genes one at a time (as in our primary simulation) until the minimal genomes are reached. Running 300 simulations using this alternative deletion method, we found that it resulted in fewer independent communities (42.7%

vs. 51%; $P < 0.01$) and more collapsed communities (16.7% vs. 10.7%; $P < 0.001$). The prevalence of commensal and mutualistic communities was similar to that observed in our primary simulation set (39.3% and 1.3%, respectively). This finding suggests that deletion of larger portions of the genome at a time could increase the frequency of cross-feeding dependence but could also destabilize such relationships.

Genome Size of Species Evolved in Mono-Culture Conditions

To study the effect of the co-culture condition on the size of the evolved minimal genomes we performed additional simulations in mono-culture growth conditions. Specifically, we ran 2000 evolution simulations that were similar to the simulations described in the main text except that they included only a single species. We found that the resulting minimal genomes of these species were similar in size to that of species in independent pairs (297.87 vs. 297.85), with no statistically significant difference in size ($P > 0.5$, two sample t-test). As with the independent species' genomes, the mono-culture species' genomes were significantly smaller than commensal provider genomes ($P < 10\text{-}6$) and significantly larger than commensal dependent genomes ($P < 10\text{-}20$).

Associating Gene Retention with Interaction Type

We compared mutualistic and independent species and identified a set of 80 genes that are deleted at a significantly higher frequency in mutualistic species (chi-square test; FDR 1%; Table S1B). This set was enriched for various pathways, including oxidative phosphorylation, histidine metabolism, lipopolysaccharide biosynthesis, and nitrogen metabolism. For example, the gene tyrA, necessary for tyrosine synthesis, was never deleted in independent species but in 31.9% of mutualistic species, reflecting frequent tyrosine cross-feeding in mutualistic

communities (as also observed above). We additionally identified a set of 132 genes that are deleted at a significantly higher frequency in independent species, though this set was not significantly enriched for any specific pathway. Interestingly, however, the gene with the greatest difference in retention rate in this set was *tyrP* (deleted in 96.3% of independent species but only in 68.2% of mutualist species) which is necessary for tyrosine transport, reflecting the need of mutualistic species to uptake this metabolite. Moreover, the larger number of genes identified to be deleted more often in independent species, despite the fact that independent species were found to retain a larger number of total genes compared to mutualistic species, suggests that a multitude of mutualist strategies exist, each involving the retention of a different subset of genes.

Similarity of Retained Gene Sets between Community Partners

We examined how similar, on average, are the sets of genes retained between the two partners in each community. We found that mutualistic species were less similar to each other than independent species ($P < 10\text{-}20$; two sample t-test). This finding suggests that the evolution of metabolic dependency is associated with a process of functional diversification, where each of the two species retains certain metabolic capacities that the other species has lost. This result is perhaps not surprising given the phenotypic differences that must emerge to give rise to mutual dependence on cross-feeding. To further investigate this diversification process we turned our attention again to commensal communities, in which the two species can be labeled as dependent and provider and therefore the direction of dependency is clear. Such communities also represent an intermediate level of interaction as compared to mutualistic communities, in which each species is acting as both provider and dependent, which complicates dissection of the mechanism of interaction. Indeed, the two partner species in commensal communities were more similar to

one another than species in mutualistic communities (P = 1.4 x 10-7) but more divergent than species in independent communities (P < 10-20).

The Dynamics of Gene Deletion Events

We set out to examine the dynamics of gene deletion events in commensal communities, specifically focusing on the order in which deletions occurred in the provider and dependent species and aiming to detect dependencies between these deletion events that could highlight key evolutionary steps on the route to cross-feeding. To this end, we used a permutation-based analysis to identify instances where a gene in one species tended to be deleted after another gene was deleted in the partner species (see Supporting Methods). We identified 9 such gene pairs (at 1% FDR), all of which involved one gene in the provider tending to be deleted first before a second, different gene was deleted in the dependent. Specifically, deletion of the *tyrA* gene in the dependent often followed deletion of a set of genes (*talA*, *talB*, *aroP*, and *pheP*) in the provider. *talA* and *talB* catalyze a reaction connecting glycolysis to the pentose phosphate pathway, and their deletion likely disrupts central carbon metabolism and diverts excess flux toward aromatic amino acid biosynthesis. Similarly, *aroP* and *pheP* are both transporters capable of transporting phenylalanine, and their deletion potentially prompts the excretion of tyrosine instead of phenylalanine. These deletions therefore promote over production and excretion of tyrosine by the provider, allowing the dependent to lose *tyrA*, a gene necessary for tyrosine synthesis. The deletion of the *pheA*, a gene necessary for phenylalanine synthesis in the dependent, was also found to follow the deletion of *talA* and *talB* in the provider, which is not surprising given the similarity in the biosynthesis pathways of these two amino acids. Finally the deletion of *pyrG* in the dependent tended to follow the deletion of *cdd*, *cmk*, and *codA* in the provider. The deletion of *cmk* (necessary for recycling CMP into CTP) and of *cdd* and *codA* (catalyzing reactions that

could convert CMP or related products into other bases) could result in cytidine excretion and accordingly allows the dependent to lose *pyrG* (a component of CTP synthase) which creates a dependency on cytidine (Figure A.2). To further examine the mechanism involved in these interactions, we tested whether the deletions of these key genes are sufficient to cause over-production and excretion of the relevant metabolites. Indeed, we found that deletion of *cdd*, *cmk*, and *codA* in the ancestral species (i.e., without any additional gene deletions) led to cytidine excretion. Deletion of *talA*, *talB*, *aroP*, and *pheP* in the ancestral species, however, was insufficient to cause excretion of either phenylalanine or tyrosine, suggesting that additional gene deletions are necessary to give rise to this phenotype.

The Number of Metabolites Excreted by Species Throughout Evolution and its Association with Evolved Interactions

We examined the total number of different metabolites being excreted by each species over time, hypothesizing that species that excrete useful metabolites early on are more likely to become provider species. Surprisingly, however, we found that during the first half of the evolutionary process future providers in fact tend to excrete a similar or even a smaller number of metabolites on average compared to future dependents (Figure A.3), and only toward the end of the evolutionary process did providers excrete more metabolites than dependent species. This pattern could suggest that species that eventually became dependent were less optimal early on, excreting more waste products, and that this wasteful behavior may have led to the development of dependence. Notably, all species gradually excrete more metabolites over the course of the evolution process, likely reflecting more complex growth strategies imposed by their shrinking genome.

Supporting Methods

Evolution Simulation

The evolution simulation was initiated with a pair of genome scale metabolic models. For this study all simulations were initiated with two identical copies of the iAF1260 *E. coli* model [89]. This model includes 1260 genes, 2382 reactions, and 1668 metabolites (which includes extracellular, periplasmic, and cytoplasmic versions of some of the same metabolites). For 304 of those metabolites the model contains transport and exchange reactions that allow them to be exchanged with the external environment. During each step in the evolutionary process, a gene in one of the two species was chosen uniformly at random from the set of all genes still retained by the two species. The chosen gene and all the metabolic reactions that can no longer be performed without this gene were deleted from the species' model. The iAF1260 model contains genes that are required for multiple metabolic reactions as well as genes with redundant effects, and as a result the set of reactions that will be lost as a result of a gene deletion is contingent on which genes have already been deleted previously. The fitness effect of this gene deletion in the context of the community was determined using the co-culture growth model (see below) to evaluate the growth rate of the reduced model when grown with the current model of the partner species. If the calculated fitness effect (when compared to the fitness of that species prior to this gene deletion) was positive, neutral, or smaller than the chosen cutoff (cutoffs used include 1%, 5%, and 10%), the deletion became permanent and the process repeated with the reduced model. However, if the fitness effect exceeded the cutoff, the deletion was considered to be too harmful to occur and the process repeated until a gene that could be deleted was found. This evolutionary process continued until deletion of any remaining gene from either of the two species would

cause a drop in fitness exceeding the cutoff, in which case the simulation ended. The simulation also ended if the chosen gene deletion in one species (i.e., a gene deletion that was relatively harmless for that species) caused the other species to drop significantly in fitness (>50%) in the co-culture. Such simulations, where a partner species was no longer being supported, represent collapsed communities.

Co-Culture Growth Simulation

The co-culture growth simulation was based on a previously introduced dynamic Flux-Balance Analysis framework and is described in more detail elsewhere [55]. Briefly, given a multi-species community inhabiting a shared medium, the framework assumed that at each time step, each species grew optimally given the current concentration of metabolites in the medium (i.e., selfish growth), and then updated the abundance of each species and the concentration of metabolites in the medium based on the predicted growth and activity of each species. Specifically, at each time step, the framework first calculated the upper bound on metabolites' uptake for each species based on the concentration of metabolites in the medium and the cell density of each species. A Flux Balance Analysis (FBA) was then used to determine the fluxes through each species' reactions given these uptake constraints by maximizing the species' biomass production (as a proxy for growth). A second optimization was performed to minimize the total flux through all reactions while keeping the biomass production fixed at the maximum rate (representing a minimization of enzyme usage). The predicted growth rate of each species and the predicted rates at which each species uptakes and excretes various metabolites were then used to update the cell density and concentrations of metabolites in the medium. The process was then repeated at the next time step.

For the purpose of this study, each co-culture simulation consisted of 8 steps of 0.125 hours followed by 4 steps of 0.5 hours. This provided a more accurate account of species growth at the initiation of any potential interaction, while still providing information about the co-culture growth at a longer time scale. The growth rates at the last time point (i.e., after 2.5 hours) were used as a measure of each species' fitness. Both species started at a biomass of 0.01 grams dry mass in 1L volume for mono-culture or 2L for co-culture, resulting in the same cell density for both (which is equal to about 4*10^7 cells per liter for *E. coli*). The species were grown on a medium based on M9 minimal media [90], containing sodium, chloride, sulfate, inorganic phosphate, potassium, magnesium, ammonia, glucose, water, hydrogen, and oxygen. In addition the metals copper, iron, molybdate, manganese, zinc, nickel, and cobalt were included as they are necessary for growth of the *E. coli* model. These metabolites were all present in the medium at an excess concentration of 10M to ensure exponential growth for the entire course of the co-culture simulation. A low concentration (0.0001 mM) of jumpstart metabolites were also included to allow growth of obligate mutualistic pairs (see below). FBA solutions were calculated using glpkmex, a Matlab interface for GLPK, GNU Linear Programing Kit. GLPK version 4.54 was used, and glpkmex version 2.11.

Jumpstarting Mutualistic Growth

Simulating the growth of species that evolved to be obligate mutualists with a dynamic FBA model has the inherent problem that neither species is able to grow initially on the minimal medium (and consequently will not excrete any of the byproducts needed to allow the other species to grow). In biological systems this problem can be overcome by heterogeneity in the growth phenotypes of individual cells, nutrients released by dead cells, or trace nutrients present in the environment. Rather than simulating diverse growth phenotypes or cell death, in this study

we jumpstarted mutualistic growth by supplementing the minimal growth medium described above with trace amounts of potentially necessary metabolites. The set of these "jumpstart metabolites" was determined by identifying metabolites that could be produced by non-transfer reactions still present in at least one of the two species (even if the pathway was not complete). This set therefore represented an upper bound on which metabolites could be exchanged. Jumpstart metabolites were initialized at a low concentration of 0.0001 mM. To ensure that species that utilized these metabolites for growth could eventually be supported by the production of these metabolites by the partner species (rather than continually relying on the trace amounts of these metabolite provided at the beginning of the simulation), at 1 hour into the growth simulation, this same low concentration (0.0001 mM) was subtracted from each jumpstart metabolite.

Filtering Completed Simulation Runs

Simulation runtime considerations necessitated using a relatively limited time resolution in the co-culture growth simulation (see above). To confirm that the growth patterns observed in evolved communities (including obligate cross-feeding interactions) were not artifacts caused by this limited time resolution, for each completed simulation we ran additional co-culture growth simulations on the resulting minimal models using a finer time resolution. Specifically, co-culture growth was simulated until the medium was exhausted with time steps of 0.1 hours, using otherwise the same conditions as the co-culture growth model employed during evolution (including removing the jumpstart metabolites at 1 hour). Community growth was deemed to have been accurately simulated if:

1. Glucose eventually ran out, indicating that the two species were able to continue growing stably.

2. Both species were able to continue growth until this exhaustion of the media. Growth of both species must have been at least 50% of their measured fitness value within the last hour before all growth ended.
3. The growth rate of both species at 2.5 ± 0.2 hours was at least 90% of their fitness value as measured during the course of the evolution.

Simulations that failed any of these three criteria were excluded from the downstream analysis. Of the 16377 completed simulation runs, 16, 317 (99.6%) passed this filtering step. Examples of an evolved mutualistic community that passed this filtering step and an evolved commensal community that failed this filtering step are illustrated in Figure A.4.

Simulation Runtime

Each simulation (i.e., a single evolutionary trajectory) required on average 12.2 (± 3.7 STD) hours on a single CPU core. This rather extensive runtime is the outcome of the many FBA estimations performed in each such simulation. Specifically, it should be noted that not only did each 'generation' (i.e., gene deletion) in the evolutionary trajectory require 24 FBA optimizations (12 time points for each species) to obtained the necessary data, but that in fact for each generation our framework may have first tested several potential gene deletions that proved too deleterious and therefore were not selected (see Methods). This step was extremely time-consuming, particularly toward the end of the evolutionary trajectory when more and more (and eventually all) remaining genes *cannot* be selected for deletion. As a result, a single non-collapsed evolutionary simulation required on average ~6,650 co-culture simulations (each involving 24 FBA optimizations), as well as ~1925 additional mono-culture simulations for each of the two species (each involving 12 FBA optimizations for each), resulting in ~205,800 FBA optimization per evolutionary simulation. Simulations were run on a shared high performance

computing cluster containing 23 Intel E5345 or E5410 CPUs with 8 cores each, 3 AMD Opteron 6168 CPUs with 48 cores each, and 3 AMD Opteron 6278 CPUs with 64 cores each. Around 80-100 simulations were run simultaneously through the cluster's job scheduling system, and 3 GB of memory was requested for each simulation. The complete set of simulations took about 3-4 months to run.

Determining Interaction Type

Interaction type was determined by comparing the fitness of each species when grown in co-culture with its fitness when grown in mono-culture. If the fitness of a species at a given time point was zero in mono-culture and non-zero in co-culture, the species was labeled as dependent at that time. If it had non-zero growth in both mono- and co-culture it was labeled as independent. Communities were labeled by the relationships between the two species: If both species were independent, the community was labeled as independent. If one species was dependent and the other was independent, the community was labeled as commensal. If both species in a community were dependent, the community was labeled as mutualistic. Within commensal communities, the dependent species was referred to as 'dependent' and the independent species was referred to as 'provider'.

Determining Metabolic Dependencies

For each evolved species we determined what metabolites it depends on (if any). To this end, we first identified metabolites that were being exchanged between the two species at the final co-culture growth time point by finding exchange reactions for which the two species had fluxes of opposite sign. The growth of each dependent species in the pair was then assayed on minimal medium supplemented with all possible combinations of these exchanged metabolites,

using a single time step mono-culture growth model, to identify the smallest set of supplement metabolites that allowed it to grow at >50% its growth rate in co-culture. If no combination of supplement metabolites allowed such growth the search was expanded to include all combinations of metabolites present in the medium at the end of the co-culture simulation and that were not part of the minimal media (such metabolites could have been excreted by the provider at previous time steps). In 359 simulations multiple sets of supplement metabolites were sufficient for growth and were neither subsets nor supersets of each other. These ambiguous simulations were excluded from any analysis of metabolite dependency or association.

Pathway Analyses

Pathways annotations for each gene in the model were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [120]. To identify enrichment of KEGG pathways in subsets of genes (e.g., those that were deleted at significantly different rates between interaction types), we generated 100,000 random subsets (out of the 551 genes retained at intermediate frequencies) of the same size and compared the total number of genes associated with each pathway in the real set to the number of genes associated with that pathway in random sets.

Measuring Genome Similarity

To compare the similarity of two genomes (e.g., in an evolved community), we used the Jaccard similarity coefficient. We then used a two-sample t-test to test for significant differences in similarity between different types of communities.

Identifying Co-Retained Gene Pairs

To identify gene-gene co-occurrence relationships, we examined all pairwise combinations of genes that were both retained and deleted at least five times. This cutoff was chosen to restrict analysis to genes for which sample size would provide enough statistical power to confidently identify co-occurrence. Since many genes perfectly co-varied with each other across simulations, we first grouped genes into sets of perfectly co-varying genes and identified co-occurrence relationships between these sets. For each pair of gene sets, we found the number of species that had retained each set and the number of species that had retained both sets and used a hypergeometric test to determine whether these sets have been co-retained significantly more or less often than expected by chance (at 1% false discovery rate; [121]). Test for enrichment of shared pathways among the significant gene set pairs was done by permuting the connections between pairs.

Identifying Significant Gene Deletion Ordering

Given a pair of genes, A and B, we recorded the number of times gene A in the dependent was deleted before gene B in the provider. We then used a permutation-based assay, permuting the time of deletion (measured as the position in the ordering of all gene deletions in that simulation) of each gene between all providers or dependents from commensal communities in which that gene had been deleted. Gaps and overlaps in the resulting permuted gene deletion histories were resolved by shifting deletions into gaps and randomly breaking ties. The number of times gene A in the dependent was deleted before gene B in the provider in the original data

was compared to this number in the permuted data to identify significantly common ordered pairs of gene deletes (at 1% false discovery rate).

Gene-Metabolite Connections

To identify correlations between retention or deletion of specific genes and metabolic phenotypes, we considered all genes that were both deleted and retained more than 10 times and all metabolites that were depended upon at least 10 times. These cutoffs were chosen to restrict analysis to genes and metabolites for which sample size would provide statistical power to confidently identify significant correlations. For every pair of such genes and metabolites, we compared the frequency of deletion of that gene in commensal species that are dependent on that metabolite to the frequency of deletion of that gene in independent species. This was repeated for commensal species that provided the metabolite their partner depends upon, and in both cases instances of genes being deleted more often or retained more often in species with that metabolic phenotype were identified (at 1% false discovery rate).

Supporting Figures

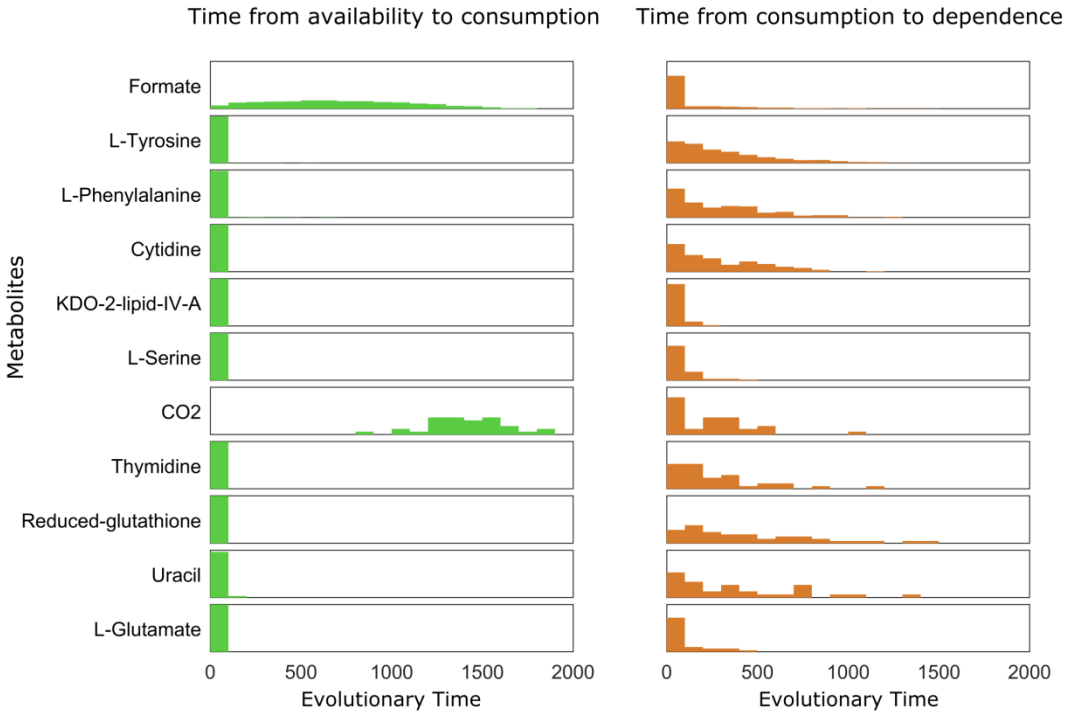

Figure A.1. Time of emergence of metabolites' availability, cross-feeding, and dependence. The distributions of evolutionary time (measured as number of gene deletions) elapsed between the different stages of metabolic interaction, for commensal species dependent on a single metabolite..

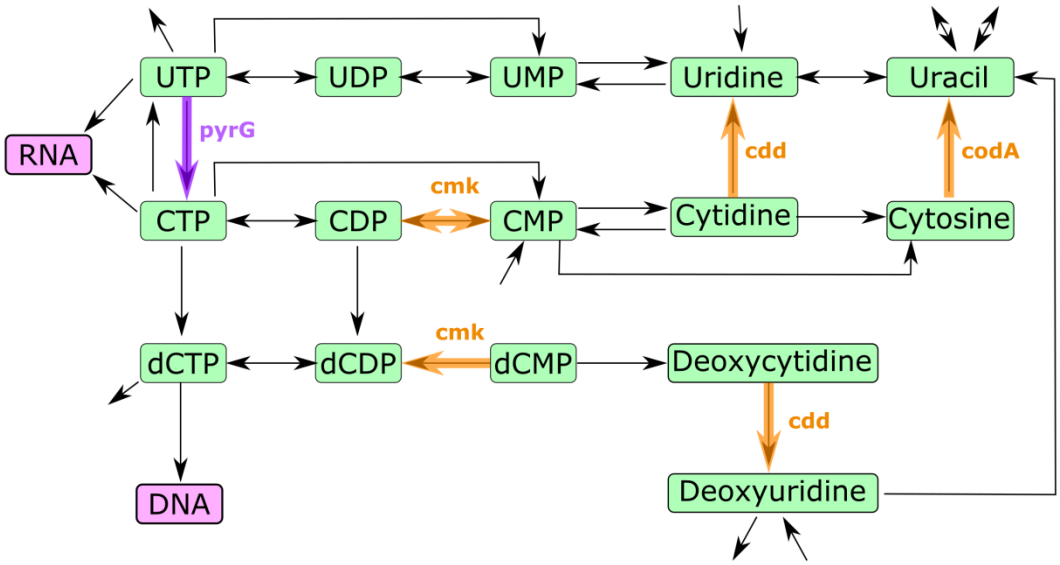

Figure A.2. Dependent deletion events in pyrimidine biosynthesis.
The subset of the pyrimidine biosynthesis network in the E. coli model is shown. Reactions catalyzed by enzymes encoded by genes that tend to be lost first in the provider are marked in orange (*cmk*, *cdd*, and *codA*), and the reaction catalyzed by an enzymes encoded by a gene that is lost later in the dependent is marked in purple (*pyrG*).

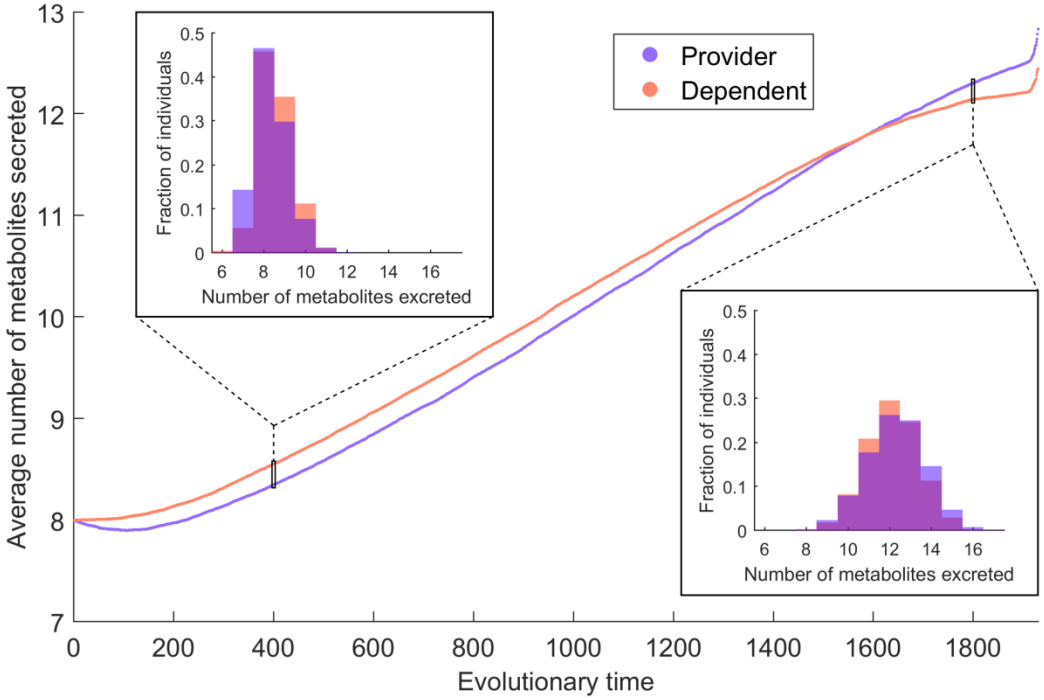

Figure A.3. The number of different metabolites excreted by species of different interaction types over evolutionary time.
Plotted is the mean number of unique metabolites excreted by dependent and provider species from commensal communities at each point throughout he evolutionary process. The two insets show the distribution of the number of metabolites excreted at time 400 and 1800 for both dependents and providers.

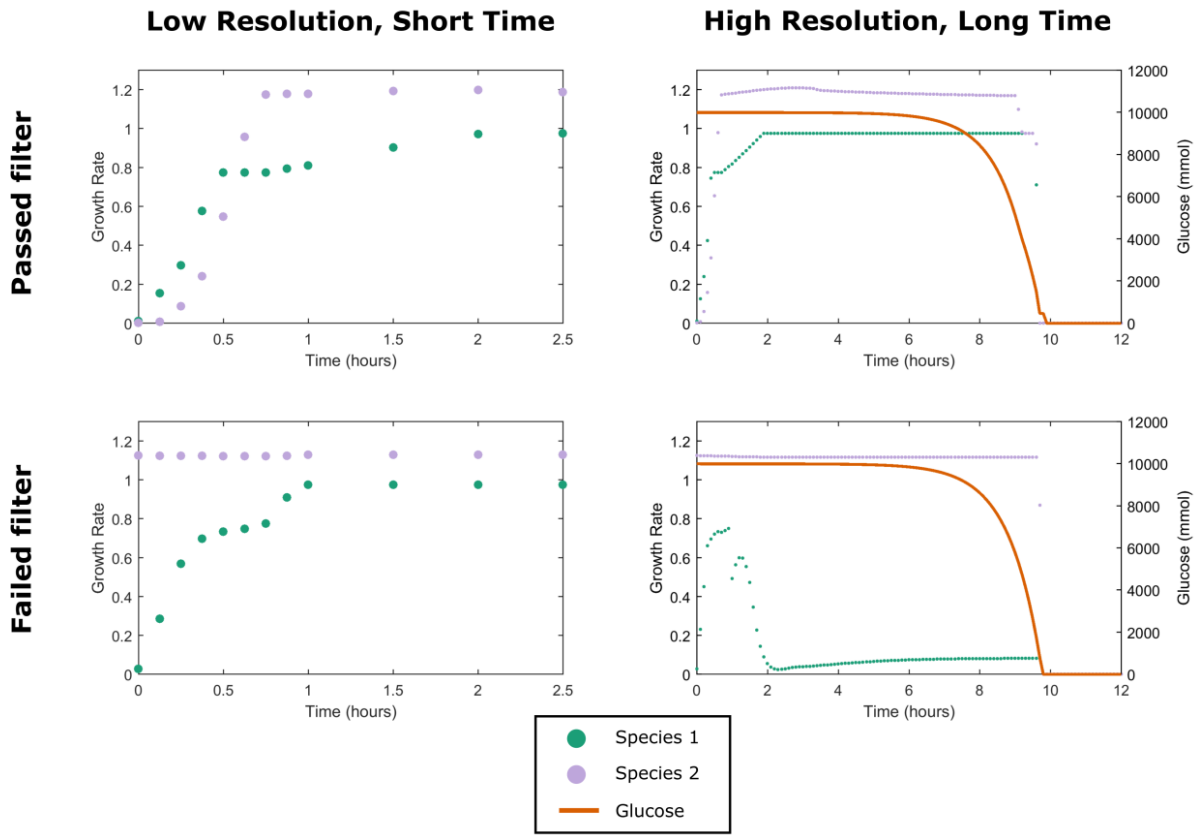

Figure A.4. Example of co-culture growth of communities that passed and failed the quality filtering step.

The co-culture growth of evolved pairs are plotted both for the settings used during the evolution simulation (left: low resolution, short time) and the settings used to validate growth during filtering (right: high resolution, long time). On the top is an evolved mutualistic community that passed the filter, and on the bottom is an evolved commensal community that failed the filtering step. When the top community was grown with high resolution time steps both species reached similar growth rates to that seen at the last time point of the low resolution simulation, and maintained roughly those growth rates until glucose was exhausted. The bottom community fails to show such consistent behavior. Specifically species 1 never reaches the growth rate it achieved in the low resolution simulation, and instead collapses to a very low growth rate shortly after the jumpstart metabolites are removed (1 hour).

REFERENCES

1. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A. National Academy of Sciences; 1998;95: 6578–83. doi:10.1073/PNAS.95.12.6578

2. Falkowski PG, Fenchel T, Delong EF. The Microbial Engines That Drive Earth's Biogeochemical Cycles. Science (80-). American Association for the Advancement of Science; 2008;320: 1034–1039. doi:10.1126/science.1153213

3. Francis CA, Beman JM, Kuypers MMM. New processes and players in the nitrogen cycle: the microbial ecology of anaerobic and archaeal ammonia oxidation. ISME J. Nature Publishing Group; 2007;1: 19–27. doi:10.1038/ismej.2007.8

4. Tiedje JM, Asuming-Brempong S, Nüsslein K, Marsh TL, Flynn SJ. Opening the black box of soil microbial diversity. Appl Soil Ecol. Elsevier; 1999;13: 109–122. doi:10.1016/S0929-1393(99)00026-8

5. Grady Jr CPL, Daigger GT, Love NG, Filipe CDM. Biological wastewater treatment. CRC press; 2011.

6. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, et al. Structure, function and diversity of the healthy human microbiome. Nature. Nature Publishing Group; 2012;486: 207–214. doi:10.1038/nature11234

7. Thaiss CA, Zmora N, Levy M, Elinav E. The microbiome and innate immunity. Nature. Nature Publishing Group; 2016;535: 65–74. doi:10.1038/nature18847

8. Keren L, Hausser J, Lotan-Pompan M, Vainberg Slutskin I, Alisar H, Kaminski S, et al. Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness [Internet]. Cell Elsevier; Aug 27, 2016 p. 1282–1294.e18. doi:10.1016/j.cell.2016.07.024

9. West SA, Diggle SP, Buckling A, Gardner A, Griffin AS. The Social Lives of Microbes. Annu Rev Ecol Evol Syst. 2007;38: 53–77. doi:10.1146/annurev.ecolsys.38.091206.095740

10. Hibbing ME, Fuqua C, Parsek MR, Peterson SB. Bacterial competition: surviving and thriving in the microbial jungle. Nat Rev Microbiol. Nature Publishing Group; 2010;8: 15–25. doi:10.1038/nrmicro2259

11. Traore AS, Fardeau M-L, Hatchikian CE, Le Gall J, Belaich J-P. Energetics of Growth of a Defined Mixed Culture of Desulfovibrio vulgaris and Methanosarcina barkeri: Interspecies Hydrogen Transfer in Batch and Continuous Cultures. Appl Envir Microbiol. 1983;46: 1152–1156. Available: http://aem.asm.org/content/46/5/1152.short

12. Boetius A, Ravenschlag K, Schubert CJ, Rickert D, Widdel F, Gieseke A, et al. A marine microbial consortium apparently mediating anaerobic oxidation of methane. Nature. 2000;407: 623–626. doi:10.1038/35036572

13. Rakoff-Nahoum S, Foster KR, Comstock LE. The evolution of cooperation within the gut microbiota. Nature. 2016;advance on. doi:10.1038/nature17626

14. Mee MT, Collins JJ, Church GM, Wang HH. Syntrophic exchange in synthetic microbial communities. Proc Natl Acad Sci U S A. 2014;111: E2149-56. doi:10.1073/pnas.1405641111

15. Wintermute EH, Silver PA. Emergent cooperation in microbial metabolism. Mol Syst Biol. Nature Publishing Group; 2010;6. doi:10.1038/msb.2010.66

16. Pande S, Merker H, Bohl K, Reichelt M, Schuster S, de Figueiredo LF, et al. Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. ISME J. 2014;8: 953–62. doi:10.1038/ismej.2013.211

17. Foster KR, Bell T. Competition, not cooperation, dominates interactions among culturable microbial species. Curr Biol. Elsevier Ltd; 2012;22: 1845–50. doi:10.1016/j.cub.2012.08.005

18. Verster AJ, Ross BD, Radey MC, Bao Y, Goodman AL, Mougous JD, et al. The Landscape of Type VI Secretion across Human Gut Microbiomes Reveals Its Role in Community Composition. Cell Host Microbe. Elsevier; 2017;22: 411–419.e4. doi:10.1016/j.chom.2017.08.010

19. Ho BT, Dong TG, Mekalanos JJ. A View to a Kill: The Bacterial Type VI Secretion System. Cell Host Microbe. Cell Press; 2014;15: 9–21. doi:10.1016/J.CHOM.2013.11.008

20. Miller MB, Bassler BL. Quorum Sensing in Bacteria. Annu Rev Microbiol. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA ; 2001;55: 165–199. doi:10.1146/annurev.micro.55.1.165

21. Nadell CD, Drescher K, Foster KR. Spatial structure, cooperation and competition in biofilms. Nat Rev Microbiol. 2016;14: 589–600. doi:10.1038/nrmicro.2016.84

22. Kaeberlein T, Lewis K, Epstein SS. Isolating “Uncultivable” Microorganisms in Pure Culture in a Simulated Natural Environment. Science (80-). 2002;296: 1127–1129. doi:10.1126/science.1070633

23. Nichols D, Cahoon N, Trakhtenberg EM, Pham L, Mehta A, Belanger A, et al. Use of ichip for high-throughput in situ cultivation of "uncultivable" microbial species. Appl Environ Microbiol. American Society for Microbiology; 2010;76: 2445–50. doi:10.1128/AEM.01754-09

24. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, et al. Emergent simplicity in microbial community assembly. Science. American Association for the Advancement of Science; 2018;361: 469–474. doi:10.1126/science.aat1168

25. Celiker H, Gore J. Clustering in community structure across replicate ecosystems following a long-term bacterial evolution experiment. Nat Commun. Nature Publishing Group; 2014;5: 4643. doi:10.1038/ncomms5643

26. Zelezniak A, Andrejev S, Ponomarova O, Mende DR, Bork P, Patil KR. Metabolic dependencies drive species co-occurrence in diverse microbial communities. Proc Natl Acad Sci. 2015;112: 6449–6454. doi:10.1073/pnas.1421834112

27. Thompson JR, Pacocha S, Pharino C, Klepac-Ceraj V, Hunt DE, Benoit J, et al. Genotypic diversity within a natural coastal bacterioplankton population. Science. American

Association for the Advancement of Science; 2005;307: 1311–3. doi:10.1126/science.1106028

28. Greenblum S, Carr R, Borenstein E. Extensive Strain-Level Copy-Number Variation across Human Gut Microbiome Species. Cell. Elsevier {BV}; 2015;160: 583–594. doi:10.1016/j.cell.2014.12.038

29. Van Valen L. A new evolutionary law. Evol Theory. 1973;1: 1–30.

30. Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, Thomson NR, et al. Antagonistic coevolution accelerates molecular evolution. Nature. Nature Publishing Group; 2010;464: 275–278. doi:10.1038/nature08798

31. Nowak MA. Five rules for the evolution of cooperation. Science. 2006;314: 1560–3. doi:10.1126/science.1133755

32. Morris JJ, Lenski RE, Zinser ER. The Black Queen Hypothesis : Evolution of Dependencies through Adaptive Gene Loss. MBio. 2012;3: 1–7. doi:10.1128/mBio.00036-12.Copyright

33. Morris JJ, Papoulis SE, Lenski RE. Coexistence of evolving bacteria stabilized by a shared black queen function. Evolution (N Y). 2014;68: 2960–71. doi:10.1111/evo.12485

34. Baumann P. Biology of Bacteriocyte-Associated Endosymbionts of Plant Sap-Sucking Insects. Annu Rev Microbiol. Annual Reviews; 2005;59: 155–189. doi:10.1146/annurev.micro.59.030804.121041

35. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol. Nature Publishing Group; 2011;10: 13–26. doi:10.1038/nrmicro2670

36. Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, Khouri H, et al. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol. 2006;4: e188. doi:10.1371/journal.pbio.0040188

37. Van Leuven JT, Meister RC, Simon C, McCutcheon JP. Sympatric Speciation in a Bacterial Endosymbiont Results in Two Genomes with the Functionality of One. Cell. Elsevier Inc.; 2014;158: 1270–1280. doi:10.1016/j.cell.2014.07.047

38. Łukasik P, Nazario K, Van Leuven JT, Campbell MA, Meyer M, Michalik A, et al. Multiple origins of interdependent endosymbiotic complexes in a genus of cicadas. Proc Natl Acad Sci U S A. National Academy of Sciences; 2018;115: E226–E235. doi:10.1073/pnas.1712321115

39. Bennett GM, Moran NA. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. Proc Natl Acad Sci. National Academy of Sciences; 2015;112: 10169–10176. doi:10.1073/pnas.1421388112

40. Moran NA, Bennett GM. The Tiniest Tiny Genomes. Annu Rev Microbiol. Annual Reviews ; 2014;68: 195–215. doi:10.1146/annurev-micro-091213-112901

41. West SA, Fisher RM, Gardner A, Kiers ET. Major evolutionary transitions in individuality. Proc Natl Acad Sci U S A. National Academy of Sciences; 2015;112: 10112–9. doi:10.1073/pnas.1421402112

42. Axelrod R. The Evolution of Cooperation. Basic Books; 1984.

43. Nowak M, Sigmund K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. Nature. Nature Publishing Group; 1993;364: 56–58. doi:10.1038/364056a0

44. Nowak MA, Sasaki A, Taylor C, Fudenberg D. Emergence of cooperation and evolutionary stability in finite populations. Nature. Nature Publishing Group; 2004;428: 646–650. doi:10.1038/nature02414

45. Hauert C, Doebeli M. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. Nature. Nature Publishing Group; 2004;428: 643–646. doi:10.1038/nature02360

46. Doebeli M. A model for the evolutionary dynamics of cross-feeding polymorphisms in microorganisms. Popul Ecol. Springer-Verlag; 2002;44: 59–70. doi:10.1007/s101440200008

47. Pfeiffer T, Bonhoeffer S. Evolution of cross-feeding in microbial populations. Am Nat. 2004;163: E126–E135. doi:10.1086/383593

48. Germerodt S, Bohl K, Lück A, Pande S, Schröter A, Kaleta C, et al. Pervasive Selection for Cooperative Cross-Feeding in Bacterial Communities. Igoshin OA, editor. PLOS Comput Biol. Public Library of Science; 2016;12: e1004986. doi:10.1371/journal.pcbi.1004986

49. Mitri S, Xavier JB, Foster KR. Social evolution in multispecies biofilms. Proc Natl Acad Sci U S A. National Academy of Sciences; 2011;108 Suppl 2: 10839–46. doi:10.1073/pnas.1100292108

50. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010;28: 245–248. doi:10.1038/nbt.1614

51. Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol Syst Biol. 2007;3: 119. doi:10.1038/msb4100162

52. Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng. 2003;5: 264–276. doi:10.1016/j.ymben.2003.09.002

53. Covert MW, Schilling CH, Palsson B. Regulation of Gene Expression in Flux Balance Models of Metabolism. J Theor Biol. Academic Press; 2001;213: 73–88. doi:10.1006/JTBI.2001.2405

54. Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh J a, et al. Metabolic modeling of a mutualistic microbial community. Mol Syst Biol. 2007;3: 92. doi:10.1038/msb4100131

55. Chiu H-C, Levy R, Borenstein E. Emergent Biosynthetic Capacity in Simple Microbial Communities. Ouzounis CA, editor. PLoS Comput Biol. 2014;10: e1003695. doi:10.1371/journal.pcbi.1003695

56. Klitgord N, Segrè D. Environments that Induce Synthetic Microbial Ecosystems. PLoS Comput Biol. 2010;6. doi:10.1371/Citation

57. Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M, et al. Competitive and cooperative metabolic interactions in bacterial communities. Nat Commun. Nature Publishing Group; 2011;2: 589. doi:10.1038/ncomms1597

58. Großkopf T, Consuegra J, Gaffé J, Willison JC, Lenski RE, Soyer OS, et al. Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment. BMC Evol Biol. BioMed Central; 2016;16: 163. doi:10.1186/s12862-016-0733-x

59. Little AEF, Robinson CJ, Peterson SB, Raffa KF, Handelsman J. Rules of engagement: interspecies interactions that regulate microbial communities. Annu Rev Microbiol. Annual Reviews; 2008;62: 375–401. doi:10.1146/annurev.micro.030608.101423

60. Wintermute EH, Silver PA. Dynamics in the mixed microbial concourse. Genes Dev. 2010;24: 2603–14. doi:10.1101/gad.1985210

61. McInerney MJ, Struchtemeyer CG, Sieber J, Mouttaki H, Stams AJM, Schink B, et al. Physiology, Ecology, Phylogeny, and Genomics of Microorganisms Capable of Syntrophic Metabolism. Ann N Y Acad Sci. Blackwell Publishing Inc; 2008;1125: 58–72. doi:10.1196/annals.1419.005

62. Sieber JR, McInerney MJ, Gunsalus RP. Genomic Insights into Syntrophy: The Paradigm for Anaerobic Metabolic Cooperation. Annu Rev Microbiol. Annual Reviews ; 2012;66: 429–452. doi:10.1146/annurev-micro-090110-102844

63. Belenguer A, Duncan SH, Calder AG, Holtrop G, Louis P, Lobley GE, et al. Two routes of metabolic cross-feeding between Bifidobacterium adolescentis and butyrate-producing anaerobes from the human gut. Appl Environ Microbiol. American Society for Microbiology; 2006;72: 3593–9. doi:10.1128/AEM.72.5.3593-3599.2006

64. McCutcheon JP, McDonald BR, Moran NA. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc Natl Acad Sci U S A. 2009;106: 15394–9. doi:10.1073/pnas.0906424106

65. McCutcheon JP, Moran NA. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc Natl Acad Sci U S A. 2007;104: 19392–7. doi:10.1073/pnas.0708855104

66. Morris JJ. Black Queen evolution: the role of leakiness in structuring microbial communities. Trends Genet. Elsevier Ltd; 2015;31: 475–482. doi:10.1016/j.tig.2015.05.004

67. D'Souza G, Waschina S, Pande S, Bohl K, Kaleta C, Kost C. Less is more: selective advantages can explain the prevalant loss of biosynthetic genes in bacteria. Evolution (N Y). 2014;68: 2559–70. doi:10.1111/evo.12468

68. D'Souza G, Kost C, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Schneider D. Experimental Evolution of Metabolic Dependency in Bacteria. Hughes D, editor. PLOS Genet. 2016;12: e1006364. doi:10.1371/journal.pgen.1006364

69. Opatovsky I, Santos-Garcia D, Lahav T, Ofaim S, Mouton L, Barbe V, et al. Modeling trophic dependencies and exchanges among insects' bacterial symbionts in a host-simulated environment. bioRxiv. 2016; doi:10.1101/086165

70. Treves DS, Manning S, Adams J. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of Escherichia coli. Mol Biol Evol. 1998;15: 789–797. doi:10.1093/oxfordjournals.molbev.a025984

71. Rosenzweig RF, Sharp RR, Treves DS, Adams J. Microbial evolution in a simple unstructured environment: Genetic differentiation in Escherichia coli. Genetics. 1994;137: 903–917.

72. Mendes-Soares H, Chia N. Community metabolic modeling approaches to understanding the gut microbiome: Bridging biochemistry and ecology. Free Radic Biol Med. Pergamon; 2017;105: 102–109. doi:10.1016/J.FREERADBIOMED.2016.12.017

73. Oliveira NM, Niehus R, Foster KR. Evolutionary limits to cooperation in microbial communities. Proc Natl Acad Sci. 2014;111: 17941–17946. doi:10.1073/pnas.1412673111

74. Bull JJ, Harcombe WR. Population dynamics constrain the cooperative evolution of cross-feeding. PLoS One. 2009;4: e4115. doi:10.1371/journal.pone.0004115

75. Pál C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD. Chance and necessity in the evolution of minimal metabolic networks. Nature. 2006;440: 667–70. doi:10.1038/nature04568

76. Yizhak K, Tuller T, Papp B, Ruppin E. Metabolic modeling of endosymbiont genome reduction on a temporal scale. Mol Syst Biol. European Molecular Biology Organization; 2011;7. doi:10.1038/msb.2011.11

77. Mahadevan R, Edwards JS, Doyle FJ. Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys J. Elsevier; 2002;83: 1331–40. doi:10.1016/S0006-3495(02)73903-9

78. Hudson GS, Wong V, Davidson BE. Chorismate mutase/prephenate dehydrogenase from Escherichia coli K12: purification, characterization, and identification of a reactive cysteine. Biochemistry. 1984;23: 6240–9.

79. Belfort M, Maley GF, Maley F. Characterization of the Escherichia coli thyA gene and its amplified thymidylate synthetase product. Proc Natl Acad Sci U S A. 1983;80: 1858–61. Available: http://www.pnas.org/content/80/7/1858.abstract

80. Fricke J, Neuhard J, Kelln RA, Pedersen S. The cmk gene encoding cytidine monophosphate kinase is located in the rpsA operon and is required for normal replication rate in Escherichia coli. J Bacteriol. 1995;177: 517–23. Available: http://jb.asm.org/content/177/3/517.abstract

81. Maughan H, Callicotte V, Hancock A, Birky CW, Nicholson WL, Masel J. The population genetics of phenotypic deterioration in experimental populations of Bacillus subtilis. Evolution (N Y). Blackwell Publishing Ltd; 2006;60: 686–695. doi:10.1111/j.0014-3820.2006.tb01148.x

82. Zomorrodi AR, Segrè D. Genome-driven evolutionary game theory helps understand the

rise of metabolic interdependencies in microbial communities. Nat Commun. Nature Publishing Group; 2017;8: 1563. doi:10.1038/s41467-017-01407-5

83. Barve A, Hosseini S-R, Martin OC, Wagner A. Historical contingency and the gradual evolution of metabolic properties in central carbon and genome-scale metabolisms. BMC Syst Biol. 2014;8: 1–20. doi:10.1186/1752-0509-8-48

84. Zomorrodi AR, Maranas CD. OptCom: A Multi-Level Optimization Framework for the Metabolic Modeling and Analysis of Microbial Communities. Rao C V., editor. PLoS Comput Biol. Public Library of Science; 2012;8: e1002363. doi:10.1371/journal.pcbi.1002363

85. Russell JB, Diez-Gonzalez F. The effects of fermentation acids on bacterial growth. Adv Microb Physiol. 1998;39: 205–34. Available: http://www.ncbi.nlm.nih.gov/pubmed/9328648

86. Gómez-Valero L, Rocha EPC, Latorre A, Silva FJ. Reconstructing the ancestor of Mycobacterium leprae: the dynamics of gene loss and genome reduction. Genome Res. Cold Spring Harbor Laboratory Press; 2007;17: 1178–85. doi:10.1101/gr.6360207

87. Nilsson AI, Koskiniemi S, Eriksson S, Kugelberg E, Hinton JCD, Andersson DI. Bacterial genome size reduction by experimental evolution. Proc Natl Acad Sci U S A. National Academy of Sciences; 2005;102: 12112–6. doi:10.1073/pnas.0503654102

88. Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, et al. Metabolic Resource Allocation in Individual Microbes Determines Ecosystem Interactions and Spatial Dynamics. Cell Rep. 2014;7: 1104–15. doi:10.1016/j.celrep.2014.03.070

89. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol. EMBO and Nature Publishing Group; 2007;3: 121. doi:10.1038/msb4100155

90. Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. Second edi. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 1989.

91. Johnson MTJ, Stinchcombe JR. An emerging synthesis between community ecology and evolutionary biology. Trends Ecol Evol. 2007;22: 250–257. doi:10.1016/j.tree.2007.01.014

92. Casey P, Zee PC, Heath KD, Miller TE, Pastore AI, Patel S, et al. Evolution in a Community Context : Trait Responses to Multiple Species Interactions *. 2018;191: 368–380. doi:10.1086/695835

93. Iwao K, Rausher MD. Evolution of Plant Resistance to Multiple Herbivores: Quantifying Diffuse Coevolution. Am Nat. University of Chicago Press ; 1997;149: 316–335. doi:10.1086/285992

94. Cothran RD, Jeyasingh PD. Condition dependence of a sexually selected trait in a crustacean species complex: Importance of the ecological context. Evolution (N Y). Wiley/Blackwell (10.1111); 2010;64: 2535–2546. doi:10.1111/j.1558-5646.2010.00998.x

95. McPeek MA, Schrot AK, Brown JM. Adaptation to Predators in a New Community:

Swimming Performance and Predator Avoidance in Damselflies. Ecology. Wiley-Blackwell; 1996;77: 617–629. doi:10.2307/2265635

96. Gómez P, Buckling A. Real-time microbial adaptive diversification in soil. Grover J, editor. Ecol Lett. Wiley/Blackwell (10.1111); 2013;16: 650–655. doi:10.1111/ele.12093

97. Hietpas RT, Jensen JD, Bolon DNA. Experimental illumination of a fitness landscape. Proc Natl Acad Sci U S A. National Academy of Sciences; 2011;108: 7896–901. doi:10.1073/pnas.1016024108

98. Manor O, Levy R, Borenstein E. Mapping the inner workings of the microbiome: Genomic- and metagenomic-based study of metabolism and of metabolic interactions in the human gut microbiome. Cell Metab. Elsevier Inc.; 2014;20: 742–52. doi:10.1016/j.cmet.2014.07.021

99. Garud NR, Good BH, Hallatschek O, Pollard KS. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. bioRxiv. Cold Spring Harbor Laboratory; 2018; 210955. doi:10.1101/210955

100. Zhao S, Lieberman TD, Poyet M, Groussin M, Gibbons SM, Xavier RJ, et al. Adaptive evolution within the gut microbiome of individual people. bioRxiv. Cold Spring Harbor Laboratory; 2018; 208009. doi:10.1101/208009

101. Ghoul M, Mitri S. The Ecology and Evolution of Microbial Competition. Trends Microbiol. Elsevier Ltd; 2016;24: 833–845. doi:10.1016/j.tim.2016.06.011

102. Hillesland KL, Lim S, Flowers JJ, Turkarslan S, Pinel N, Zane GM, et al. Erosion of functional independence early in the evolution of a microbial mutualism. Proc Natl Acad Sci U S A. National Academy of Sciences; 2014;111: 14822–7. doi:10.1073/pnas.1407986111

103. Hillesland KL, Stahl DA. Rapid evolution of stability and productivity at the origin of a microbial mutualism. Proc Natl Acad Sci U S A. 2010;107: 2124–2129. doi:10.1073/pnas.0908456107

104. Lawrence D, Fiegna F, Behrends V, Bundy JG, Phillimore AB, Bell T, et al. Species interactions alter evolutionary responses to a novel environment. PLoS Biol. 2012;10. doi:10.1371/journal.pbio.1001330

105. Meadows AL, Karnik R, Lam H, Forestell S, Snedecor B. Application of dynamic flux balance analysis to an industrial Escherichia coli fermentation. Metab Eng. Academic Press; 2010;12: 150–160. doi:10.1016/j.ymben.2009.07.006

106. Bordbar A, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. Nat Rev Genet. Nature Publishing Group; 2014;15: 107–120. doi:10.1038/nrg3643

107. Notebaart R a., Szappanos B, Kintses B, Pal F, Gyorkei a., Bogos B, et al. Network-level architecture and the evolutionary potential of underground metabolism. Proc Natl Acad Sci. 2014;111: 11762–7. doi:10.1073/pnas.1406102111

108. Chubiz LM, Granger BR, Segrè D, Harcombe WR. Species interactions differ in their genetic robustness. Front Microbiol. 2015;6: 1–9. doi:10.3389/fmicb.2015.00271

109. Calcagno V, Jarne P, Loreau M, Mouquet N, David P. Diversity spurs diversification in ecological communities. Nat Commun. Nature Publishing Group; 2017;8: 15810. doi:10.1038/ncomms15810

110. Jousset A, Eisenhauer N, Merker M, Mouquet N, Scheu S. High functional diversity stimulates diversification in experimental microbial communities. Sci Adv. American Association for the Advancement of Science; 2016;2: e1600124–e1600124. doi:10.1126/sciadv.1600124

111. Strauss SY, Strauss SY, Sahli H, Conner JK. Toward a more trait-centered approach to diffuse (co)evolution. New Phytol. 2004;165: 81–90.

112. Faith JJ, McNulty NP, Rey FE, Gordon JI. Predicting a human gut microbiota's response to diet in gnotobiotic mice. Science. 2011;333: 101–4. doi:10.1126/science.1206025

113. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat Biotechnol. Nature Publishing Group; 2016;35: 81–89. doi:10.1038/nbt.3703

114. Clark D, Cronan JE. Escherichia coli mutants with altered control of alcohol dehydrogenase and nitrate reductase. J Bacteriol. American Society for Microbiology; 1980;141: 177–83. Available: http://www.ncbi.nlm.nih.gov/pubmed/6986356

115. Ruczinski I, Kooperberg C, LeBlanc M. Logic Regression. J Comput Graph Stat. Taylor & Francis; 2003;12: 475–511. doi:10.1198/1061860032238

116. Brockhurst MA, Colegrave N, Hodgson DJ, Buckling A. Niche occupation limits adaptive radiation in experimental microcosms. Borevitz J, editor. PLoS One. Public Library of Science; 2007;2: e193. doi:10.1371/journal.pone.0000193

117. Gogarten JP, Townsend JP. Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol. Nature Publishing Group; 2005;3: 679–687. doi:10.1038/nrmicro1204

118. Endler JA. Multiple-trait coevolution and environmental gradients in guppies [Internet]. Trends in Ecology & Evolution. Elsevier Current Trends; 1995. pp. 22–29. doi:10.1016/S0169-5347(00)88956-9

119. Xu Z, Knight R. Dietary effects on human gut microbiome diversity. Br J Nutr. Cambridge University Press; 2015;113: S1–S5. doi:10.1017/S0007114514004127

120. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. Oxford University Press ({OUP}); 2000;28: 27–30. doi:10.1093/nar/28.1.27

121. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. [Royal Statistical Society, Wiley]; 1995;57: 289–300. Available: http://www.jstor.org/stable/2346101

122. Nowak MA, May RM. Evolutionary games and spatial chaos. Nature. Nature Publishing Group; 1992;359: 826–829. doi:10.1038/359826a0

123. Peaudecerf FJ, Bunbury F, Bhardwaj V, Bees MA, Smith AG, Goldstein RE, et al. Microbial mutualism at a distance: The role of geometry in diffusive exchanges. Phys Rev E. American Physical Society; 2018;97: 022411. doi:10.1103/PhysRevE.97.022411

124. Kim HJ, Boedicker JQ, Choi JW, Ismagilov RF. Defined spatial structure stabilizes a synthetic multispecies bacterial community. Proc Natl Acad Sci U S A. National Academy of Sciences; 2008;105: 18188–93. doi:10.1073/pnas.0807935105

125. de Jong IG, Haccou P, Kuipers OP. Bet hedging or not? A guide to proper classification of microbial survival strategies. BioEssays. Wiley-Blackwell; 2011;33: 215–223. doi:10.1002/bies.201000127

126. Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science. American Association for the Advancement of Science; 2018;361: 1380–1385. doi:10.1126/science.aau0730

127. Brenner K, You L, Arnold FH. Engineering microbial consortia: a new frontier in synthetic biology. Trends Biotechnol. Elsevier Current Trends; 2008;26: 483–9. doi:10.1016/j.tibtech.2008.05.004

128. Basu S, Mehreja R, Thiberge S, Chen M-T, Weiss R. Spatiotemporal control of gene expression with pulse-generating networks. Proc Natl Acad Sci U S A. National Academy of Sciences; 2004;101: 6355–60. doi:10.1073/pnas.0307571101

129. Tamsir A, Tabor JJ, Voigt CA. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires.' Nature. 2011;469: 212–215. doi:10.1038/nature09565

130. Chen Y, Kim JK, Hirning AJ, Josić K, Bennett MR. Emergent genetic oscillations in a synthetic microbial consortium. Science. American Association for the Advancement of Science; 2015;349: 986–9. doi:10.1126/science.aaa3794

VITA

Colin was born in San Francisco and grew up in Davis, California. He enjoys scuba diving, aquariums, and movies, and is confident that food equals love.