2015

Controlling microbial community dynamics through engineered metabolic dependencies

BOSTON UNIVERSITY

COLLEGE OF ENGINEERING

Dissertation

CONTROLLING MICROBIAL COMMUNITY DYNAMICS THROUGH ENGINEERED METABOLIC DEPENDENCIES

by

MICHAEL TRAVIS MEE

B.Eng., McGill University, 2009

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2015

Approved by

First Reader

George M. Church, Ph.D.
Robert Winthrop Professor of Genetics
Harvard Medical School

Second Reader

James J. Collins, Ph.D.
Termeer Professor of Bioengineering
Massachusetts Institute of Technology

Third Reader

Ahmad S. Khalil, Ph.D.
Assistant Professor of Biomedical Engineering
Associate Director, Center of Synthetic Biology
Boston University

Fourth Reader

Daniel Segrè, Ph.D.
Associate Professor of Biomedical Engineering
Associate Professor of Bioinformatics
Associate Professor of Biology
Boston University

Fifth Reader

Muhammad H. Zaman, Ph.D.
Professor of Biomedical Engineering
Professor, Howard Hughes Medical Institute
Boston University

ACKNOWLEDGMENTS

There are many people who have impacted my Ph.D. journey. First and foremost I am grateful to my advisor, George, for giving me the opportunity to be a part of one of the most inspiring places to think, play and be a scientist. He has curated an enriching environment that is second-to-none that had a critical influence on my development. Many colleagues in the lab have been fantastic mentors. Harris Wang has been an inspiration to work alongside and a wealth of insight. Without him as a sounding board, my thoughts on microbial interactions would never have crystalized as they did. Dan Mandell and Marc Lajoie were a joy to work with and learn from on a truly thrilling collaboration. I would also like to thank my committee, for the time and thought they put towards advancing my project. Without my parent's unwavering support and the curiosity they instilled in me this journey may never have begun. My brothers and extended family have always been inspirational role models. I've also been very lucky to have the support of close friends: Raffi Afeyan for his loyalty and reminding me what friendship is all about; Phil Dawson for his encouragement and counsel on all maters of life; and Nick D'Antono for enduring my many frantic late night Skype calls. Finally, Nadège, who puts everything into perspective, entertains my inner geek and arouses my creative side – I couldn't have done it without you. To you all, I am eternally grateful.

CONTROLLING MICROBIAL COMMUNITY DYNAMICS THROUGH ENGINEERED METABOLIC DEPENDENCIES

MICHAEL TRAVIS MEE

Boston University College of Engineering, 2015

Major Professor: George M. Church, Ph.D., Robert Winthrop Professor of Genetics, Harvard Medical School

ABSTRACT

Metabolic cross-feeding is an important process that can broadly shape microbial communities. Comparative genomic analysis of >6000 sequenced bacteria from diverse environments provides evidence to suggesting that amino acid biosynthesis has been broadly optimized to reduce individual metabolic burden in favor of enhanced cross-feeding to support synergistic growth across the biosphere. Still, little is known about specific cross-feeding principles that drive the formation and maintenance of individuals within a mixed population. Here, we devised a series of synthetic syntrophic communities to probe the complex interactions underlying metabolic exchange of amino acids. We experimentally analyzed multi-member, multi-dimensional communities of *Escherichia coli* of increasing sophistication to assess the outcomes of synergistic cross-feeding. We find that biosynthetically costly amino acids including methionine, lysine, isoleucine, arginine and aromatics, tend to promote stronger cooperative interactions than amino acids that are cheaper to produce. Furthermore, cells that share common intermediates along branching pathways yielded more synergistic growth, but exhibited many instances of both positive

and negative epistasis when these interactions scaled to higher-dimensions. This system enabled the identification of synergistic pairings and optimal expression levels of amino acid exporters of arginine, threonine and aromatics towards drastic improvements of ecosystem productivity. Tradeoffs identified in these mutualistic systems between secretion, relative abundance and absolute community productivity have implication in the evolution of cooperative behaviors. Long-term evolution of these synthetic communities highlight transporter over-expression, amino acid pool redistribution, and perturbations to nitrogen regulation as strategies to circumvent imposed metabolic dependencies. To address this potentially problematic genomic plasticity, a genetically reassigned organism is leveraged to investigate synthetic metabolic dependencies showing improved biocontainment and potential for microbial consortia control. These results improve our basic understanding of microbial syntrophy while also highlighting the utility and limitations of current approaches to modeling and controlling the dynamic complexities of microbial ecosystems. This work sets a foundation for future endeavors in microbial ecology and evolution, and presents a platform to develop better and more robust engineered synthetic communities for industrial biotechnology.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

1.1 Motivation For Microbial Ecosystem Engineering

Microbes constitute the most abundant and diverse set of organisms on Earth(Achtman & Wagner, 2008; Schloss & Handelsman, 2004). By generating and turning over organic material, they play a dominant role in performing key biochemical reactions essential to sustaining the biosphere(Falkowski, Fenchel, & Delong, 2008). As such, these micron-sized cells have evolved an impressive array of strategies that have allowed them to grow in almost any environment on the planet(Fraser, Alm, Polz, Spratt, & Hanage, 2009). Microbes, however, do not live alone. Rather, they live in crowded environments in association with other microbes, competing for resources, sharing metabolism, and forming a complex, dynamic and evolving microbial ecosystem(Hibbing, Fuqua, Parsek, & Peterson, 2010; Klitgord & Segre, 2011).

In nature, stable microbial consortia are generally composed of members that have specialized physiologies and are tasked with different roles. These intertwined roles transform individuals that would otherwise compete, into a group that lives in concert(Pocock, Evans, & Memmott, 2012). Many such microbial ecosystems have evolved to be highly refractory to perturbations in the environment and are able to repopulate themselves when depleted in numbers. We are now beginning to appreciate the myriad of sophisticated processes and behaviors that manifest in microbial consortia, some of which mirror many essential features found in higher-level metazoans and multicellular

organisms(Foster, 2011). Understanding how individual microbes form communities will bring new and important insight to the evolution of multicellularity(Ispolatov, Ackermann, & Doebeli, 2012). A grand challenge in applied biology is to develop the knowledge and technology necessary to build these self-adaptive systems that can perform complex tasks at the micron-scale. Therefore, engineering microbial communities is an important endeavor, ripe for pursuit by synthetic biologists.

Over the past decade, the field of synthetic biology has aimed to make biology easier to engineer(Endy, 2005; Khalil & Collins, 2010). Under the paradigm of traditional engineering, new conceptual frameworks were devised to describe the organization of genetic regulation and cellular machinery to build new metabolisms(Canton, Labno, & Endy, 2008; Medema, van Raaphorst, Takano, & Breitling, 2012). New tools for the synthesis, assembly, and engineering of genes have been scaled to whole genomes to enable faster prototyping of biological designs(Carr & Church, 2009). Standardized inventories of useful genes and other biological components are growing rapidly(Muller & Arndt, 2012). All of these efforts help us develop a better understanding of the cell and the underlying design principles for engineering it. Scaling these efforts to communities of cells will require the development of new frameworks, methods and technologies (Figure 1).

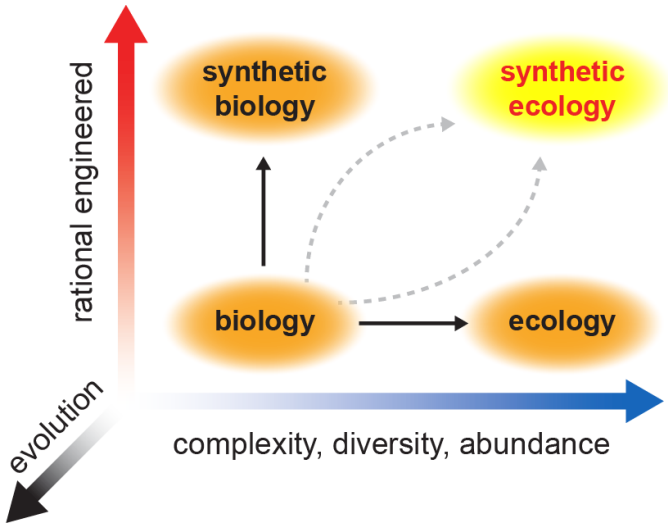

Figure 1: Development of synthetic ecology requires insights gained through manipulating simple biological systems and analyzing complex ecological systems. Evolution must be factored into these pursuits, not only as a destabilizing force but also as a means to optimize our engineered designs.

1.2 Engineering Parameters

What goes on in microbial communities can be quite complicated to understand, appearing almost irreducibly complex. Therefore, engineering such a system is a daunting task. Even when grossly approximating a cell as a linear input-output unit, we are confronted with the observation that interactions between cells generate behaviors that are non-linear, asynchronous, and heterogeneous. Toward building a framework for engineering synthetic microbial ecosystems, we outline a set of essential parameters that we believe are core features of a microbial community. These parameters should be the subject of analysis, perturbation, and optimization when building synthetic ecosystems *de novo*. Based on recent literature about natural and engineered ecologies, we

highlight these parameters with regard to their significance, relationship with one another, and tunability from a synthetic perspective. These parameters help to build a framework for microbial communities where the individual members interact with one another through exchange of material, energy, and information (Figure 2).

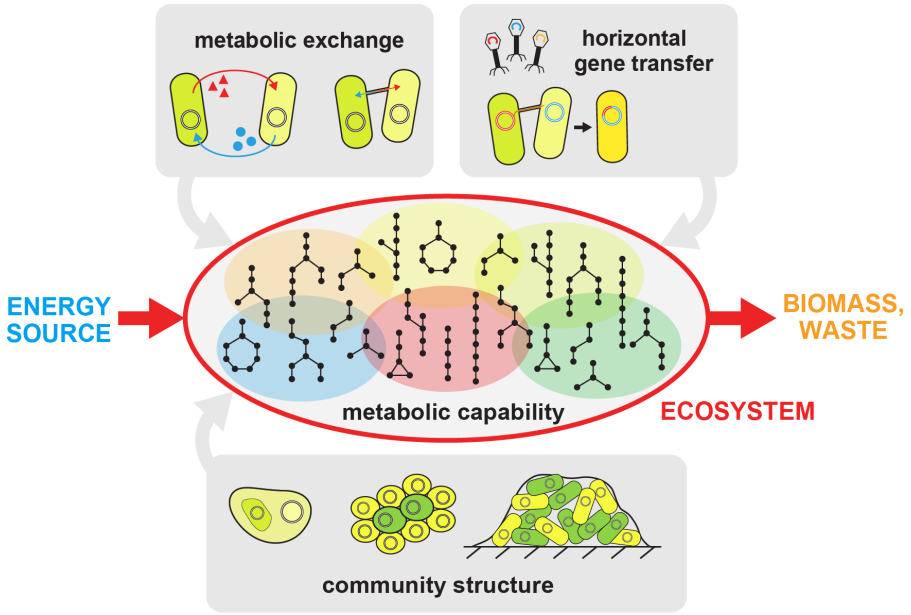

Figure 2: A summary of the crucial parameters that impact a microbial ecosystem. These parameters determine the ecosystem's ability to convert an energy source into biomass and waste, and are prime targets for engineering and optimization. Metabolic capabilities are distributed across different members as defined by metabolotypes (shaded and colored ovals). Metabolic exchange can occur via metabolite transport across cellular membranes or through intercellular bridges. Community structure can be tuned by adjusting the degree of aggregation and formation of extracellular structures such as biofilms. Horizontal gene transfer enables genomic innovation and the rise of new capabilities within the population.

1.2.1 Metabolic Capabilities and Metabolotypes

Metabolism is the core essence of life at all scales, from individual enzymatic reactions in each cell all the way to the ecosystem as a whole. In

nature, the goal of metabolism is to extract energy from substrates, use them to synthesize biomass, and leave behind waste byproducts. For any given environment, we can argue that the residing consortium of cells performs a set of input-output operations to generate biomass and waste from an initial source of energy (e.g. sunlight, sugar, other biomass, etc.). The black-box operation that the consortium performs may in fact be very complicated depending on the metabolic capability and efficiency of the members, as well as their abundance and diversity. In fact, many different arrangements can be functionally equivalent because microbes house a staggering array of metabolic capabilities in a near infinite number of combinations. Over the past decades, we have cataloged a significant portion of all possible chemical and enzymatic reactions that biology can perform in databases such as KEGG(Kanehisa, Goto, Sato, Furumichi, & Tanabe, 2012) or MetaCyc(Caspi et al., 2012). With computers and *in silico* models, we can now recreate cellular metabolism for well-studied organisms(Duarte et al., 2007; Duarte, Herrgard, & Palsson, 2004; Oh, Palsson, Park, Schilling, & Mahadevan, 2007; Reed, Vo, Schilling, & Palsson, 2003; Schilling et al., 2002). Therefore, a deeper understanding of how metabolism scales to communities of cells can now be achieved.

The total metabolic capability of a microbial community arises from the summation of capabilities of each individual member. Identification of a cell's metabolism is not a trivial task, however. Traditional taxonomic classification of microbial species by 16S rRNA(Woese & Fox, 1977) profiling is a poor reflector

of metabolic functionality. For example, communities that are only 15% similar as profiled by 16S may be 70% similar in terms of metabolic capability as determined by metagenomic sequencing(Burke, Steinberg, Rusch, Kjelleberg, & Thomas, 2011). Furthermore, we have a poor understanding of how metabolic capabilities that are distributed across different individuals can impact the community as a whole. We do know that with sufficient functional redundancy in the population, system-level behavior can be stably maintained even though individuals may vary in abundance(Fernandez et al., 1999; Turnbaugh et al., 2009). Therefore, to have a clear picture of community-level metabolism, it is essential to identify the total list of metabolic genes, how they are allocated among individual members, and the level of redundancy in the system. We believe that the *metabolotype*, or the range of metabolic capabilities of any individual cell, may be a more relevant identifier of consortium members than the standard 16S phylogenetic signature. Metabolotype can be derived from the genotype via comparative genomic analyses(Goh et al., 2006) or from the phenotype via experimental characterizations(Bochner, Gadzinski, & Panomitros, 2001). Engineering metabolotypes may provide important avenues to tune the metabolic capacity, dynamics, and diversity of the ecosystem.

1.2.2 Intercellular Exchange of Metabolites and Signals

In order to understand intercellular metabolic interactions (i.e. those occurring between cells), we need to understand the trafficking of metabolites

across the cell membrane. The cell membrane provides an essential function: trapping enzymes and metabolites within the cytosol to increase their effective local concentration, thereby increasing their rate of catalysis. Any metabolic interaction between cells must require metabolites and intermediates to cross the membrane barrier. For most valuable metabolites, passive diffusion across the membrane barrier is very limited and active transport systems are needed. These molecular transport pumps vary in terms of specificity (general vs. specific pumps), directionality (symport, antiport), and energy requirement (ATP-dependency)(Borths, Poolman, Hvorup, Locher, & Rees, 2005; Patzlaff, van der Heide, & Poolman, 2003; Saier, 2000). Controlling these transport processes is an important thrust in microbial ecosystem engineering.

While most cells have a myriad of transporters that import metabolites, far fewer transporters that export metabolites out of the cell have been identified. It is thought that most exporters (or efflux pumps) mainly serve to remove toxic or antagonistic compounds such as antibiotics from the cell(X. Z. Li & Nikaido, 2009). More recent studies have suggested that these exporters are important in the maintenance of cellular homeostasis by regulating intracellular metabolite concentrations(Burkovski & Kramer, 2002). For example, a number of exporters exist to prevent excessive accumulation of different amino acids such as R, Y, W, F, L, M, K, I(Cruz-Ramos, Cook, Wu, Cleeter, & Poole, 2004; Doroshenko et al., 2007; Eggeling & Sahm, 2003; Franke, Resch, Dassler, Maier, & Bock, 2003; Kutukova et al., 2005; Peeters, Nguyen Le Minh, Foulquie-Moreno, & Charlier,

2009). From the microbial community perspective, these transport systems are critical in enabling selective, and potentially programmable, metabolite sharing between cells with different metabolotypes. In addition to extracellular exchange, other strategies for metabolite sharing exist. Nano-tubules or pilus-based structures enable direct cell-to-cell exchange by establishment of cytosolic bridges(Dubey & Ben-Yehuda, 2011; C. S. Hayes, Aoki, & Low, 2010). These systems allow larger macromolecules such as polypeptides, proteins and DNA/RNA to be exchanged, thus providing additional means to metabolically connect individual cells within a community.

Microbes interact not only through interdependent metabolisms, but also by coordinated behaviors. Group behavior differentiates microbial communities that are merely collections of individuals from those that truly work in a concerted fashion. Coordinating behavior at the population level requires chemical signals and intercellular communication systems such as quorum sensing(Bassler & Losick, 2006). Quorum sensing is the ability of cells to detect population density by measuring the concentration of a membrane-permeable chemical signal. These communication molecules serve to trigger genetic programs across the cell population to elicit synchronized behavior, such as cell division, differentiation, and aggregation(Rath & Dorrestein, 2011; Shank & Kolter, 2011; Straight & Kolter, 2009). From an engineering perspective, we can co-opt these chemical communication systems for synthetic ecosystems. Using synthetic quorum sensing circuits, Weiss et al. generated cell communities that exhibited

different spatially-defined phenotypes in response to chemical gradients(Basu, Gerchman, Collins, Arnold, & Weiss, 2005). These circuits have been further developed for edge detection systems that allow cells to sense the state of adjacent neighbors and respond accordingly(Tabor et al., 2009), as well as for macro-scale synchronization of behavior across physical distances 1000 times greater than the length of a cell(Prindle et al., 2012). These examples of engineered synthetic communities illustrate that controllable cell-cell signaling can enable the design of even more complex systems.

1.2.3 Aggregation and Physical Structure

Metabolic exchange and intercellular interactions require cells to be in close proximity. Cellular aggregation, by cell-cell contact or generation of extracellular matrices (known as biofilms), is a common strategy that natural microbial communities use to increase their local cell density(Hall-Stoodley, Costerton, & Stoodley, 2004). Often, cell aggregates directly lead to the formation of biofilms(Hall-Stoodley & Stoodley, 2002). Biofilm structures are particularly common as they anchor communities to a surface, allowing them to thrive more stably than in an otherwise mixed environment. By strengthening the local interactions in a community, these extracellular structures further enrich for ecosystems that behave cooperatively and in concert. Biofilms also decrease permeability of toxins and antimicrobial compounds thereby protecting the entire community(Fux, Costerton, Stewart, & Stoodley, 2005). These structures provide

tantalizing opportunities for synthetic engineering. For example, Brenner and Arnold et al. developed an engineered biofilm community with increased cooperative growth and resilience to fluctuating environments(Brenner & Arnold, 2011). These systems should be further engineered for directed reciprocity – the ability for individuals to recognize and foster cooperative partners. Directed reciprocity is often found in naturally structured communities such as plant-mycorrhizal ecosystems(Kiers et al., 2011) and other symbiotic systems(Ruby, 2008).

An extreme case of cell-cell association is endosymbiosis(McCutcheon & Moran, 2012). The engulfment of one cell by another and the sustainment of such association can lead to the development of complementary physiologies. It is thought that eukaryotic organelles such as the chloroplast and the mitochondria were the result of endosymbiosis(Margulis, 1971). Metabolic interdependency of endosymbionts often rely on exchange of essential metabolites (e.g. amino acids) as is the case for insect endosymbionts such as Tremblaya & Moranella in mealybugs(McCutcheon & von Dohlen, 2011), Buchnera in aphids(Hansen & Moran, 2011) and Sulcia in cicadas(McCutcheon, McDonald, & Moran, 2009). While these systems clearly present fascinating examples of extreme interdependency, we have yet to fully understand the evolutionary processes that lead to endosymbiosis(Dyall, Brown, & Johnson, 2004). Therefore, forward engineering of such systems remains a significant challenge.

1.2.4 Mutation and Gene Flow

The genetic makeup of the cell is not static but subject to constant change. In a microbial consortium, an individual's metabolic capabilities can change over time due to evolution and horizontal gene transfer (HGT)(Burrus & Waldor, 2004; Frost, Leplae, Summers, & Toussaint, 2005). Small changes to the genome arise from mutations generated during replication or from DNA-damaging agents. Larger changes may arise from mobile genetic elements that move around the same genome and between different genomes(Frost et al., 2005). Small-scale mutations (e.g. point mutations, indels) generally affect the activity, specificity, or expression of proteins, so they are more likely to impact the cell's physiology incrementally(Gogarten & Townsend, 2005). Truly novel traits rarely evolve independently and are more likely to be acquired horizontally from another cell(Frost et al., 2005; van Passel, Marri, & Ochman, 2008). HGT enables the cell to adopt new traits that require large leaps in sequence space, such as new biosynthesis capabilities. These processes can occur via conjugation, natural transformation, recombination, or transduction(Gogarten & Townsend, 2005). So what influences the rate of genetic exchange in communities? Using comparative genomics, Smillie et al. argued that shared ecology is the most important factor that facilitates genetic exchange(Smillie et al., 2011). The rate of HGT can also be accelerated in structured environments when neighboring cells are in close proximity and are more related phylogenetically(Smillie et al., 2011). The level at which Darwinian selection

occurs will affect the distribution and abundance of metabolotypes in the population. In order to effectively engineer ecosystems that behave predictably and stably over time, we must be able to either insulate the system from genetic mutations or harness natural selection to help maintain the engineered and desired state.

1.3 Theoretical and Quantitative Models

Theoretical and quantitative models are valuable analysis tools for studying natural and synthetic microbial ecosystems(Koide, Pang, & Baliga, 2009; Raes & Bork, 2008). While numerous important contributions have been made in this area, they have been for the most part limited by analytical, computational or algorithmic complexity. Since natural ecosystems are highly heterogeneous and nonlinear, molecular-resolution simulations of population-level interactions remain infeasible with current computational resources. Nonetheless, significant progress has been made for *in silico* reconstruction of cell physiology(Feist, Herrgard, Thiele, Reed, & Palsson, 2009). Scaling these models from single cells to ecosystem, however, often demands a compromise in generality. Certain models may highlight individual population-level behavior better than others, but are doing so by sacrificing consideration of another important parameter. Here, we describe four classes of quantitative models that have been developed for understanding microbial ecosystems (Figure 3) and highlight the importance of each.

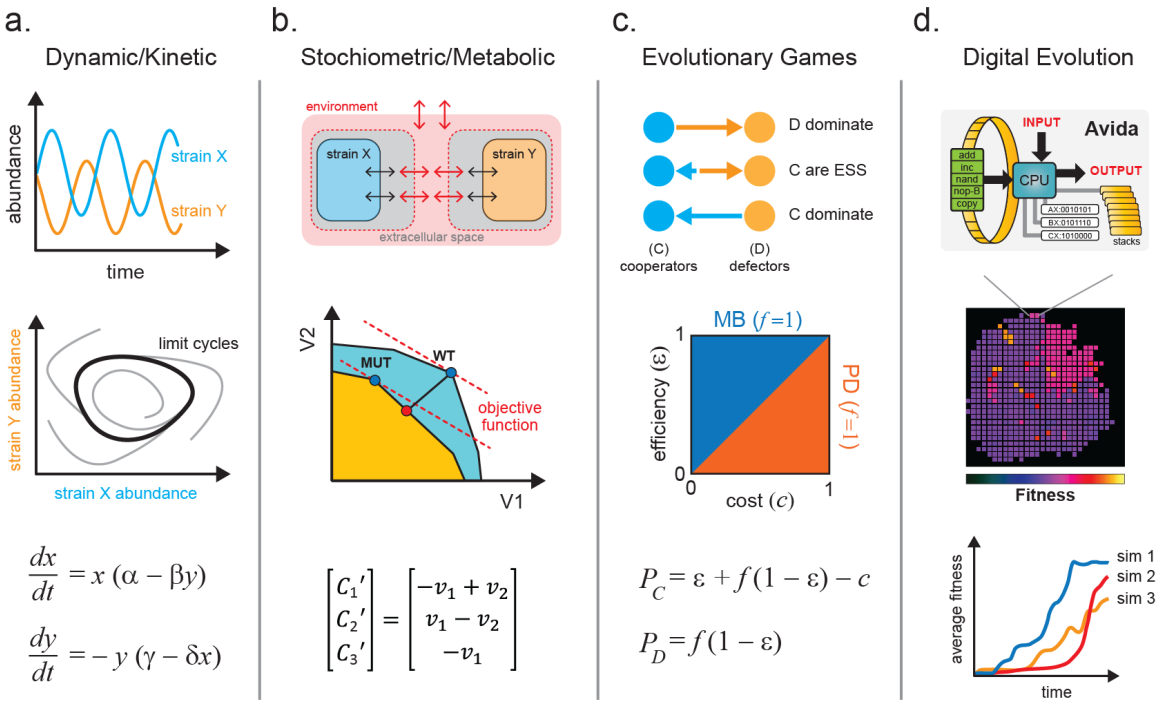

Figure 3: The four main classes of quantitative models that are used to study microbial ecosystems. (a.) Kinetic models describe changes in system variables (e.g. abundance) with simple differential equations that can exhibit interesting dynamics such as oscillations and limit cycles. (b.) Stoichiometric models can be applied to study optimal metabolic flux using objective functions to guide the design of intercellular metabolite exchange. (c.) Evolutionary games can be used to analyze phenotypic strategies within a microbial community using payoff calculations. These models aid in elucidating key variables that influence the domination or coexistence of microbial strategies. (d.) Digital evolution systems help to simulate microbial evolution, traversal of fitness landscapes, development of complex traits, and contributions of epistatic and pleiotropic effects to fitness.

1.3.1 Dynamic Models

Dynamic models are used to predict changes in a system as a function of time. They can be used at various scales from individual metabolites, to proteins, all the way to groups of cells(Chen, Niepel, & Sorger, 2010). In general, concentration or abundance of each component in the system is tracked over time as they interact with one another. In dynamic models, every process in the

system is described by a differential equation. Variables in the equations represent the time-varying parameters being modeled. Coefficients in the equations define the type (e.g. positive or negative) and strength of each interaction. The classical example of such a model is the Lotka-Volterra predator-prey system(van den Ende, 1973). In this system, two subpopulations exist, the predator and the prey. The predator consumes the prey, which leads to depletion of the prey population. A significant depletion of the prey population leads to starvation and decline of the predator population. When the predator population is low, the prey population is then able to thrive, thereby bringing the ecosystem through cycles of boom and bust. The dynamic model is able to capture the expected phasic oscillation in abundance of predator and prey subpopulations and determine parameters in which such associations may exist (Figure 3a)(Balagadde et al., 2008). This model can be scaled to whole populations as long as proper assumptions are made (e.g. linear vs. nonlinear parameter relationships). For example, dynamic models have been successfully applied to study macro-scale systems such as freshwater lake ecosystems(Sahasrabudhe & Motter, 2011). These models also enable perturbation studies where starting conditions (such as population size) can be varied, and solutions are obtained. The largest limitation to these models is that analytical solutions for most nonlinear differential equations with more than two variables are not readily available. Numerical solutions require additional mathematical and computational tools that need to be further developed. Nonetheless, these models are helpful

for us to develop first order intuition about the dynamics of the system.

1.3.2 Stoichiometric Metabolic Models

Stoichiometric models have been developed to study metabolism at the cellular level(Lewis, Nagarajan, & Palsson, 2012). These models describe metabolism of individual cells using matrices containing stoichiometric coefficients of all metabolic reactions and sets of optimization constraints. Stoichiometric representation of metabolism can be analyzed by various approaches(Papin et al., 2004; Price, Reed, Papin, Wiback, & Palsson, 2003) such as Flux Balance Analysis (FBA)(Schilling & Palsson, 1998). In contrast to dynamic models, FBA assumes that the system is at steady state such that all metabolite concentrations are time-invariant. This assumption is likely valid for cells grown in exponential phase(Varma & Palsson, 1994). The solution to the system is described by a series of steady state fluxes for each reaction. By combining all possible fluxes, we can generate a multidimensional flux space that describes the entire metabolic capacity of the cell (Figure 3b). An objective statement is used to define a given flux or criterion, such as flux to biomass (approximating growth rate), for which the multidimensional flux space can be optimized. Through linear optimization, the model predicts metabolic fluxes that maximize the objective function (e.g. biomass). This model has been extensively applied to *in silico* metabolic reconstruction of a variety of organisms(Duarte et al., 2007; Duarte et al., 2004; Oh et al., 2007; Reed et al., 2003; Schilling et al.,

2002). Stolyar et al. used a FBA model to describe a methanogenic community of *M. maripaludis* and *D. vulgaris* that exchanged metabolites hydrogen and formate(Stolyar et al., 2007). The metabolisms of the two strains are divided into two separate compartments which exchange metabolites via a third common compartment. This model successfully predicted the ratio of *M. maripaludis* to *D. vulgaris* during growth and suggested that hydrogen was essential for syntrophy while formate could be removed from the co-culture interaction(Stolyar et al., 2007).

Two developments have greatly improved stoichiometric models of microbial communities: the application of multi-level objective statements(Lewis et al., 2012; Zomorrodi & Maranas, 2012), and inclusion of dynamics(Mahadevan, Edwards, & Doyle, 2002). Multi-level objective statements can be formulated to describe different and potentially competing flux conditions. This approach has been used to model synthetic ecosystems of three or more members, where objective statements are defined separately for both the strain and the community(Zomorrodi & Maranas, 2012). By simultaneously optimizing these objective functions, the model captures the selective forces that act on individuals and the community. For example, growth of individual species can be sacrificed to promote maximal community growth(Zomorrodi & Maranas, 2012). Thus, models with multi-level objectives more accurately describe metabolite exchange. To account for dynamics in the system, population abundance and metabolite concentrations can be separated into different FBA models and

solved independently at every time step in an approach called dynamic multi-species metabolic modeling (DMMM)(Zhuang et al., 2011). As substrate concentrations change over time, DMMM is able to adjust the substrate utilization mode of each strain to the present conditions by switching to the appropriate stoichiometric matrix. This method is able to capture scenarios of resource competition and identify metabolites whose limited exchange affect population dynamics(Zhuang et al., 2011). These and other stoichiometric models, such as elementary mode analysis (EMA)(Taffs et al., 2009), enable full-scale quantitative models of ecosystems that are predictive and important for forward engineering.

1.3.3 Evolutionary Game Models

In contrast to dynamic and metabolic models, evolutionary game models focus on describing strategic decision-making of interacting agents and successfulness of their strategies (Figure 3c)(West, Griffin, Gardner, & Diggle, 2006). Rules of the evolutionary game define the payout that each player receives for every possible combination of strategies (e.g. cooperate, cheat). Each player's payout represents the individual's fitness, and the highest value "wins" the game. For example, microbial phenotypes can often be described as altruistic (A) or selfish (S); evolutionary games can model how such behaviors arise(West et al., 2006). While we would assume that selfish exploitation of the environment may be a winning strategy, the natural world is

paradoxically filled with organisms that exhibit cooperative behavior(Sachs, Mueller, Wilcox, & Bull, 2004). For microbial communities, the fitness of every individual in a population is determined by the net payout from all pairwise games with all other individuals. The initial proportion of individuals adopting a given strategy is an input for this model. These games are then iterated over time with a given strategy changing in abundance based on the fitness of individuals who hold the strategy compared to the average population fitness. As the marginal cost of cooperating and benefit of cheating lead to changing payouts, the two strategies will dynamically vary and affect the outcome of the game(Hauert, Michor, Nowak, & Doebeli, 2006). From these models, we find that populations that are dominated by altruists will often have a higher fitness than those dominated by selfish exploiters(Nowak, 2006).

For microbial ecosystems, evolutionary game theory models allow us to investigate how system parameters impact microbial interactions and dynamics of competing strategies. These models have been used to predict the evolutionary steady state of engineered yeast populations that exhibit altruistic or selfish strategies through the snowdrift game(Gore, Youk, & van Oudenaarden, 2009). In such a game, the altruists secrete an invertase enzyme that hydrolyses a polysaccharide to generate diffusible glucose products that are available to the entire population. The selfish individuals forgo the cost of secreting the enzyme, but rely on the glucose generated by the altruistic strains. Modulating the cost of cooperation resulted in shifts in the final population structure. Altruists dominated

when cost of cooperation was very low. Altruists and cheaters coexisted at median costs of cooperation, while cheaters dominated at high costs(Gore et al., 2009). To further take into account spatial structures, agent-based game models are used to restrict interactions to individuals in close physical proximity(Nadell, Foster, & Xavier, 2010). Clusters of cells that exhibit cooperative strategies will derive more benefit due to spatial confinement, and thus will be further enriched in the population. These and other evolutionary game models(Nowak, 2006) will be important quantitative tools to guide ecosystem engineering.

1.3.4 Digital Evolution

Long-term bacterial evolution experiments have been used to track how phenotypes and genotypes change in a constantly selective environment(Conrad, Lewis, & Palsson, 2011). Similarly, *in silico* simulations of evolution have been developed (Figure 3d)(Wilke, Wang, Ofria, Lenski, & Adami, 2001). Earlier forms of these simulations derive from cellular automata approaches, such as the Game of Life(M. Gardner, 1970). Cells in the cellular automata live in a two-dimensional environment. Reproductive success or cell death is governed by the density and configuration of the local population. Discrete time steps are iterated over the population to simulate the process of life. A more sophisticated implementation of digital evolution, called Avida, has been described(Lenski, Ofria, Pennock, & Adami, 2003). Avida is inspired by an earlier system Tierra, in which digital organisms contain computer programs that

compete for Central Processing Units (CPUs) and access to memory in order to reproduce(Ray, 1992). In Avida, digital organisms have their own memory space and virtual CPUs to perform tasks(Ofria, 1998). Each digital organism has a circular “genome” composed of a collection of 26 possible discrete basic programs (Nand, IO, swap etc.) that are executed in series. When certain combinations of these basic programs are executed in the correct order, one of several logic operations is performed. Strains able to execute higher complexity operations are rewarded with more energy and therefore replicate faster. As cells replicate, mutations are introduced, which result in programs being added, removed, or moved. This leads to new operational capabilities. Because the history of each organism’s genotype and phenotype are chronicled, digital evolution models enable better understanding of how individuals traverse a fitness landscape as complex traits evolve. These artificial life models also enable the reversion of individual and combinations of mutations to study epistasis. Key conclusions(Lenski et al., 2003) reinforced by these models include: 1) deleterious mutations may be needed to develop complex traits; 2) even though complex traits are fragile to mutations, they fix in the population because they provide significant fitness benefit, and 3) development of complexity requires selection of traits with intermediate complexity to allow gradual transition through the fitness landscape. Since complex phenotypes are a hallmark feature of microbes, this framework will likely provide useful insights to improve engineering of ecosystems through digital simulations. These

approaches are now being extended to simulation population-level behavior(Chow, Wilke, Ofria, Lenski, & Adami, 2004; Yedid, Ofria, & Lenski, 2008).

1.4 Experimental Tools

Over the last decade, the field of microbial ecology has been swept by a wave of new technologies, significantly reshaping the traditional investigative approach. These advances have centered on key developments in microfabrication, high-throughput sequencing, genome engineering, and synthetic circuit design. These new methods allow for better *in vitro* and *in vivo* models, culture-independent identification and quantification of individual species across populations, and generation of targeted genotypes for functional studies (Figure 4). Forward engineering of synthetic microbial ecosystems will rely heavily on these techniques.

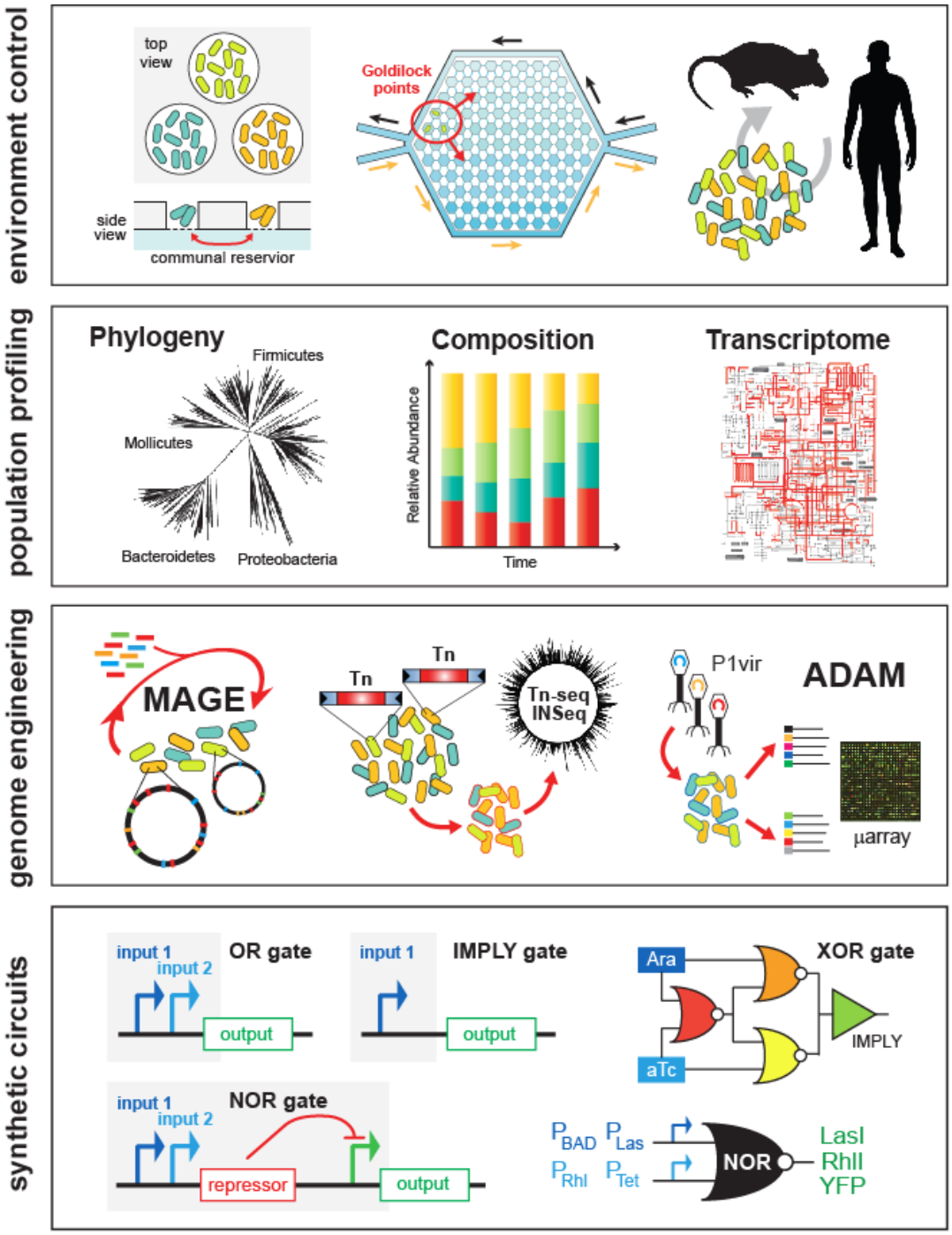

Figure 4: Experimental tools enable engineering of microbial ecosystems from the population level down to the DNA level. In vitro tools such microfluidics and microchambers or in vivo mice models enable precise control of the environment. High-throughput sequencing and transcriptomics enable parallel interrogation of phylogeny, composition, and gene expression of cell populations. Techniques such as multiplexed genome engineering and transposon mutagenesis enable forward engineering and

accelerated evolution of cell populations at the genetic level. New genetic circuitry and synthetic biology frameworks enable the development of multi-component genetic programs that are executed across populations of cells.

1.4.1 In vitro Models

Going beyond traditional cultivation techniques using petri dishes and culture flasks, advances in microfabrication and microfluidics have produced a variety of cheap lab-chip devices that can be used to cultivate and analyze microbes grown in massively parallel micron-sized chambers and channels(Balagadde, You, Hansen, Arnold, & Quake, 2005; Link, Jeong, & Georgiou, 2007). These devices are particularly useful for generating physicochemical conditions found in heterogeneous ecological niches to study behaviors such as quorum sensing or antibiotic susceptibility. For example, Zhang et al. developed a microfluidic chip that contained 1200 interconnected wells to probe the development of ciprofloxacin antibiotic resistance(Zhang, Lambert, et al., 2011). Local antibiotic gradients generated “Goldilocks points” in the microchamber where motile strains gathered and developed notable ciprofloxacin resistance (10 mg/ml) – 200 times the minimum inhibitory concentration. This phenomenon was not observed in the absence of such antibiotic gradients when grown in standard flasks as no resistance strains developed. This work highlights the importance of local heterogeneity in the evolution of microbial populations and development of antibiotic resistance.

Microfluidic chambers can also be used to study chemical signaling and nutritional cross-feeding between different microbes. Hyun et al. developed a

fluidic chip that contained arrays of spatially separated micro-wells with selectively permeable bottoms placed over a common liquid reservoir(Kim, Boedicker, Choi, & Ismagilov, 2008). Through size exclusion, metabolites could diffuse to neighboring wells while the bacteria producing them remained in each well. Using this system, the authors built a synthetic consortium of three bacteria, *Azotobacter vinelandii*, *Bacillus licheniformis*, and *Paenibacillus curdlanolyticus*, which normally do not grow together in nature. In a defined environment that is nitrogen and carbon depleted, and in the presence of antibiotics, the consortium exhibited reciprocal syntrophy because each species performed a specialized function that benefited the entire group. *A. vinelandii* fixed gaseous nitrogen into amino acids. *B. licheniformis* degraded the antibiotic penicillin. *P. curdlanolyticus* generated carbon sources needed by the consortium by degrading carboxymethyl-cellulose. In this co-culture, spatial structures and local interactions amongst the members defined the viability of the ecosystem. These interactions can be further elucidated at the single-cell level by using agarose tracks in channels that are the width of one cell(Balaban, Merrin, Chait, Kowalik, & Leibler, 2004; Moffitt, Lee, & Cluzel, 2012). Through optical microscopy, growth of individual cells by linear extension along the channel can be tracked over 40 generations. Syntrophic exchange between strains of *E. coli* auxotrophic for different amino acids enabled growth in separate parallel channels(Moffitt et al., 2012). Highlighting the importance of locality in syntrophic exchange, the co-culture growth rate was shown to decrease sharply when the distance between

complementary strains in neighboring channels increased by more than a few cell lengths.

In addition to microchambers and microchannels, microdroplet technology is also useful in probing interspecies interactions(Park, Kerner, Burns, & Lin, 2011). Groups of cells can be encapsulated in monodispersed aqueous-phase droplets using a T-junction microfluidic channel with an oil-phase. Through syntrophic cross-feeding, auxotrophic *E. coli* strains can grow in these microdroplets and be analyzed by microscopy(Park et al., 2011). These approaches will improve cultivation of new microbes by recapitulating microenvironments in which otherwise unculturable microbes can grow in the presence of metabolically compatible partners.

1.4.2 In vivo Models

Experimental models that recapitulate natural environments lend crucial insights into structure and function of microbial communities in their native habitats. Tractable live animal models, such as gnotobiotic germ-free (GF) mice, have been used extensively to investigate the relationship between the mammalian gut and the resident microbial community(Faith et al., 2010). Gnotobiotic mice can be inoculated with defined and sequenced microbes that are trackable to investigate processes of gut colonization, food metabolism, and community stability. In one such recent study, Faith et al. introduced 10 representative strains of the human microbiota into GF mice that are fed with

defined diets of macronutrients(Faith, McNulty, Rey, & Gordon, 2011). Four classes of foods were given to mice: proteins, fats, polysaccharides, and sugars. The 10-member microbial consortium was tracked by analysis of fecal samples after transition to different diets. The researchers found that a simple linear model could predict over 60% of the variation in species abundance due to diet perturbations. The use of synthetic microbial communities in live animal models provides a feasible way to untangle the web of complex interactions that may go on in the population. Furthermore, *in vivo* mice models are amenable to genetic modifications to produce important disease phenotypes such as ob/ob(Ley et al., 2005) or Tlr2(-/-)(Kellermayer et al., 2011), which can be used to tease out host-microbe interactions.

Simple evolutionary models of antibiotic antagonism, such as the classic non-transitive rock–paper–scissors (RPS) game, have also been demonstrated by studying engineered *E. coli* strains in GF-mice. Kirkup and Riley(Kirkup & Riley, 2004) used three types of strains: one that produces bactericidal colcins (P) that preferentially kill off sensitive strains (S) versus resistant strains (R). Sensitive strains can outcompete resistant strains, which in turn can outcompete colcin-producing strains. GF-mice associated with the microbial consortium showed cycling between the three phenotypes, which illustrated the RPS model and the *in vivo* role of colcin as an antibiotic. More interestingly this synthetic consortium model suggests that antibiotic-mediated antagonism can serve to promote microbial diversity in the mammalian gut.

1.4.3 Population Quantification Techniques

Precipitous reduction in cost and exponential growth in throughput of next-generation DNA sequencing technologies have revolutionized molecular biology(Medini et al., 2008). Sequencing has been used extensively for cataloging the composition, abundance, and metabolic potential of microbes from a variety of natural environments such as soil(Mackelprang et al., 2011), ocean(Biers, Sun, & Howard, 2009), acid mines(Simmons et al., 2008), and the human body(Peterson et al., 2009). Molecular barcoding allows large numbers of samples to be multiplexed and can be combined with time-series measurements to capture temporal changes across the entire population(Hamady & Knight, 2009; MacLean, Jones, & Studholme, 2009). Furthermore, transcriptome sequencing methods such as RNA-seq allow us to measure detailed transcriptional profiles of consortium members under different environmental conditions(Turnbaugh et al., 2010). Resequencing genomes from long-term evolution studies have also increased in popularity(Conrad et al., 2011). These investigations help to identify genetic mutations that arise due to adaptation to new environments(Chou, Chiu, Delaney, Segre, & Marx, 2011; Khan, Dinh, Schneider, Lenski, & Cooper, 2011) and help to reveal genetic heterogeneity within the population(Tenaillon et al., 2012). Goodarzi et al. developed the genetic footprinting technique, array-based discovery of adaptive mutations (ADAM), which enabled selective identification of mutations that provide a competitive advantage within a cell population(Goodarzi, Hottes, & Tavazoie,

2009). Combining sequencing and functional measurements, this method reconstructs beneficial phenotypes to increase the scope of adaptive lab evolution studies and enhance understanding of genetic interactions in complex populations.

1.4.4 Genome engineering

Construction and engineering of sophisticated synthetic ecosystems require facile modification of microbial genomes. Transposable elements have long been used as an efficient way to produce mutants of various phenotypes by random insertion into the genome(F. Hayes, 2003). Libraries of such transposon-mutated strains diverge in genotype and phenotype, but when pooled together can begin to resemble a microbial consortium. Using high-throughput DNA sequencing, large libraries of transposon mutants can be interrogated efficient. Goodman et al. combined the use of transposon mutagenesis, high-throughput sequencing and gnotobiotic mice in a technique called Insertion Sequencing (IN-Seq) to probe the function of *Bacteroides thetaiotaomicron* in the mouse gut(Goodman et al., 2009). Populations of *B. thetaiotaomicron* cells that were mutated by Himar1 mariner transposons were assessed by Illumina sequencing. The modified Himar1 inverted repeat sites contained MmeI-compatible sequences. Upon MmeI digestion of genomic DNA from the mutant population, high-throughput sequencing can be used to determine two 18-bp pairwise genomic fragments that correspond to the transposon insertion. Abundance

levels of each mutant can be tracked and distinguished from one another, as well as from defined microbes in other phylum such as Firmicutes or Actinobacteria. Other similar techniques for high-throughput transposon sequencing include Tn-seq(van Opijnen, Bodi, & Camilli, 2009), high-throughput insertion tracking by deep sequencing (HITS)(Gawronski, Wong, Giannoukos, Ward, & Akerley, 2009) and transposon-directed insertion-site sequencing (TraDIS)(Langridge et al., 2009) have also been developed.

Often, engineering members of a synthetic consortium requires precise genetic manipulation of the genome instead of random mutagenesis. Recent advances in oligo-mediated genomic engineering such as Multiplex Automated Genome Engineering (MAGE) has enabled efficient, parallel, and site-specific modification of genomes across many target sites(H. Wang, Kim, HB, Cong, L, Bang, D, Church GM., 2012; H. H. Wang & Church, 2011; H. H. Wang et al., 2009). By using pools of oligos, MAGE can generate genetic diversity in the population at a rate of $4.3x10^9$ modified bases per day, which enables combinatorial generation of divergent and complementary phenotypes within population clades(H. H. Wang et al., 2009). MAGE relies on the transformation of small chemically synthesized oligonucleotides (~50–90 bp) into the genome that then proceed to integrate into the chromosome during replication in an Okazaki-like fashion. Single-stranded DNA binding proteins and recombinases greatly facilitate this process and are often found as a part of viral integration machinery(Datta, Costantino, Zhou, & Court, 2008). Rapid generation of cells

that exhibit a variety of physiologies is not only feasible but can be automated. Therefore, these approaches are crucial to the construction of viable and stable synthetic communities. Oligo-mediated genomic engineering has shown promise in a variety of organisms including *Escherichia coli*(Ellis, Yu, DiTizio, & Court, 2001), *Pseudomonas syringae*(Swingle, Bao, Markel, Chambers, & Cartinhour, 2010), *Pantoea ananatis*(Katashkina et al., 2009), and other gram-negative bacteria(Swingle, Markel, et al., 2010), as well as *Mycobacterium tuberculosis*(van Kessel & Hatfull, 2007), lactic acid bacteria(van Pijkeren & Britton, 2012), and yeast(Kow, Bao, Reeves, Jinks-Robertson, & Crouse, 2007).

1.4.5 Synthetic Computing Circuits

Construction of genetic circuits that perform computational operations has been a long-standing goal in synthetic biology(Lu, Khalil, & Collins, 2009). Recent advances in genetic circuit design have now been extended to libraries of cells, which can be modularly combined to perform basic logic functions. Earlier work demonstrated that population-level behavior can be programmed using feedback genetic circuits and quorum sensing molecules(Brenner, Karig, Weiss, & Arnold, 2007; You, Cox, Weiss, & Arnold, 2004) but needed precise population-synchronization for robust behavior(Danino, Mondragon-Palomino, Tsimring, & Hasty, 2010). More recently, two groups developed multicellular computing systems(Regot et al., 2011; Tamsir, Tabor, & Voigt, 2011). Regot et al. constructed a library of engineered yeast cell-types that could sense different

extracellular input signals such as NaCl, doxycycline, galactose, oestradiol and produce chemical ‘wiring molecules’ such as pheromones to communicate with one another(Regot et al., 2011). These cell-types were made into AND and inverted IMPLIES logic functions to implement Boolean operations. For example, Cell 1 when presented with an input such as NaCl, will produce the wire molecule, pheromone, which is received by Cell 2. Cell 2 will produce a detectable fluorescence output only when it senses the pheromone and a second input such as oestradiol. The NaCl AND oestradiol operation is achieved with this two-cell implementation. By combining different cell-types, the authors generated a variety of logic gates (AND, NOR, OR, NAND, XNOR, XOR). More impressively, complex circuits including a multiplexer and a 1-bit adder with carry were built using additional chemical wires and cell-types. Based on a similar design scheme, Tasmir et al. constructed libraries of *E. coli* cells with simple NOR logic gates and connected them using quorum sensing molecules(Tamsir et al., 2011). The NOR gate was built using two tandem promoters that served as orthogonal inputs to drive the transcription of a repressor element. This simple implementation was used to build more complex circuits, which the authors demonstrated by performing logic operations on solid plates with different spatially defined colony types(Tamsir et al., 2011). These results support the notion that cellular consortia may be used to perform complex tasks more efficiently than single-cell implementations, further advocating the development of synthetic consortia as a platform technology.

1.5 Applications of Synthetic Consortia

Microbial consortia can potentially be programmed to perform useful tasks in both natural and artificial environments at spatial and temporal scales well beyond the capabilities of any individual member. Numerous applications may warrant such systems, ranging both in sophistication and in scale. Engineered microbes have long been used for industrial production of chemicals and pharmaceuticals(Alper & Stephanopoulos, 2009). These reactions tend to occur in fermentation chambers using genetically identical strains. All multi-step reactions need to be carried out intracellularly or would require separate fermentation pipelines. For complex feedstocks such as cellulosic biomass, single-strain fermentation reactions are unlikely to suffice. On the other hand multi-species communities can degrade these complex substrates efficiently(Kato, Haruta, Cui, Ishii, & Igarashi, 2005). Thus, future microbial fermentation systems are likely to shift to more heterogeneous population of engineered strains with diversified metabolic capabilities(Shong, Jimenez Diaz, & Collins, 2012).

Engineered consortia can be designed to degrade complex feedstock while simultaneously producing valued products. Using a symbiotic co-culture of engineered yeasts and *Actinotalea fermentans*, a cellulolytic bacterium, Bayer et al. were able to convert unprocessed switchgrass, corn stover, sugar cane bagasse, and poplar into methyl halide, a biofuel precursor(Bayer et al., 2009). *A. fermentans* fermented cellulose to acetate and ethanol, but its growth was

inhibited by these toxic waste products. However, engineered yeast was used to reduce acetate level by utilizing it for energy to produce methyl halide through heterologous expression of a methyl halide transferase. Thus, interdependence was established between the two strains to alleviate growth inhibition toward production of a biofuel. This type of division of labor is a powerful approach for processing complex substrates – a strategy commonly adopted in natural microbial consortia(Warnecke et al., 2007).

Applications in coordinated toxin detection and bioremediation may also benefit from synthetic consortia. By engineering auto-synchronization in populations of oscillating cells, Prindle et al. developed a liquid crystal display (LCD)-like macroscopic clock that could sense arsenic concentrations and respond by changing the oscillatory period(Prindle et al., 2012). The researchers nested two modes of cell signaling to expand the scale at which coordinated events manifest across the population. Slower local synchronization proceeded via a well-established quorum sensing genetic circuit to form colonies called “biopixels.” Arrays of these small colonies were synchronized across a large scale with a weaker but faster redox signaling system using hydrogen peroxide. Using an extra positive-feedback element that was linked to an arsenic-responsive promoter, the oscillatory system became a macroscopic arsenic biosensor that fluoresced at different periods depending on the arsenic concentration. By combining the two modes of cellular communication across thousands of microwell channels, the authors developed a proof-of-principle

biochip that may potentially be used as a handheld arsenic detector.

For applications in medical therapeutics, engineered microbial gut consortia will likely be an important area of development. Recent studies have highlighted the important role of human-associated microbial communities in maintaining health and causing diseases(Dethlefsen, McFall-Ngai, & Relman, 2007; Nicholson, Holmes, & Wilson, 2005; Turnbaugh et al., 2007), especially in the gastrointestinal (GI) tract where food and drugs are metabolized. The gut environment is home to the highest density of microbes in the body (up to 10^{11} cells/gram) and irregularities in the microbial composition are linked to diseases including Crohn's(Manichanh et al., 2006; Sokol et al., 2009), inflammatory bowel disease(Nell, Suerbaum, & Josenhans, 2010), obesity(Turnbaugh et al., 2009), diabetes(Giongo et al., 2011), infections(Walk & Young, 2008), and maldigestion(He et al., 2008). Traditional therapeutic strategies using probiotics have failed to generate consistent results largely due to a lack of understanding for the design principles needed to maintain engineered microbes *in vivo*. New approaches in synthetic consortia engineering will likely succeed where previous attempts have failed. Few successes in this area are already encouraging. Steidler et al. engineered an orally administered *Lactococcus lactis* strain that excreted human interleukin-10 in the GI tract(Steidler et al., 2000). This engineered probiotic strain reduced the degree of induced colitis in mice models, paving the way for human clinical trials for IBD(Steidler, Rottiers, & Coulie, 2009). Saeid et al. showed that engineered *E. coli* could detect the human pathogen

Pseudomonas aeruginosa via a quorum sensing pathway(Saeidi et al., 2011). *P. aeruginosa* often colonize the respiratory and GI tracts, leading to chronic and fatal diseases. Upon pathogen detection, the programmed *E. coli* self-lyse and release pyocin, a narrow-spectrum bacteriocin that kills *P. aeruginosa*. Future applications of human-microbiome engineering may include enhancing catabolism of troublesome but common metabolites (e.g. lactose and gluten), precise microbial modulation of the immune system, and removal of multi-drug resistant pathogens by selective toxin release.

We outline different applications of synthetic microbial communities to highlight their potential in improving areas of biosensing, biosynthesis, and biodegradation where the capabilities of homogeneous populations of genetically identical cells are insufficient (Figure 5).

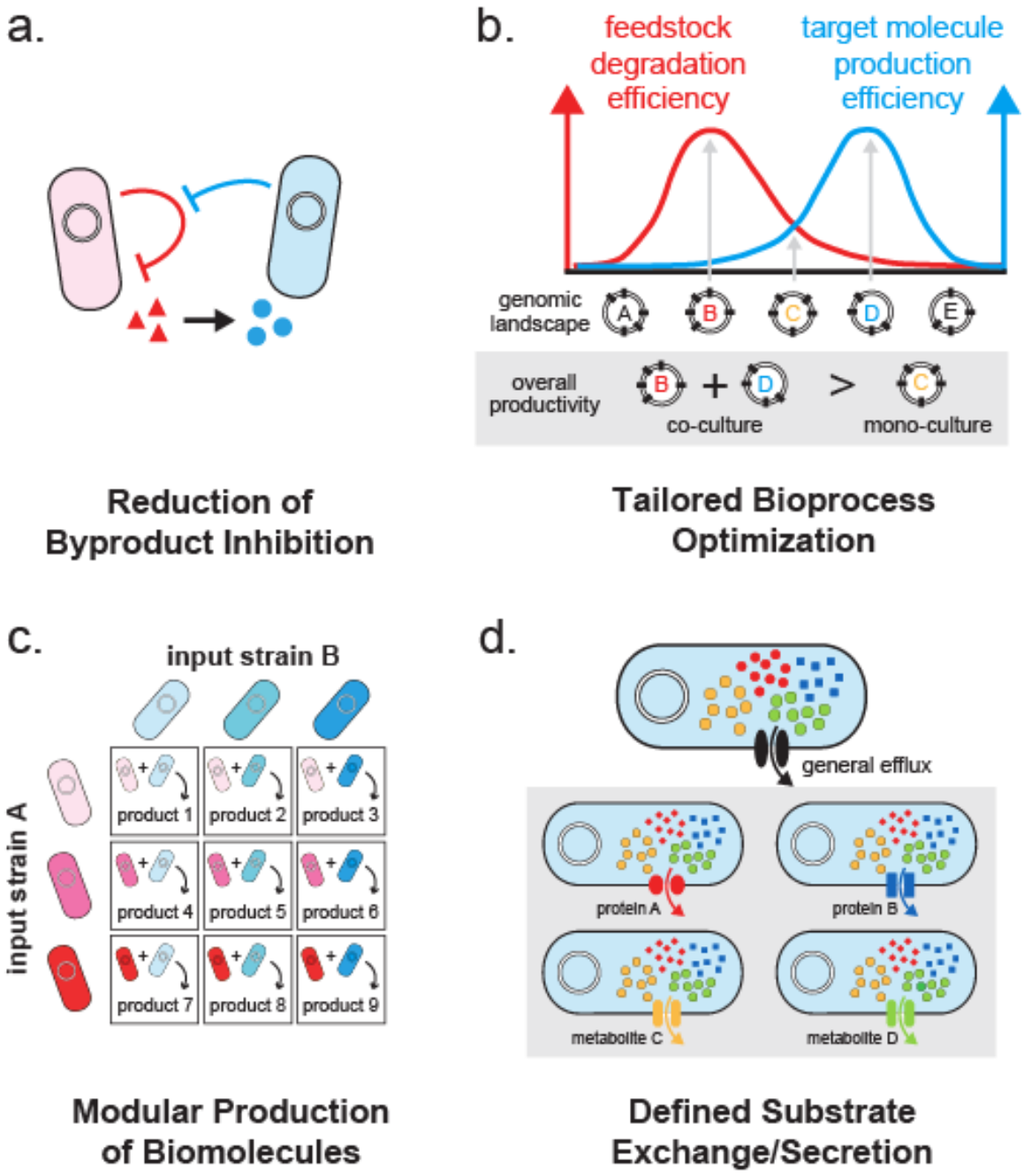

Figure 5: Engineering improvements for synthetic consortia. (a) A synthetic consortium can be designed to reduce by-product inhibition that accompanies over-accumulation of a toxic intermediate. (b) A co-culture bioprocessing strategy can be used where multiple steps are separately optimized in different cells to maximize overall productivity. (c) Modular assembly of synthetic consortia using common intermediate metabolites enables reprogrammable bioprospecting. (d) Production and secretion of multiple metabolites may saturate general cellular machinery. Specific metabolite export across different cell-types may increase productivity.

1.5.1 Biosensing

Abilities to sense diverse environmental signals and actuate appropriate responses are necessary and key features of engineered microbial communities. For example, autodetection of changes among networks of gut microbes in the intestine would allow for real-time monitoring and pinpointed responses to alarming events such as infections or toxins. Such capabilities would present a marked improvement over current monitoring strategies where symptoms are only recognized once an infection has fully developed and treatment requires indiscriminant depletion of the native community using antibiotics. These population-level behaviors are only now been demonstrated using synthetic communication circuits with quorum sensing modules. Non-pathogenic *Escherichia coli* have been engineered to recognize specific QS molecules diffusing from virulent strains of *Vibrio* (Duan & March, 2010) or *Pseudomonas* (Saeidi et al., 2011). Upon detection, pathogen-specific antimicrobial proteins or compounds are released, resulting in 99% reduction in the pathogen load (Saeidi et al., 2011).

Synthetic consortia can also be designed to detect and respond to other compounds to regulate programmed behaviors. Consortia growth rate and relative abundances of different members can be tuned in response to the environment (Hu, Du, Zou, & Yuan, 2010; Kerner, Park, Williams, & Lin, 2012). These approaches can be used to engineer biofilms to alter its physical architecture and membership composition to optimize bioprocesses that rely on

spatially associated communities (Shong et al., 2012). Engineered communities can also be used not only to microscopically sense low-level metabolites but also to amplify the signal for macroscopic detection. Building on an oscillatory fluorescence-generating circuit, Prindle *et al* synchronized local and global sensing mechanisms to generate periodically synchronized signals that changed in response to arsenic concentration, thereby generating a macroscopic biosensor using populations of cells (Prindle et al., 2012). Further demonstrations of synthetic consortia for biosensing applications are needed.

1.5.2 Biodegradation

Microbial communities naturally degrade various compounds into nutrients to sustain metabolism. Synthetic communities are increasingly being used to degrade xenobiotic and recalcitrant compounds. Similarly to the process of synthesis, degradation can be improved through careful engineering of organisms with desired functionalities that may be modular and complementary in physiology, resulting in overall improvement in performance of the community. The sequestration of undesirable compounds or pollutants can be augmented by reducing inhibition of cell growth that result from accumulation of inhibitory intermediates. For example, Li et al demonstrated that an engineered *E. coli* and *Ochrobactrum* consortia can enhance the degradation of methyl parathion (MP), an insecticide and toxin, through removal of the growth inhibitory intermediate p-nitorphenol (PNP), resulting in 98% MP removal (L. Li et al., 2008). Degradation

profiles can further be improved using engineered microbes that supplement a consortium with limiting metabolites, such as biotin, thiamine, cobalamine and siderophores that help facilitate growth and bioconversion. Examples of these communities of cyanobacteria or microalgae with bacteria have been documented to greatly improved degradation of hydrocarbons in oil spills (Tang et al., 2010).

Microbial communities play a significant role in digestion and metabolism of foods in the mammalian gut and its dysfunction may lead to diseases of maldigestion (Ley et al., 2005). Perturbation of synthetic communities of gut microbes in gnotobiotic mice using altered diets demonstrated that digestive capabilities may be a viable avenue of forward engineering through synthetic biology (Faith et al., 2011). For example, gut communities that additionally carry methanotrophic Archaea can lead to overall increase in microbial metabolism though removal of inhibitory levels of hydrogen that are otherwise generated. This has the direct effect of increased degradation of nutrient into absorbable nutrition, leading to elevated nutritional uptake by the host and when in excess can cause obesity (Ley, Turnbaugh, Klein, & Gordon, 2006). Engineering and altering the degradation capacity of gut-associated microbial communities will likely be an important avenue to develop for synthetic microbial ecosystems.

1.5.3 Biosynthesis

Synthesis of new compounds or existing ones using safer and better approaches is critically needed – a task well suited for engineered consortia. More efficient use of otherwise waste feedstocks as input materials into fermentation bioreactors is a highly desirable objective. Many of these materials, such as cellulosic biomass, are complex feedstocks that are not well-suited for current bioproduction pipelines. Use of engineered communities (Zhang, He, et al., 2011) presents a better solution than current monoculture production approaches as excretion of different cellulases from different strains can improve degradation of complex cellulose polymers into smaller monomers . Additionally, cells that optimally excrete these cellulases may not be well-suited for bioproduction due to inherent metabolic costs. Shin et al demonstrated the advantage of using a synthetic consortium with a divided labor structure for ethanol production from hemicellulosic feedstock (Shin, McClendon, Vo, & Chen, 2010). Two *E. coli* strains were co-cultured; one genetically optimized for cellulase production and excretion, and the other for utilization of the digested substrate for conversion to ethanol. Ethanol production reached 70% of theoretical yield in the co-culture compared to 26-28% with single strains. The use of orthologous secretion systems can further improve specificity of secretion and improve efficiency of secretion by limiting saturation through dividing different processes across multiple strains (Eiteman, Lee, & Altman, 2008).

Medical applications of engineered microbes include *in situ* biosynthesis

and excretion of therapeutic compounds such as cytokines and immunomodulating proteins at the site of injury (Wells & Mercenier, 2008). Introduction of non-pathogenic engineered *Lactococcus lactis* that can produce interleukin 10 in the mouse gut ameliorated autoinflammatory diseases such as colitis, Crohn's and inflammatory bowel disease (Steidler et al., 2009). Improving the engineering of complex microbial ecosystems to stably maintain desired strains in challenging environments such as the human gastrointestinal tract will increase the longevity and effectiveness of these therapies.

Microbial consortia with modular architecture may enable more programmable reconfiguration of biosynthesis objectives and optimization conditions. Degradation strains and production strains can be combined modularly using shared common intermediate metabolites to generate useful products such as biofuels or biomaterials. Metabolic interactions, inhibit or beneficial, across the microbial networks must be carefully engineered (Kato et al., 2005; Kato, Haruta, Cui, Ishii, & Igarashi, 2008). For example, by-product inhibition occurs when growth or productivity of a strain is impaired by the compound it produces (Bizukojc, Dietz, Sun, & Zeng, 2010). For example, *Actinotalea fermentan* can efficiently process cellobiose feedstocks (switchgrass, corn stover, bagasse, etc.) into acetate, but its growth rate is significantly impaired by even moderate levels of acetate. Bayer *et al.* demonstrate that acetate by-product inhibition of *A. fermentan* can be removed by addition of an engineered *Saccharomyces cerevisiae* strain which utilizes acetate for growth

(Bayer et al., 2009). The yeast is then engineered to produce methyl halides, which is a useful biofuel precursor. A 12,000-fold improvement was achieved using this approach compared to levels from single culture bioreactors.

Synthetic consortia additionally enable membrane-bound enzyme complexes such as those for engineering H_2 production in *E. coli* to be maximally utilized (Waks & Silver, 2009). Integration of strain into an engineered consortium through metabolic cross-feeding is a more modular approach that allows optimization of partitioned functions such as protein engineering of membrane-bound complexes. Similarly, membrane-associated extracellular mini-cellulosomes that spatially co-localize can improve reaction rate and efficiency to improve performance of synthetic consortia (Tsai, Goyal, & Chen, 2010). A four-member cellulosome-generating yeast consortia was recently demonstrated for ethanol production, reaching 87% of theoretical yield – a 3-fold increase over monoculture strain that expressed all four enzymes(Goyal, Tsai, Madan, DaSilva, & Chen, 2011). Thus, utilizing synthetic consortia for modular and programmable biosynthesis of useful compounds remain very promising.

1.6 Project Overview

This project aims to build upon many of these great advances to our understanding of the mechanism driving the complex behaviors and capabilities of microbial communities. The main feature of this work will be investigating how metabolic dependencies implemented at the individual level can drive behaviors

of the community. By mining the vast resources of publicly available sequencing data we will investigate the prevalence of biosynthetic capabilities, and lack there of, in hopes of identifying signals of naturally defined evolutionary rules on metabolic streamlining. We will then generate engineered dependencies to study increasingly complex defined cross-feeding interactions. The scalability of computational models parameterized from simple pairwise system information will be tested against higher order interactions. We have no illusions that these models will be perfect predictors of more complex behaviors. The assumptions they are built on will break down as epistatic and evolutionary effects accumulate. However, we will also strive to study the rich information that can be derived from evolutionary modifications of these cross-feeding systems to improve our models and enable more robust engineering of microbial ecosystems. Leveraging the power of DNA synthesis, a large diversity of amino acid secretion phenotypes will be explored. In this case, rather than waiting for novel mutations to appear *de novo*, the dynamics of the community will studies to improve our mechanistic understanding of the paradoxical maintenance of cooperativity in these ecosystems. Finally, metabolic dependencies beyond the 20 amino acids and even beyond naturally produced metabolites will be explored as alternate means of controlling these communities.

2 Auxotrophies and Cross-feeding in natural ecosystems

2.1 Overview of Bioinformatic Analysis

On the population level, just as for individual cells, biosynthesis is optimized relative to cost and utility(Smith & Chapman, 2010). Redundant or unnecessary biosynthetic pathways may reduce the metabolic efficiency of the population and are likely removed through Darwinian evolution(McCutcheon & Moran, 2012). Using comparative genomics, we can computationally predict the biosynthetic capabilities of organisms that have fully sequenced genomes. Presence or absence of genes needed for biosynthesis of essential metabolites can be tabulated. Using the Integrated Microbial Genomes (IMG) database of sequenced organisms(Markowitz et al., 2012) (3062 Bacteria, 121 Archaea, 124 Eukarya as of 2011) and an algorithm for biosynthesis prediction, we discovered huge variation in biosynthetic capabilities for essential metabolites such as amino acids. The algorithm annotates an organism's biosynthetic capabilities based on sequence homology of its genome to genes in established databases(Hunter et al., 2011; Petersen, Brunak, von Heijne, & Nielsen, 2011; Punta et al., 2012; Tatusov, Koonin, & Lipman, 1997).

2.2 Results

2.2.1 Distribution of Biosynthetic Capabilities Across Kingdoms

When plotting a histogram of organisms that are capable of biosynthesizing zero to all 20 standard amino acids, we find a wide distribution (Figure 6). The Bacteria domain tends to have organisms that on average can

completely biosynthesize 7.9 out of 20 amino acids *de novo*. The average is 8.3 amino acids for Archaea and 4.1 for Eukarya. The histogram for Bacteria seems to be bimodal (Figure 6a), suggesting that further classification is needed. Organisms in the Archaea domain on average have a slightly higher biosynthetic range for amino acids. This perhaps is due to their more ancient origin as a domain. Unsurprisingly, organisms in the Eukarya domain appear to make fewer amino acids since they derive most essential amino acids from nutrient-rich diets. As a reference, humans can only make 10 out of the 20 amino acids.

For each amino acid, we can further analyze whether the full biosynthetic pathway is intact across different organisms (Figure 6b). We find that glutamic acid (E), glycine (G), and asparagine (N) tend be synthesized in most organisms while tyrosine (Y), phenylalanine (F), lysine (K), and histidine (H) tend to be made in few organisms. These trends appear to hold across Bacteria, Archaea, and Eukarya suggesting more universal processes at play. It is interesting to note that the more infrequently synthesized amino acids are also more costly to produce than those that are synthesized by most organisms, suggesting a level of cost-to-utility optimization(Akashi & Gojobori, 2002).

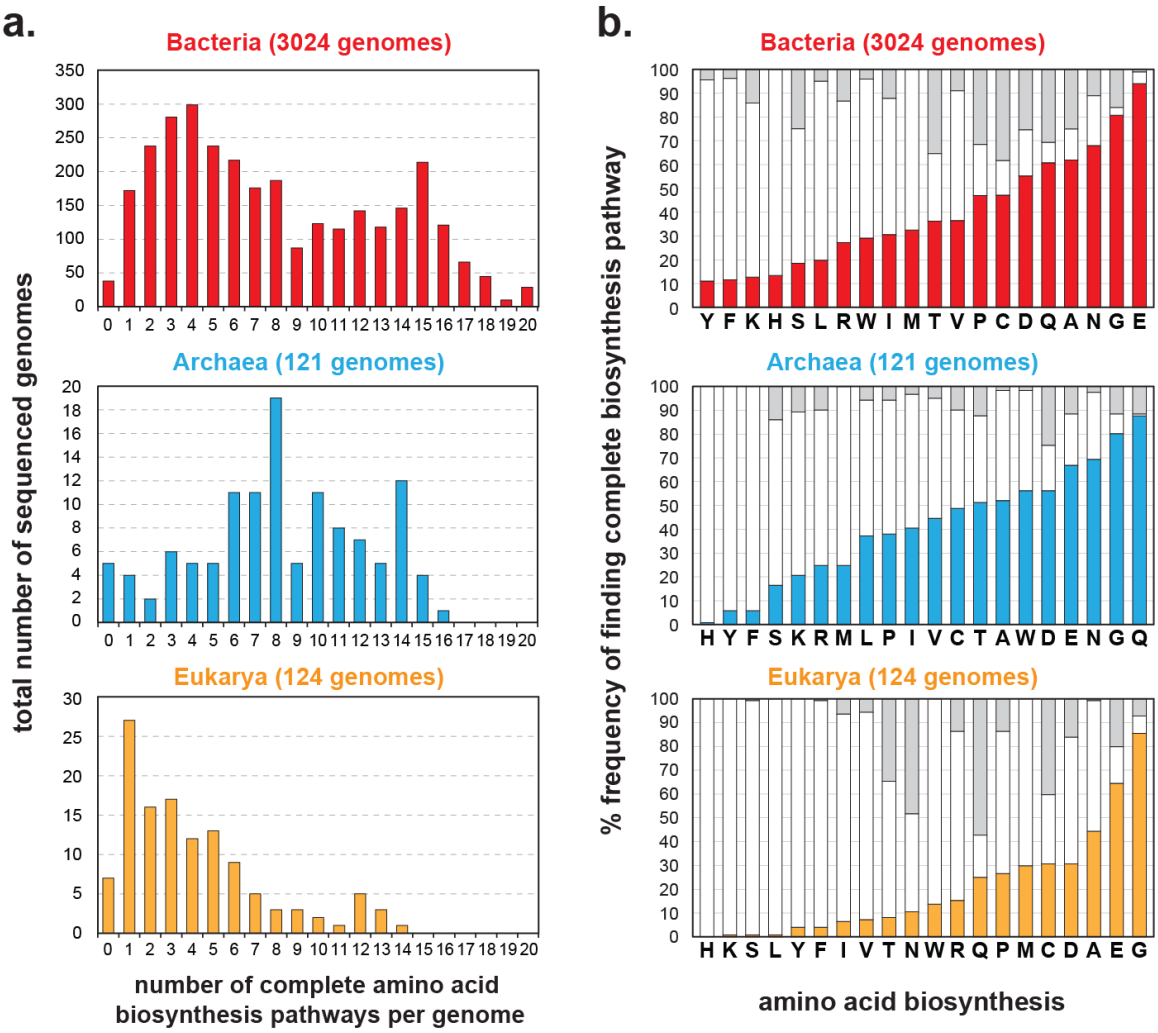

Figure 6: Diversity of amino acid biosynthetic capabilities across all sequenced organisms from the Integrated Microbial Genomes (IMG) database160, separated based on the three domains (Bacteria, red, top panel; Archaea, blue, middle panel; Eukarya, orange, bottom panel). (a.) Predicted frequencies at which species have the ability to synthesize zero to all 20 standard amino acids. (b.) For each amino acid, frequencies at which complete biosynthetic pathways are found across each domain are shown in solid colored bars (Bacteria, red, top panel; Archaea, blue, middle panel; Eukarya, orange, bottom panel). White bars indicate fractions in each domain where one or more biosynthetic gene is missing. Gray bars indicate unknown annotations.

2.2.2 Bacteria Specific Biosynthetic Capability Distribution

The general patterns arising from published studies of amino acid cross-feeding in *E. coli* led us to hypothesize that amino acid exchange may be an important property across many microbial communities in the natural biosphere. To evaluate this hypothesis, we compiled the frequencies at which bacterial genomes were predicted to contain complete and intact biosynthetic pathways for each of the 20 essential amino acids (see Methods). We analyzed 6120 sequenced genomes through the Integrated Microbial Genomes (IMG) database and pipeline (Markowitz et al., 2012) (Figure 7b). When we plotted the amino acid biosynthetic potential, we find that most bacteria are able to produce amino acids *E*, *G*, *N* and *Q*, while only a surprisingly small fraction of bacteria are able to produce amino acids *K*, *H*, *F* and *Y*. When the biosynthetic potential is plotted against the estimated metabolite cost to produce each amino acid, we find that the more expensive amino acids (*e.g. F*, *W*, *Y*) tend to be made less prevalently than inexpensive ones (*e.g*. *E*, *G*, *N*, *Q*) (Figure 7c). Other studies have found the costs associated with metabolically expensive processes (siderophore production, N fixation etc.(Church, Jenkins, Karl, & Zehr, 2005; Morris, Lenski, & Zinser, 2012)) are often shared among populations of bacteria. Additionally, the presence of exporters for these products and the fact that the lipid bilayer is up to 100 times more permeable to the expensive highly hydrophobic aromatic amino acids suggests high potential for exchange of these molecules(Chakrabarti & Deamer, 1992). Exchange of these goods provides ripe opportunity for loss of

function of these expensive biosynthesis capabilities(Church et al., 2005). While the genomic sequence distribution doesn't necessarily imply the same functional distribution in the biosphere, this sampling of bacterial genomes is most likely dominated by the most abundant organisms in the biosphere and we are only now beginning accumulate sequences for rare bacteria. Together these results suggest that the microbial biosphere has been optimized such that costly but essential resources (*i.e*. amino acids) are made by only a small fraction of the members and that while exchange may not be driving the evolutionary loss of these biosynthesis function, it is necessary for the phenomena to occur.

As many sequenced bacterial genomes come from closely-related organisms it may be important to normalize this data for phylogeny to avoid over-counting biosynthetic capabilities of related strains many times over. In fact, we found that the amino acid prototrophy distribution trend presented is quite robust to normalization at various phylogenetic levels. In figure 8a, we present the distribution normalized at the genus level derived by averaging the amino acid biosynthesis profiles across all individuals in each group. This results in reducing the 6120 species to 565 genera encompassing from 1 to a maximum of 94 (*Streptococcus*) individual species. The trend for prototrophy distribution normalized at this resolution is quite similar to that presented in figure 7b. Additionally we also found that that species of the same phylogenetic classification could have quite varied biosynthetic capabilities. Figure 8b presents the distribution of pairwise hamming distances for all biosynthetic profiles of the

most represented genus, *Streptococcus*. While many strains' amino acid biosynthesis profiles differ by less than 20%, there is a sizable fraction that have quite varied profiles within this closely related group.

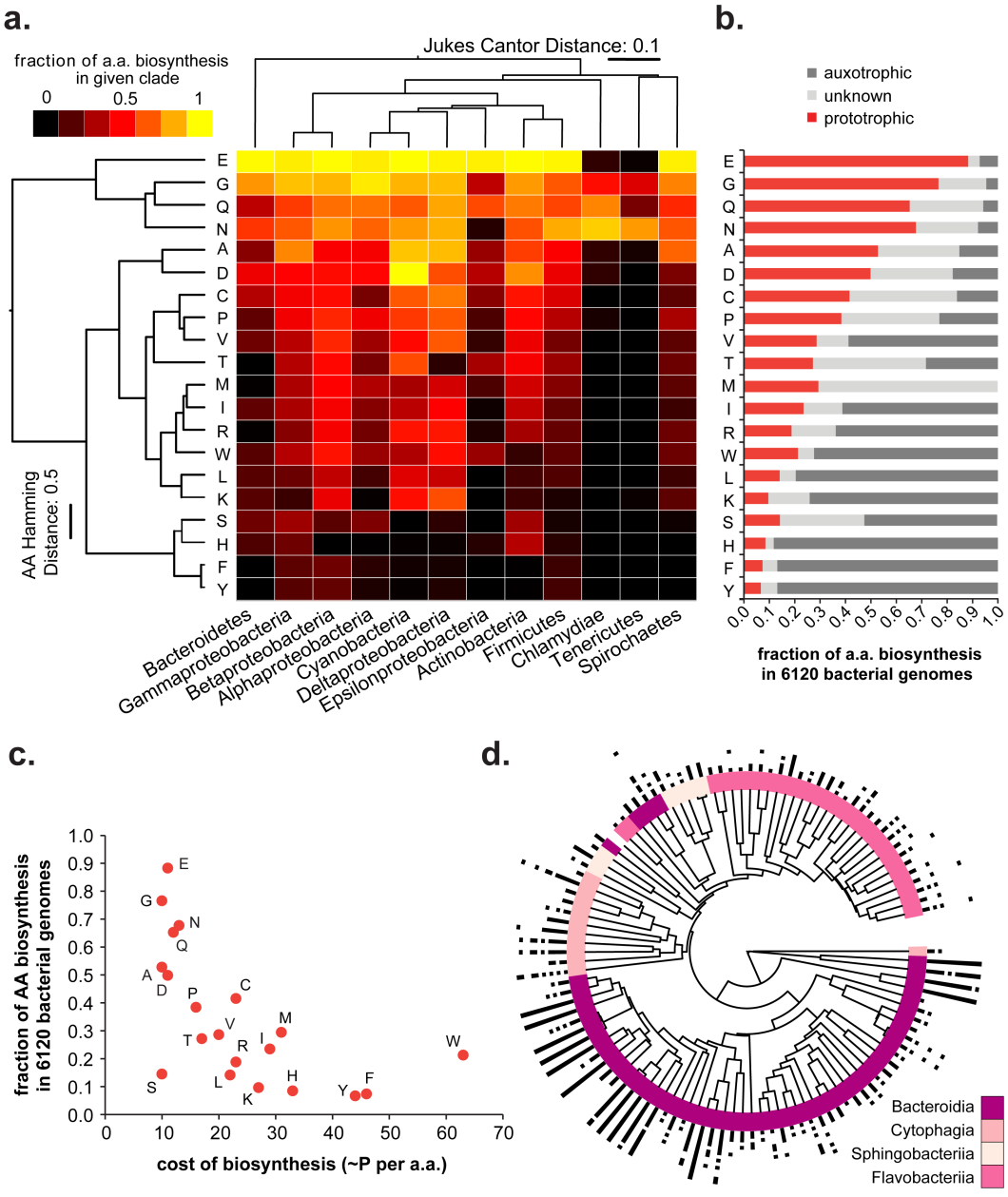

Figure 7: Amino acid biosynthesis in the microbiome. (a) Heat map of amino acid biosynthetic capabilities of the indicated phyla and classes. Prototrophy predictions for each amino acid are averaged within groups. Phyla and classes with fewer than 20 out of

the 2099 sequenced bacterial are excluded from the analysis. Value of 1/0 indicates all/none of the species in the group are prototrophic. Dendograms represent clustering of both phylogenetic distribution (based on median 16S sequence from each clade) and amino acid production profiles. Phylum/class leaf branches are not to scale to enhance higher order relationships. (b) Distribution of amino acid (AA) biosynthetic capability of 6120 sequenced bacteria. Red bars indicate complete pathway present. Incomplete or unknown pathways are denoted in black and grey bars respectively. (c) AA biosynthesis distribution plotted against metabolic cost of synthesizing each AA in terms of number of phosphates required. (d) Prototrophy distribution in Bacteroidetes Black rings indicate biosynthesis of each amino acid in increasing order of prevalence from inner to outer rings.(E,G,N,D,Q,C,A,S,V,I,P, K,L,W,H,M,R; T,F,Y not present)

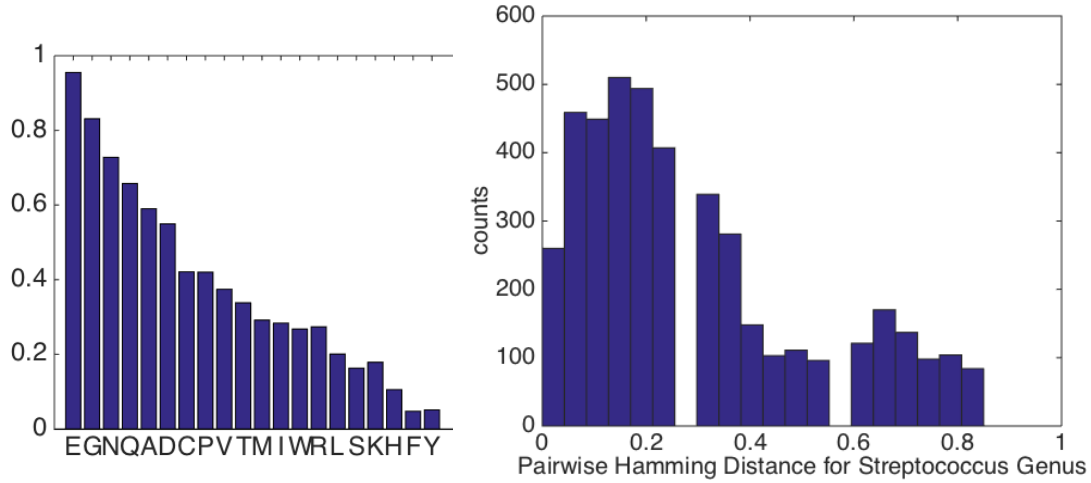

Figure 8: Genus level analysis of microbial biosynthesis. Left panel: Proportion of genera with predicted prototrophic capabilities for the indicated amino acids. Right panel: Pairwise Hamming distances for predicted amino acid biosynthetic capabilities of the most represented genus in sequenced bacteria, *Streptococcus*.

To understand the structure of amino acid biosynthesis at the biosphere scale, we mapped the prototrophy prediction distribution across a phylogenetic tree of ~2000 bacteria from diverse environments (Methods). Interestingly, we find that biosynthesis capabilities are highly structured when clustered at the phylum and class level (Figure 7a). For example, production of the most costly amino acids (*e.g. F,Y*) is concentrated in the closely related *Beta-* and *Gamma-proteobacteria* whereas production of lysine is highly enriched in *Deltaproteobacteria* and *Cyanobacteria*. Similarly, the closely related

intracellular parasitic phyla *Tenericutes* and *Chlamidiae* have similar biosynthesis profiles, making almost uniquely the amino acids *N, G* and *Q*. Similar structure is apparent when looking at biosynthesis distribution on the species level. Within the observed *Bacteroidetes* (Figure 7d), amino acid production profiles closely follow species relationships with certain clades predicted to have broad production abilities (several clades within the *Bacteroidia* class) while others have similar limited capabilities (*Flavobacteria*). Taken together, these results provide evidence suggesting that microbial genomes optimize their metabolic potential to reduce biosynthetic burden, and that microbes may be tactically leveraging the specialized biosynthetic capabilities of their neighbors while reciprocating through mutualistic trade of essential metabolites.

2.2.3 Serine Prototrophy Predictions

The low abundance of serine prototrophy among bacterial genomes predicted by the IMG pipeline raised some flags for a couple reasons. First, serine is readily synthesized by short pathways from 3-phosphoglycerate and from glycine via serine hydroxymethyl transferase. Second, serine is the most common amino acid in proteins and therefor it is somewhat in contradiction to biochemical intuition that most bacteria are serine auxotrophs. The IMG predictions only take into account serine biosynthesis from 3-phosphoglycerate. The limiting reaction, in terms of presence in the fewest bacterial genomes, seems to be the terminal reaction catalyzed by phosphoserine phosphatase (Table 1). As the standard

IMG prediction pipeline for serine biosynthesis ability doesn't include the pathway from glycine via serine hydroxymethyl transferase, this metabolic capability was assessed independently. To assess the frequency of prototrophy resulting from this biosynthesis pathway we first determine which sequenced bacteria are prototrophic for glycine. To do this the following logical test for enzyme presence predictions is implemented. The bacterial strain must be predicted to have:

1. acetaldehyde-lyase (via L-Threonine)

OR

2. NAD+ oxidoreductase **AND** glycine C-acetyltransferase (via Acetyl-CoA and L-Threonine)

OR

3. Glyoxylate aminotransferase (via L-alanine)

Finally, to be called as a serine prototroph (via glycine), a bacterial strain has to satisfy the above logical conditions **AND** be predicted to have glycine/serine hydroxymethyltransferase. Including these criteria along with those for serine biosynthesis via 3-phosphoglycerate results in an unchanged overall predicted serine prototrophy frequency of 15%. This most likely indicates limitations with the current characterization of serine biosynthesis predictions. The predictive model used will most likely improve over time.

Enzyme Name	EC #	Reaction	Percent of Sequenced Bacterial Genomes
D-3-phosphoglycerate dehydrogenase	EC 1.1.1.95	3-Phospho-D-glycerate + NAD+ <=> 3-Phosphonooxypyruvate + NADH + H+	16%
phosphoserine aminotransferase holoenzyme	EC 2.6.1.5	O-Phospho-L-serine + 2-Oxoglutarate <=> 3-Phosphonooxypyruvate + L-Glutamate	73%
phosphoserine phosphatase	EC 3.1.3.3	O-Phospho-L-serine + H2O <=> L-Serine + Orthophosphate	15%

Table 1: Three reactions evaluated by the IMG system for serine prototrophy predictions.

2.3 Conclusions

Here we provide evidence for widespread trends of metabolic cross-feeding based on comparative genomic analysis of amino acid biosynthesis across thousands of sequenced genomes. It is important to note that these computational estimates of prototrophy are likely to be at the low end. More accurate comparative genomic analysis using better-populated and more annotated databases will likely identify more biosynthetic genes. Nonetheless, the observation that most organisms cannot make all of their essential metabolites importantly highlights the interrelatedness of ecosystems. These results also highlight amino acids as a versatile set of metabolites whose exchange can enrich for consortium-level associations. Interdependencies can be engineered by exploiting biosynthetic configurations of these essential metabolites, which can further be tuned with transporters systems. These engineered communities present a framework for programming structures and

dynamics into microbial ecosystems and serve to improve our ability to engineer metabolism at the population-level.

2.4 Methods

2.4.1 Phylogenetic analysis

Predicted amino acid biosynthesis phenotype was extracted from the IMD database(Markowitz et al., 2012) for all available sequenced bacterial genomes. Phylogenetic linkage between these strains is determined using the corresponding aligned 16S sequences from the SILVA database(Yilmaz et al., 2014). Distances between sequences are determined with the standard Jukes-Cantor metric and hierarchal clustering performed with unweighted average distances. Phylogenetics trees were constructed using the iTol web application(Letunic & Bork, 2011). Amino biosynthesis profiles were clustered at the species level using Hamming distance and complete linkage. Biosynthesis profiles at the phylum and class level were determined by averaging prototrophy predictions for each amino acid across all represented species. Only the 12 clades with more than 20 species were clustered with Euclidian distance and complete linkage.

2.4.2 Biosynthesis predictions from the JGI IMG database

Amino acids biosynthetic predictions are pulled from the IMG database “phenotype function” system(Markowitz et al., 2012). This bioinformatic analysis

pipeline first predicts the enzymatic capabilities of a given organism. Next, logical testing is used for prototrophy calls to ensure that for each reaction step in amino acid biosynthesis from a common precursor (e.g. pyruvate for Leucine biosynthesis), there is a predicted enzyme that can catalyze the reaction. Similarly for auxotroph calls, an organism must be lacking enzymes catalyzing at least one of the steps in the biosynthesis pathway. Attempts have been made to take into account alternate biosynthesis pathways (e.g. Methionine synthesis via homocysteine or methanethiol intermediates), although undoubtedly not every exception will be captured and these computational approaches will not have 100% accuracy. Nonetheless, this is a powerful tool to make general trend observations. For details on prediction rule, please see the IMG website.

3 Syntrophic Exchange in Synthetic Microbial Communities

3.1 Background

3.1.1 Previous Findings in Metabolic Exchange

Microbes are abundantly found in almost every part of the world, living in communities that are diverse in many facets. While it is clear that cooperation and competition within microbial communities is central to their stability, maintenance, and longevity, there is limited knowledge about the general principles guiding the formation of these intricate systems. Understanding the underlying governing principles that shape a microbial community is key for microbial ecology, but is also crucial for engineering synthetic microbiomes for various biotechnological applications (Brenner, You, & Arnold, 2008; Mee & Wang, 2012; Shong et al., 2012). Numerous such examples have been recently described including the bioconversion of unprocessed cellulolytic feedstocks into biofuel isobutanol using fungal-bacterial communities (Minty et al., 2013) and biofuel precursor methyl halides using yeast-bacterial co-cultures (Bayer et al., 2009). Other emerging applications in biosensing and bioremediation against environmental toxins such as arsenic (Prindle et al., 2012) and pathogens such as *P. aeruginosa* and *V. cholera* have been demonstrated using engineered quorum sensing *E. coli* (Duan & March, 2010; Saeidi et al., 2011). These advances paint an exciting future for the development of sophisticated multi-species microbial communities to address pressing challenges and the crucial need to understand the basic principles that enables their design and

engineering.

An important process that governs the growth and composition of microbial ecosystems is the exchange of essential metabolites, known as metabolic cross-feeding. Entomological studies have elucidated on a case-by-case basis the importance of amino acids in natural inter-kingdom and inter-species exchange networks (McCutcheon & von Dohlen, 2011; Russell, Bouvaine, Newell, & Douglas, 2013; Wu et al., 2006). Recent comparative analyses of microbial genomes suggest that a significant proportion of all bacteria lack essential pathways for amino acid biosynthesis (Mee & Wang, 2012). These auxotrophic microbes thus require extracellular sources of amino acids for survival. Understanding amino acid (AA) exchange therefore presents an opportunity to gain new insights into basic principles in metabolic cross-feeding. Recently, several studies have used model systems of *S. cerevisiae* (Shou, Ram, & Vilar, 2007), *S. enterica* (Harcombe, 2010) and *E. coli* (Kerner et al., 2012; Pande et al., 2013; Wintermute & Silver, 2010b) to study syntrophic growth of amino acid auxotrophs in co-culture environments. Numerous quantitative models have also been developed to describe the behavior of these multispecies systems, including those that integrate dynamics (Bull & Harcombe, 2009; Estrela & Gudelj, 2010), metabolism (Klitgord & Segre, 2010; Mazumdar, Amar, & Segre, 2013; Stolyar et al., 2007), and spatial coordination (Nadell et al., 2010). While these efforts have led to an improved understanding of the dynamics of syntrophic pairs and the energetic and benefits of cooperativity in

these simple systems (Wintermute & Silver, 2010a), larger more complex syntrophic systems have yet to be explored.

3.1.2 This Work

Here, we use engineered *E. coli* mutants to study syntrophic cross-feeding, scaling to higher dimensional synthetic ecosystems of increasing sophistication. We first devised pairwise syntrophic communities that show essential and interesting dynamics that can be predicted by simple kinetic models. We then increased the complexity of the interaction in 3-member synthetic consortia involving cross-feeding of multiple metabolites. To further increase the complexity of our system, we devised a 14-member community to understand key drivers of population dynamics over short and evolutionary timescales. This large-scale and systematic effort represent an important foray into forward and reverse engineering synthetic microbial communities to gain key governing principles of microbial ecology and systems microbiology.

Our overall goal is to develop and understand a simple microbial model of metabolic cross-feeding that can be scaled in a tractable stepwise manner, towards reconstituting the complexity and dynamics exhibited by natural ecosystems (Figure 9a). To this end, we devised a series of syntrophic microbial communities of increasing diversity and complexity using the simple model bacterium *E. coli*. Our system is based on the syntrophic behavior of amino acid exchange between auxotrophic mutants to facilitate co-culture growth.

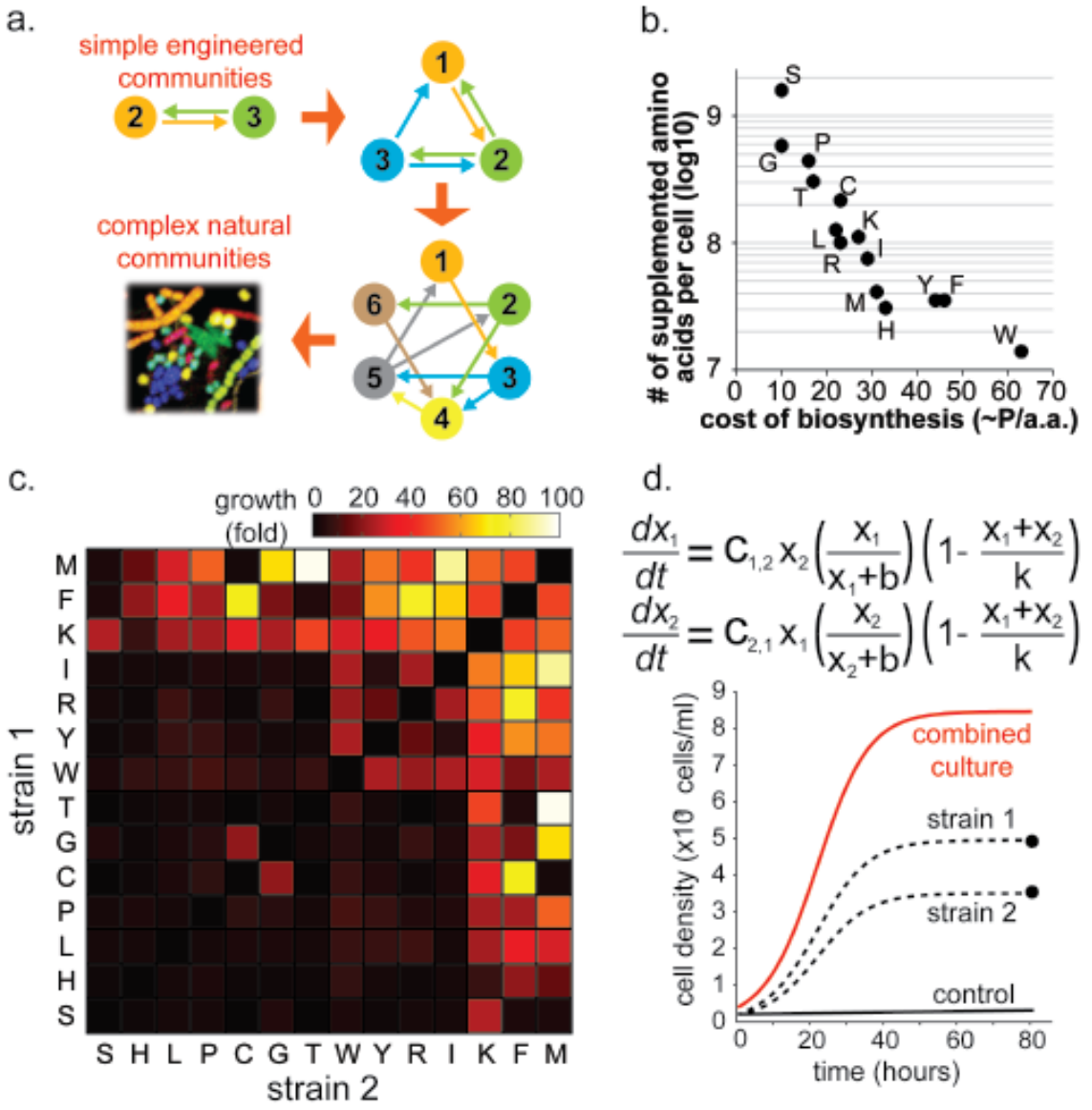

Figure 9: Metabolic cross-feeding in syntrophic communities. (a) An illustration of engineered syntrophic interactions between microbial communities of increasing complexity towards network hierarchies matching those of natural systems. (b) Relationship between number of supplemented AA needed to make one E. coli cell in log10 units versus biosynthetic cost to produce each AA. (c) Syntrophic growth yield after 84 hours between 14 single-KO auxotrophs (strain 1) and all pairwise combinations (strain 2). Color intensity indicated in the color bar denotes fold growth after 84 hours over initial population. (d) Simple two-equation dynamic model that captures the essential features of the pairwise consortium. Cooperativity coefficients C1,2 and C2,1 can be determined through the total co-culture growth curve (solid red line), the end point cell density of each strain (solid dots) and the simulated growth profile of each strain (black dotted lines). Control populations of only strain 1 or strain 2 (respectively ΔM and ΔF in this example) separately show no growth (solid black line).

3.2 Results

3.2.1 Amino Acid Utilization and Biosynthetic Cost

We first investigated the energetics involved in amino acid (AA) utilization and exchange. Starting from a prototrophic *E. coli* derivative MG1655, we generated 14 strains, each containing a gene knockout that lead to an auxotrophic phenotype of one of 14 essential amino acids (Methods). The remaining 6 essential amino acids were left out of our study because they either did not have single-gene targets that would render them auxotrophic or the resulting mutants carried significant growth defects even in richly supplemented media. By convention, we designate each auxotrophic strain by the amino acid they need – for example, the methionine Δ*metA* auxotroph is strain *M*. The 14 auxotrophs (*C*, *F*, *G*, *H*, *I*, *K*, *L*, *M*, *P*, *R*, *S*, *T*, *W*, *Y*) were confirmed to show no growth in M9-glucose minimal media after 4 days and growth only when supplemented with each AA needed. Using a microplate spectrophotometer, we performed kinetic growth curve analysis for each of the 14 auxotrophs grown in M9-glucose supplemented with varying initial AA levels. Under these AA-limiting conditions, an auxotrophic strain will grow exponentially until the AA supplementation is exhausted (Figure 10b). Saturating cell densities (*i.e.* carrying capacities) plotted against initial seeding AA levels show a strong linear relationship (Figure 10b) indicating that external AA levels can determine cell growth in a linear and predictable manner. We estimated the number of extracellular amino acids needed to generate a cell for each of the 14 AAs (Figure 10a, Table 2). The

estimated biosynthetic cost to produce each amino acid (Akashi & Gojobori, 2002) shows a strong inverse relationship with the amount of AA needed to produce a cell (Figure 9b), suggesting that the *E. coli* proteome has been optimized for amino acid usage to maximize energetic efficiency. Inexpensively produced amino acids (*e.g. S*, *G*, *P*) are used more frequently than expensive ones (*e.g*. *M*, *H*, aromatics). While previous computational approaches have suggested such relationships (Akashi & Gojobori, 2002), this is the first experimental measurement we are aware of regarding this important property. Since auxotrophic genotypes are prevalently observed in nature, our results suggest that many microbes may be subjected to this energetic optimization and that AA exchange may be an essential link guiding the evolution of these microbial ecosystems.

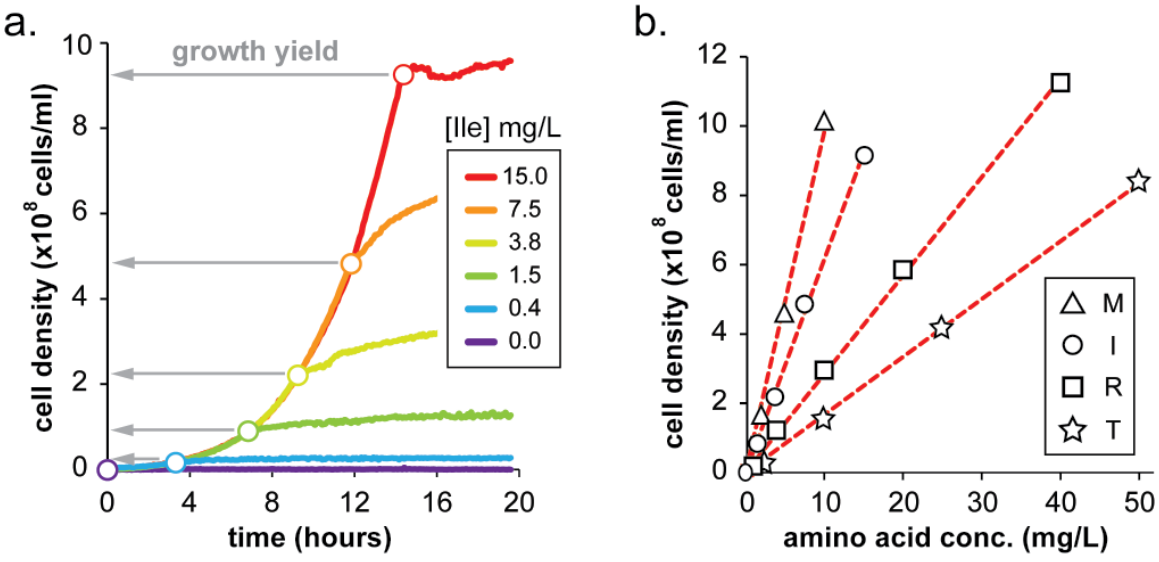

Figure 10: Calculation of amino acid utilization during growth. (a) Example growth curves of *E. coli* isoleucine auxotroph (Δ*ilvA*) in M9-glucose supplemented with varying initial concentrations of free L-isoleucine. Open circles denote the end of exponential phase resulting from supplemented amino acid exhaustion and marks final growth yield. (b)

Linear relationship between amounts of supplemented amino acid and the observed growth yield for auxotrophs M, I, R and T. Beyond the highest indicated amino acid concentrations the yield response saturated due to other metabolites becoming growth limiting. The slope of the linear relationship between amino acid concentration and cell density is used to calculate the amino acids required per auxotrophic cell in M9 media presented in table 2.

amino acid	KO target	extracellular aa needed per cell [# aa/cell]	biosynthetic cost [~p]
C	cysE	2.2E+08	23
F	pheA	3.7E+07	46
G	glyA	5.8E+08	10
H	hisB	3.1E+07	33
I	ilvA	7.5E+07	29
K	lysA	1.1E+08	27
L	leuB	1.3E+08	22
M	metA	4.1E+07	31
P	proA	4.4E+08	16
R	argA	1.0E+08	23
S	serA	1.6E+09	10
T	thrC	3.0E+08	17
W	trpC	1.5E+07	63
Y	tyrA	3.7E+07	44

Table 2: List of amino acid auxotrophs, the amount of supplemented amino acids needed to reproduce one cell, and the metabolic cost to biosynthesize each molecule of amino acid in units of phosphate bonds used. Amino acid requirement per cell is calculated using the slope from figure 10 as follows: $(1/(slope \times AA\ F.W.\ in\ g/mol \times 10^6)) \times 6.022 \times 10^{23}$.

3.2.2 Quantifying Pairwise Interactions

To more deeply investigate the properties associated with metabolite exchange of amino acids in microbial communities, we developed synthetic ecosystems using the characterized auxotrophic *E. coli*. Studying synthetic

communities using *E. coli* offers the benefits of robust and fast growing cells, tractable genetics, and well-developed *in silico* models while maintaining a standardized and reproducible genetic background. We first developed a 2-member syntrophic community composed of two different species from our 14 characterized auxotrophs. Each auxotroph is unable to grow in M9-glucose, but could potentially grow as a co-culture when paired with a different auxotroph. We probed all 91 possible pairwise syntrophic interactions in M9-glucose minimal media. In agreement with results from previous efforts (Wintermute & Silver, 2010b), we observed significant synergistic and cooperative growth in a subset of these pairwise co-cultures after 84 hours (Figure 9c). Pairings that involved cross-feeding of *M*, *F* or *K* were highly cooperative with most of the 14 partners, while *I*, *R*, *Y*, *W* had moderate cooperativity. *M*'s high cross-feeding productivity highlights the intuitive rule that partnering with direct biosynthesis precursors (*C* and *S* in this case) is unproductive. Cross-feeding between *T*, *G*, *C*, *P*, *L*, *H* or *S* was generally non-productive. Using quantitative PCR against the unique knock-out chromosomal junction in each auxotrophic strain, we determined the relative abundance of each member of the pair. In general, we did not find a significant relationship between co-culture fold-growth and the ratio between the pairwise consortium members, in contrast to previous results (Wintermute & Silver, 2010b) (Figure 11). Furthermore, our results suggest that expensive and rarer amino acids tended to crossfeed better than cheaper and more common amino acids.

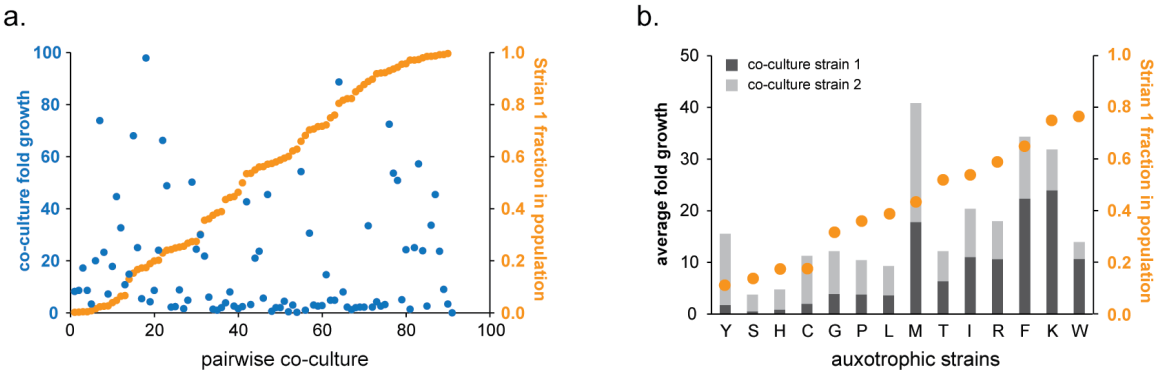

Figure 11: Growth yield of 91 pairwise co-cultures shows lack of relationship between relative population abundance and syntrophic growth. (a) Co-culture growth of each of 91 pairwise 2-member consortia. Left ordinate axis (in blue) shows whole co-culture fold growth after 84 hours for each of 91 consortia (data in blue dots). Right ordinate axis (in orange) shows fraction of Strain 1 in population for each consortia (data in orange dots). Consortia are ranked from lowest to highest based fraction of population dominated by Strain 1. The pairwise co-culture index number indicated on the x-axis refers to the co-culture rank ID. (b) Calculated average co-culture growth for each auxotroph paired with all 13 other possible auxotrophs. Left ordinate axis (in black) shows whole co-culture fold growth after 84 hours for each auxotroph, which is designated Strain 1. Color stacked bars indicates average fold growth contributed by Strain 1(in black) and Strain 2 (in gray). Right ordinate axis (in orange) shows average fraction of Strain 1 in population for each auxotroph group (data in orange range).

3.2.3 Quantifying 3-member Cross-feeding systems

To further explore the scalability of our synthetic ecosystems, we turned to higher-order syntrophic interactions. We devised 3-member synthetic consortia where each member is auxotrophic for two amino acids. Growth of each member can only occur when both amino acids are provided by partner strains. Member 1 (*e.g.* double auxotroph *MF*) does not grow with Member 2 (*e.g.* double auxotroph *MK*) because both are auxotrophic for *M*, methionine, but they can potentially co-culture together with a third member (*e.g.* double-auxotroph *FK*) that can provide *M*. At the same time, Member 3 can only grow in the presence of Member 1 and 2, thereby forming a syntrophic community. We experimentally probed these 3-

member syntrophies by first generating all 91 double-AA auxotrophic derivatives based on the 14 single-AA auxotrophs (Methods). All 91 strains showed no growth in the absence of extracellular supplementation of both needed AAs. Using these 91 double-auxotrophs, we systematically measured the growth profiles of all 364 possible 3-member consortia. After co-culture for 84 hours, we find a significant number of 3-member consortia to grow synergistically (Figure 12). As a control, growth was not observed when only 2 of the 3 members are co-cultured. Given that many microbes in nature are unable to synthesize multiple amino acids, our results demonstrated that higher-dimensional cross-feeding can yield productive syntrophic groups and are likely relevant also in natural microbial ecosystems.

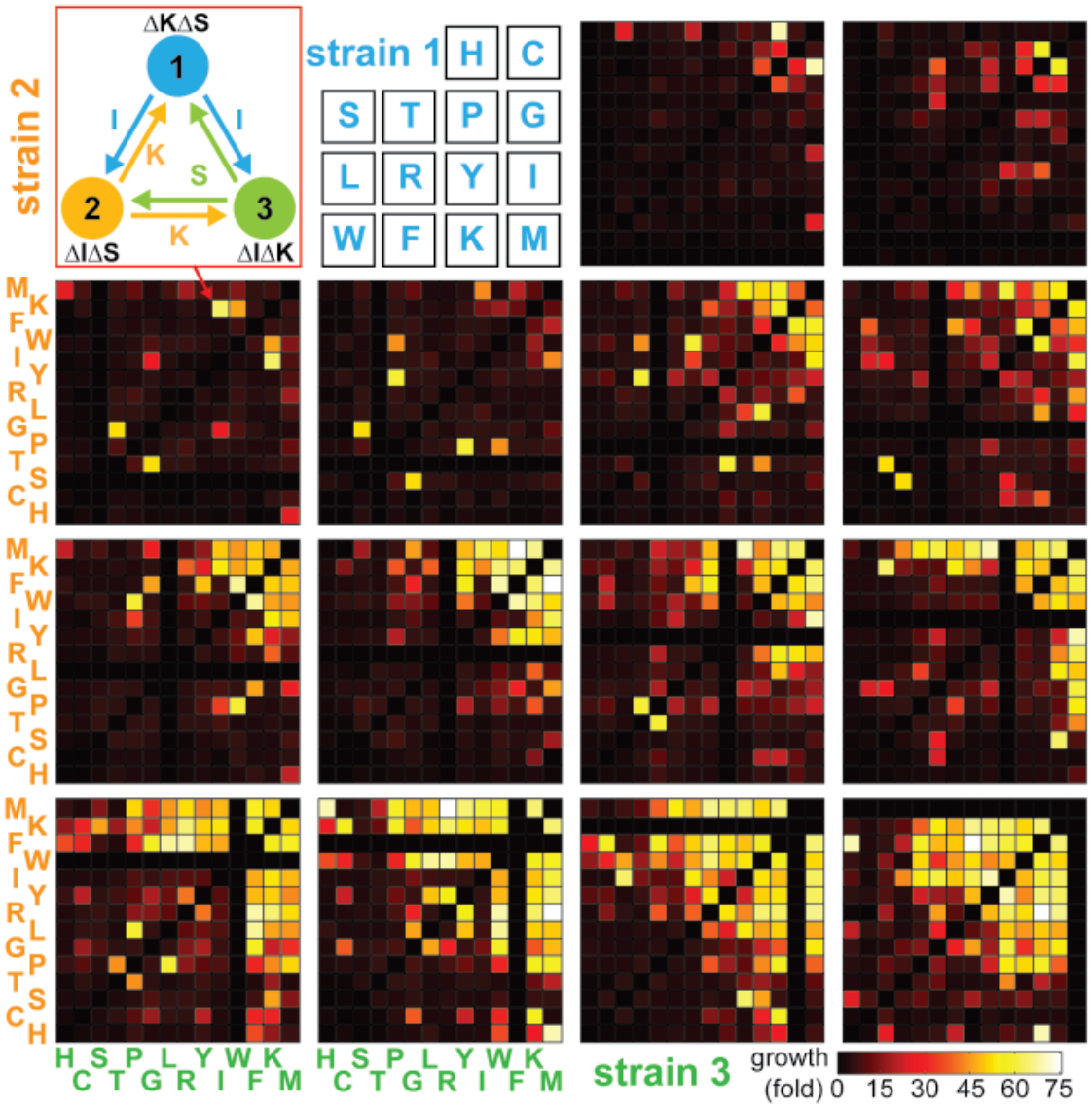

Figure 12: Three-member syntrophic consortia with each strain being auxotrophic for 2 amino acids. All combinations of 14x14x14 three-way interactions are measured after 84-hour of growth. Fourteen 14x14 panels are presented showing the growth yield of each 3-member group. Each 14x14 panel corresponds to a fixed strain 1 (blue color) against all combination of strains 2 and 3. The ordinate axis denotes different strain 2 (orange color) and the abscissa axis denote different strain 3 (green color). The key for strain 1 designation is shown in the second panel. The first panel illustrates an example consortium of KS-IS-IK with the cross-feeding amino acids shown by the correspondingly colored arrows. Color intensity indicated in the color bar denotes fold growth after 84 hours over initial population.

The overall growth profiles of the doubly-auxotrophic 3-member consortia match those of the mono-auxotrophic 2-member consortium (Figure 13a). Co-cultures involving amino acids *M*, *F*, and *K* tend to exhibit strong cooperative growth in contrast to poorly syntrophic amino acids *H*, *C* and *S*. In order to assess the predictability of syntrophic interactions when scaled to higher dimensional communities, we compared the observed fold-growth of all 364 3-member consortia with the fold-growth of each of their 2-member subset--for example, comparing doubly-auxotrophic triplets MF-MK-FK versus each of the mono-auxotrophic pairs M-F, M-K or F-K. We found four general classes of observations (Figure 13b, Zones 1-4). First, as expected, a majority of nonproductive 2-member cross-feeding interactions when scaled to 3-members also yield nonproductive growth (Figure 13b, Zone 1). Conversely, 3-member interactions where all 2-member subsets are productively cross-feeding also generate highly productive co-cultures (Figure 13b, Zone 2). Third, 3-member interactions where one of the 2-member subsets is nonproductive generally resulted in nonproductive 3-way co-cultures (Figure 13b, Zone 3). For example, the RMT triplet (RM-RT-MT) has a very limited growth of 3-fold even though the individual 2-member R-M pair yields 43-fold growth and the M-T pair yields 98-fold growth. The limiting group is the 2-member pair R-T, which yields less than 1-fold growth in co-culture. Thus the RMT triplet does not grow due to limited R-T cross-feeding despite robust cross-feeding by R-M and M-T. Finally, we find that a small group of nonproductive 2-member co-culture when placed together

showed positive epistatic synergy, resulting in more productive 3-way co-cultures (Figure 13b, Zone 4). Figure 3c summarizes the top consortia that exhibited this higher-dimensional synergy. For example, the PYT triplet culture (PY-PT-TY) results in 58-fold growth, in comparison to pairwise cultures P-Y, P-T, and T-Y that only grew by 8-fold, 2-fold, and 2-fold, respectively. These interactions highlight the surprising potential for both positive and negative epistatic interactions that exist for higher-dimensional syntrophic communities.

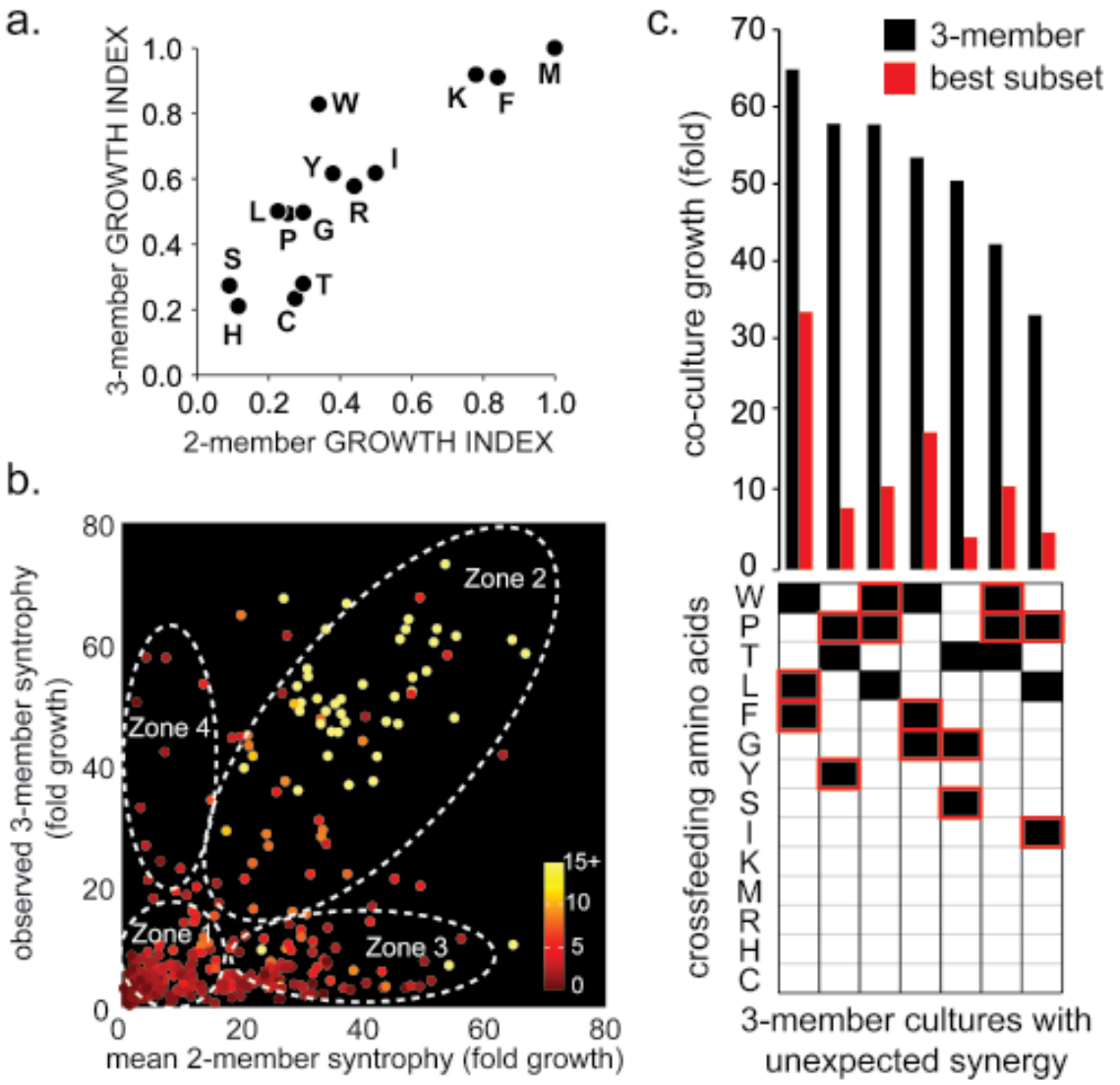

Figure 13: Comparison of 3-member syntrophies composed of double-auxotrophs against 2-member composed of single auxotrophs. (a) The sums of the final OD values for all 2- or 3-member communities containing a given auxotroph are normalized to the highest value (ΔM for both 2- and 3-member systems) to represent the syntrophic exchange growth potential of a given amino acid. This is termed the growth index for the 3-member or 2-member scenarios and shows consistent relationship when cross-feeding is scaled to higher dimensions. (b) Observed 3-member growth (ordinate axis) for all 364 triplets versus the mean growth of their three corresponding 2-member subsets (abscissa axis). Each point corresponds to a specific 3-member group. Color intensity of each point designates the lowest growth yield of the three 2-member subset and is mapped based on color bar using numerical values of fold growth (0 to 15+). Zones 1-4 are designated in the dotted regions, see text for detail. (c) Growth yield of top 3-member consortia that grow better than their corresponding 2-member subsets. Black bar indicates growth of 3-member. Red bar is the highest growth yield of the three 2-member subsets. The

corresponding 3 amino acids are shown in the bottom panel (read vertically) for each triplet. Two red boxes for each triplet designate the best 2-member subset.

3.2.4 ODE Modeling of 2- and 3-member Cross-feeding

A set of simple dynamic equations was used to model these syntrophic interactions (Figure 9d). In our model, growth of Strain 1 is linearly proportional to the amount of Strain 2 through the cooperativity coefficient term (*c*) for each pairing. The co-culture reaches saturation at carrying capacity (*k*) and a buffer term (β) is used for low density growth. We determined the cooperativity coefficients that best fit our experimental dataset, which provided the foundation for modeling higher-dimensional interactions.

To assess whether a naïve model can predict the observed growth profile of our 3-member consortia, we used the cooperativity coefficients (*c*) derived from the pair-wise 2-member interaction to build a simple 3-member dynamic model, which is described by the following system of equations:

$$\frac{dX_1}{dt} = \frac{X_1}{X_1 + \beta} \cdot \min\left(c_{1,2}X_2,\ c_{1,3}X_3\right)\left(\frac{1-(X_1+X_2+X_3)}{k}\right)$$

$$\frac{dX_2}{dt} = \frac{X_2}{X_2 + \beta} \cdot \min\left(c_{2,1}X_1,\ c_{2,3}X_3\right)\left(\frac{1-(X_1+X_2+X_3)}{k}\right)$$

$$\frac{dX_3}{dt} = \frac{X_3}{X_3 + \beta} \cdot \min\left(c_{3,1}X_1,\ c_{3,2}X_2\right)\left(\frac{1-(X_1+X_2+X_3)}{k}\right)$$

where X_1, X_2, and X_3 are the population sizes of the three members, β is the buffer term, and *k* is the carrying capacity of the population. Growth of strain 1 (dX_1/dt) is dictated either by the amount of strain 2 (X_2) times its cooperative

coefficient ($c_{1,2}$) or the amount of strain 3 (X_3) times its cooperative coefficient ($c_{1,3}$), whichever is the limiting value. The $X_i/(X_i+\beta)$ term is used such that at very low X_i levels, growth is proportional to X_i/β as with standard exponential growth, but $X_i/(X_i+\beta)$ becomes 1 at moderate X_1 values. The last term $(1-(X_1+X_2+X_3))/k$ is used to limit the density of the saturating culture. Using this simple dynamic model, we find that the predicted fold-growth profile showed statistically significant correlation with the observed 3-member fold-growth (Pearson coefficient r = 0.51, p-value = $5.2\text{x}10^{-25}$). When the zones representative of positive and negative epistatic interactions are removed (3&4), the fit of the model is greatly improved (Pearson coefficient r = 0.655, p-value = $1.\ 2\text{x}10^{-35}$). This suggests that our 3-member model can capture a significant fraction of syntrophic interactions based on the 2-member interactions. Discordance between the model and observed results highlight the potentially nonlinearly synergistic or antagonistic interactions of certain syntrophic consortia worthy of follow-up studies.

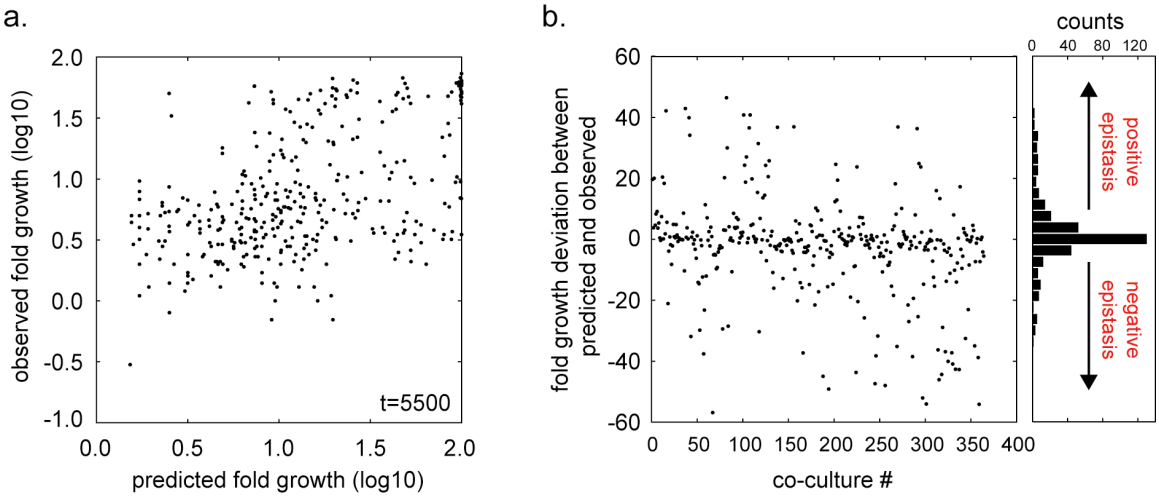

Figure 14: Evaluation of dynamic growth model of 3-member consortia. (a) Scatter plot comparing predicted fold growth of 3-member consortia (in log_{10} units) versus observed fold growth of 3-member consortia (in log_{10} units). Each black dot corresponds to each of the 364 possible unique consortia. (b) Scatter plot showing the difference between predicted and expected fold growth for all 364 consortia. Histogram on the right shows the distribution across all consortia indicating general agreement between predictions and observations. Positive values indicate potential presence of positive epistasis while negative values indicate potential negative epistasis.

3.2.5 Partially Syntrophic Higher Order Systems

The microbial communities explored thus far are *strictly syntrophic (ss)* because all strains rely on each other to grow and no subset grouping can grow alone. On the other hand, most natural ecosystems are likely composed of *partially syntrophic (ps)* interactions where one or more of the subset grouping can grow by themselves. For example, in contrast to the strictly syntrophic 3-member group MF-MK-FK that requires all members present, the partially syntrophic group MF-MK-HF contains a 2-member subset (MK-HF) that can potentially grow without the third member (MF). At longer timescales, one might predict that all *partially syntrophic (n)-member* interactions are reduced to a more minimal

strictly syntrophic consortium of fewer members. Thus, the evolution of partial syntrophy to strict syntrophy is of great importance to formation of sustainable cross-feeding consortia.

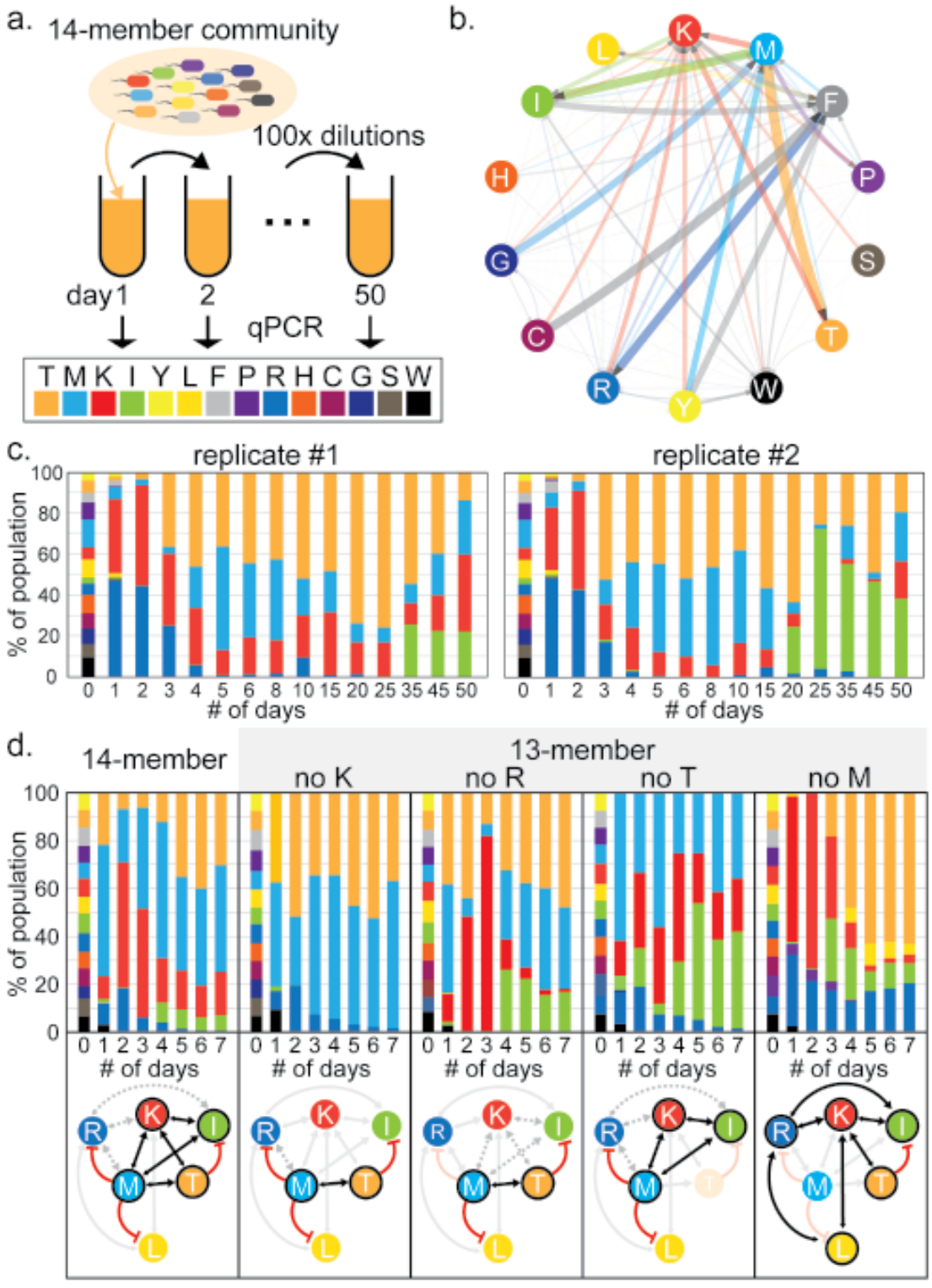

Figure 15: Dynamics of a 14-member syntrophic consortium. (a) Fourteen different single amino acid auxotrophs where combined in equal ratios to form a pooled mixture and

passaged daily in minimal media over 50 days. Samples of the population were periodically measured to determine the absolute and relative abundance of each of the 14 auxotrophs. (b) Syntrophic interaction map generated from all measured 91 pairwise cross-feeding experiments. Each auxotroph is designated by a circle and a different color. Arrowed lines correspond to the directional interaction from each strain to all 13 corresponding partners. Lines are color-coded according to the directional benefit the receiving strain is gaining from the donor strain (e.g. all incoming lines to the K auxotroph are red, designating the benefit gained by K from each donor). Increased thickness and opacity of the lines quantitatively denote increased cooperative benefits. (c) Population distribution of two biological replicate 14-member populations over 50 daily passages. Each colored bar section denotes the fractional composition of each auxotroph in the population. Color coding is the same as that of (a) and (b). (d) Subsequent short-term 7-day experiments of the 14-member population as well as 13-member populations that excluded one of four dominant amino acids (K, R, T or M) from the initial population. The syntrophic interaction network is shown below each panel. We denoted cooperative interactions with bidirectional black arrows and competitive (seemingly inhibitory) effects by directional blunted red arrows. Each auxotroph dropout and their associated interactions are shown in faded colors. Transient cooperative interactions are shown as dotted gray arrowed lines. Black circles around each amino acid designate final fixation to a stable community of 2-5 members. Values are derived from the average of three to four biological replicates.

We tested the predicted reduction from partial to strict syntrophy by devising a synthetic consortium using the 14 mono-auxotrophs that had been characterized. Combining all 14 auxotrophs into one pool produces a *partially syntrophic 14-member* consortium (Figure 15a). In principle, each auxotroph requires only one other partner to survive and thus may reduce over time to a *strictly syntrophic* 2-member population. Pairs that are able to grow the fastest are likely enriched. This system highlights the scenario where balancing between competitive growth and maintenance of cooperativity is key. Based on our prior 2-member cross-feeding results, we first constructed a 14-member interaction map (Figure 15b) and devised the following dynamic model to predict the possible outcomes of this 14-member consortium:

$$\dot{X}_i = \frac{X_i}{X_i + \beta}\left(\sum_{j=1, j\neq i}^{14} c_{ij} X_j\right)\left(\frac{1 - \sum_{j=1}^{14} X_j}{k}\right)$$

Here, X_i is the population size for strain i and c_{ij} is the cooperative coefficient between each 2-member pair i and j that we determined previously and terms β and *k* are as described previously. Several general predictions are noted. First, the resulting population is likely to be dominated by strains *M*, *F*, *K*, *I*, *R*, *T* and *W* strains based on their cooperativity profiles (Figure 16). Strain *M* is likely to sustain *I*, *T* and *K* due to the directionality of the cooperativity and thus may be a hub for any syntrophic interaction. Strain *F* is likely to derive benefit from multiple sources including *Y*, *C*, *I* and *K*, while only contributing to the growth of *R*. Strain *K* is likely to benefit from many strains including *M*, *T*, *R*, *C* and *Y*. Strains *L*, *W*, *P* is expect to have modest contributions. Finally, we expect strains *H*, *G*, *C*, *Y* and *S* to not be major components of the consortium because they do not generally yield productive pairwise cross-feeding.

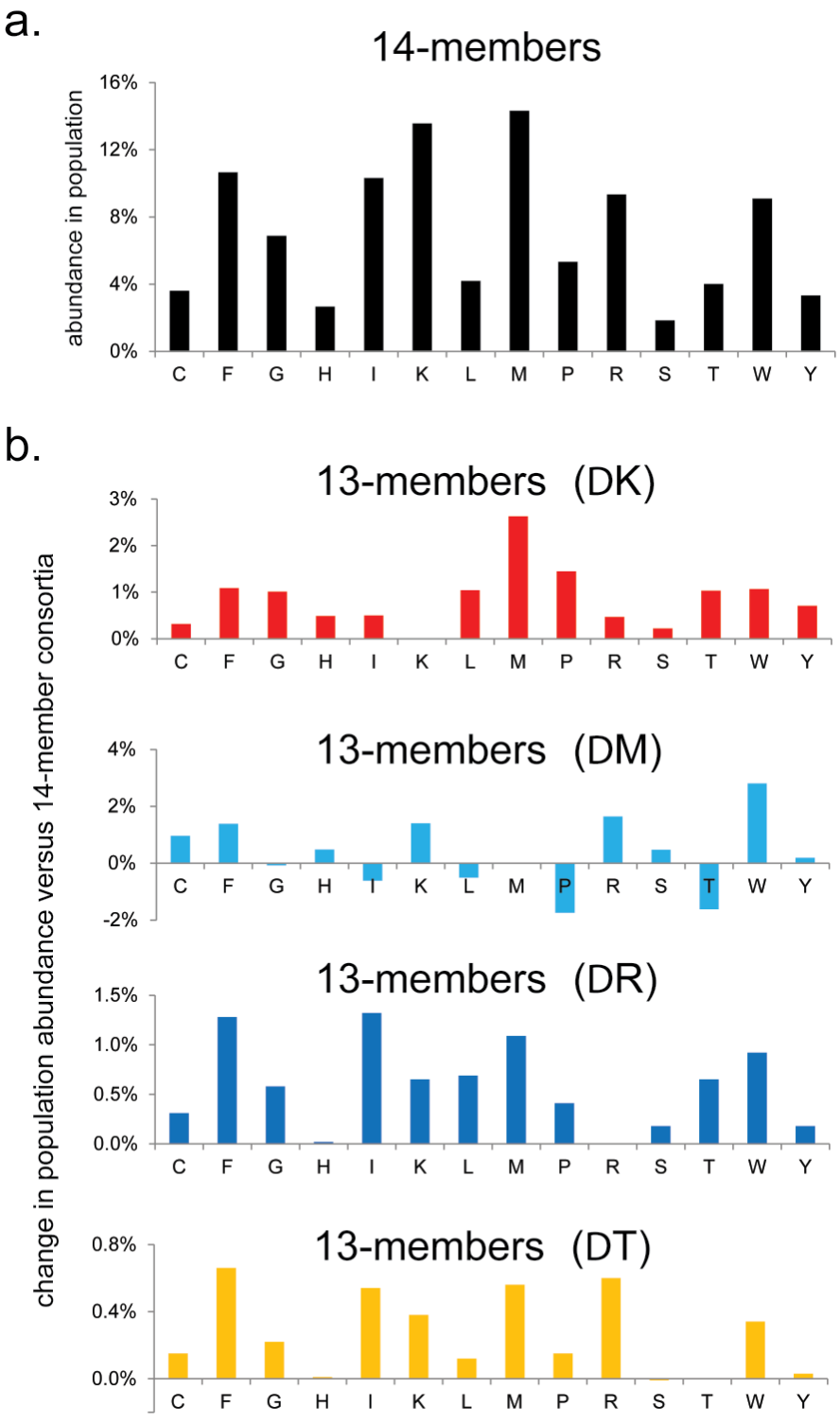

Figure 16: Predictions of dynamic growth model of 14-member consortia. (a) Bar graph showing predicted steady state population abundance of 14-member mono-auxotrophs. Values are percentage of whole population. (b) Bar graph showing predicted change in

steady state population abundance of four 13-member dropout consortia (no K, no M, no R, or no T) versus the 14-member consortium. Values are percentage of whole population.

Starting from a 10^7 cells/ml population composed of equal number of each of the 14 mono-auxotrophic strain, we experimentally passaged two identical replicate populations in M9-glucose for 50 days (~400 generations) and tracked the population abundance using quantitative PCR (Figure 4a). At the end of each 24-hour period, the population reached saturation and was diluted 100-fold for the next passage. At short time scales, we find that the 14-member systems undergo a drastic population shift towards a consortium dominated by 4 members (*R*, *K*, *M*, *T*) after only 2-3 daily passages (Figure 15c). Eventually, *R* is replaced by *I* and the population ratio varied from time to time but the member composition remained stable over the course of 50 days. The experiment was terminated at 400 generations when mutants were discovered in both replicates that confounded basic interpretations. We thus focused on the short-term dynamics of the system that are not subject to mutational events at evolutionary time-scales. Satisfactorily, the *K*, *M*, *R*, *I* and *T* strains dominant in the passaged population are also predicted to be predominant by our dynamic model. However, aromatic auxotrophs (*F*, *W*) were not seen despite their capabilities to crossfeed with others during individual pairwise matchups (Figure 9b). These pairwise interactions almost always benefited the *F* or *W* strain more than the partner strain, which may be in part responsible for their absence in the total mixture. Nonetheless, the model was able to capture a majority of the basic and important

features of this otherwise complex community.

To further probe the structure of the syntrophic network, we systematically tested 13-member consortia where one of the four initial dominant strains (*R*, *K*, *M*, *T*) was left out of the population. The composition of the mixture was tracked over 7 daily passages (Figure 15c). A number of interesting observations were seen. In contrast to the 14-member population, a 13-member consortium absent of the *K* auxotroph resulted in stable dominance of the *M*-*T* co-culture only (Figure 15c, panel 2). We note that the biosynthesis pathways of methionine and threonine converge upstream at a single common precursor, L-homoserine, and speculate the *M* auxotroph (Δ*metA*) could result in shunting biosynthetic flux in this branched synthesis pathway towards increasing *T* production. Conversely, the T auxotroph (Δ*thrC*) could result in the opposite shunt leading to increased *M* biosynthesis and has been shown to lead to increased M excretion (Usuda & Kurahashi, 2005). Indeed, we find that the *M*-*T* pair yielded the highest fold-growth amongst all 91 pairwise co-cultures (Figure 9c), suggesting that *M* and *T* could provide additional cooperative benefit to one another specifically for this pairing. Furthermore, the presence of characterized exporters for L-methionine and L-threonine could further facilitate syntrophy in this subgroup (Zakataeva, Aleshin, Tokmakova, Troshin, & Livshits, 1999).The synergistic effect of branched pathway shunt is also seen for the M-I, K-I, K-T and G-C pairs although to a lesser extent. Lower synergism in these pairs may be due to increasing numbers of intermediates and potential for branch down-regulation as previously

suggested(Wintermute & Silver, 2010b). Additionally, non-exclusive shunting could diminish the effect size for cases such as G-C where ΔglyA and ΔcysE can redirect metabolic flux to a third parallel pathway, towards W in this case. Similarly, we found that the *R*-absent population also led to the dominant *M*-*T* co-culture, although the dynamics of the population was very different (Figure 15c, panel 3). In the *R*-absent population, the *I* auxotroph bloomed, making up 20% of the population at Day 4, but was eventually outcompeted in subsequent days. In contrast, the *I* auxotroph was not detected in the *K*-absent population. Interestingly, a *T*-absent population (Figure 15c, panel 4) resulted in expansion of the *I* subpopulation as well as the maintenance of the *K* subpopulation, thus producing a stable *M*-*K*-*I* consortium. Together, these results suggest that the *T* and *I* auxotrophs are competing for similar cooperating partners such that the presence of *T* limits the growth capability of *I*, but absence of *T* allows growth of *I*. *T*'s competitive advantage over the *I* strain in this mixed environment could be explained by the fact that biosynthesis of *I* requires *T* as an essential precursor and therefor *I* is likely sharing an additional costly metabolite (*T*) with the other strains. Finally in *M*-absent population, *L* is present in addition to *K*, *I* and *R*. In contrast to the other mixtures, the *R* subpopulation was stably sustained over the course of 7 days. We integrated all these subtractive experiments to form a reconstructed topology of the syntrophic network (Figure 15c). This interaction network recapitulates the important properties of the partially syntrophic

community and highlights the role of transient dynamics in the development of syntrophic community.

3.3 Discussion

Pairwise amino acid cross-feeding experiments showed that these simplified interactions can be quantitatively modeled. When these interactions are scaled to 3-membered communities requiring cooperative behavior by all individuals involved, we also found many examples of syntrophic growth. The general trends in cooperative exchange were conserved in these higher dimensional communities. Furthermore, we observed positive epistasis within a small subset of 3-member consortia that performed better than their 2-member constituents. It is feasible to further increase the complexity of the system by developing triply-auxotrophic strains that interact as 4-member strictly syntrophic communities. These co-cultures demonstrate that microbes with multi-auxotrophic phenotypes can rely on direct cross-feeding for survival.

Our 14-member consortium showed that cross-feeding interactions can often be quite complex and the system may not necessarily converge to an expected simpler 2-member community through hundreds of generations. Various resulting 3- or 4-member populations were often stable over evolutionary timescales, but removal of certain keystone members (*e.g.* K or R) from the initial population resulted in convergence to the best pair M-T. While our dynamic models were able to capture general features of this system, specific quantitative predictions were less accurate, thus highlighting the current limitations to model

ecosystems of even moderate complexity (Klitgord & Segre, 2011).

It is important to note that microbial cross-feeding studied here relies on export of amino acids into the shared extracellular environment. Such membrane transport systems have been recently characterized and many more are being found with the help of metagenomic sequencing. Our *E. coli* genome encodes several amino acid exporters for excretion of different amino acids including L-threonine (*RhtA* and *RhtC*) (Zakataeva et al., 1999), L-leucine (*YeaS*) (Kutukova et al., 2005), L-aromatic amino acids (*YddG*) (Doroshenko et al., 2007), L-arginine (*YggA*) (Nandineni & Gowrishankar, 2004), L-alanine (*alaE*) (Hori et al., 2011), and L-homoserine (*RhtB*) (Zakataeva et al., 1999). Other exporters have been documented in related organisms including *lysE* for L-lysine export and *brnFE* for L-isoleucine and L-methionine export in *C. glutamicum* (Kennerknecht et al., 2002; Vrljic, Sahm, & Eggeling, 1996). In addition to active transport, some level of passive transport may also be involved since hydrophobic amino acids such as I, F, Y, W have membrane permeability that is 100 times greater than hydrophilic amino acids (Chakrabarti, 1994). Recently, the *YddG* aromatic amino acid exporter has been exploited to tune microbial cross-feeding in an *E. coli* Y-W syntrophic system (Kerner et al., 2012). These systems present an opportunity to study long-term evolution of microbial ecosystems and the enhancement of cooperative phenotypes. While metabolic cross-feeding could be exploited by selfishly cheating phenotypes, the formation of spatial architectures such as biofilms and aggregates may help to prevent such scenarios. Furthermore,

quorum sensing and response to the presence of cooperators may further help to drive the development of multi-species syntrophic growth.

We believe that the multi-dimensional syntrophic system presented here provide a useful foundation for studying and engineering microbial communities of increasing sophistication. These synthetic approaches can be used to study natural microbial communities such as the human microbiome in specific ways towards unraveling the complex interactions at play (Faith et al., 2011). Advances in synthetic and systems ecology will offer new avenues to explore and exploit natural and defined microbiota to develop sustainable solutions to global health, energy, and environmental issues.

3.4 Materials and Methods

3.4.1 Strain construction and verification

All strains used were based on the EcNR1 *E. coli* derivative of MG1655, which carried an integrated, temperature inducible, λ-Red prophage for recombineering (H. H. Wang et al., 2009). Each of the 14 amino acid (AA) auxotroph was generated by Red-recombineering as previous described (Yu et al., 2000) of a chloramphenicol resistance cassette into each of the following targets separately: *argA* (R), *cysE* (C), *glyA* (G), *hisB* (H), *ilvA* (I), *leuB* (L), *lysA* (K), *metA* (M), *pheA* (F), *proA* (P), *serA* (S), *thrC* (T), *trpC* (W) and *tyrA* (Y). Double AA auxotrophs were generated by introducing a kanamycin resistance cassette into the same set of targets in the 14 single AA auxotrophs to generate the 91 double auxotroph strains. All single AA knockout strains were confirmed to

be auxotrophic as evidenced by the lack of growth in minimal M9-glucose media after 4 days. All double AA knockout strains were confirmed to not grow in M9-glucose supplemented with only one of the two amino acids. Presence of the inserted antibiotic cassettes was additionally verified in each strain via allele specific PCR.

3.4.2 Co-culture growth conditions

Strains were first picked from an overnight colony into LB-Lennox medium (LBL, 10 g/L bacto tryptone, 5 g/L NaCl, 5 g/L yeast extract) with selective antibiotics as appropriate (chloramphenicol 20 ug/ml; kanamycin 50 ug/ml). Late exponential phase cells were harvested and washed twice in M9 salts (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl) by centrifugation at 17,900x g. Cell concentrations were determined based on OD600 readings from a spectrophotometer. Prior to co-culture experiments, all cell concentrations were adjusted to 10^7 cells/mL using M9 media. Co-culture growth was performed by equal volume inoculation of each strain at a seeding density of 10^7 cells/mL. All 2-member and 3-member co-cultures were grown in 200ul of M9-glucose media (M9 salts supplemented with 1 mM MgSO4·7H2O, 0.083 nM thiamine, 0.25 ug/L D-biotin, and 0.2% w/v glucose) in 96-well microtiter plate format in an incubator or a platereader at 30 °C to maintain λ-Red prophage repression. Microtiter plates were shaken at >500 rpm to maintain aerobic growth. Growth of 13- and 14-member co-cultures was done in 3 mL cultures in a 30 °C rotating drum and passaged without washing by 100-fold dilution every 24 hours as the cultures

reach saturation. The growth and fold-growth metrics mentioned throughout the text refer to the yield of the community calculated by final cell density/initial density. Biological replicates were performed by splitting a single well-mixed initial seeding population.

3.4.3 Kinetic growth assays and strain identification

For precise determination of the cell density at amino acid limited conditions, OD600 readings were taken every 5 minutes during exponential growth at 30 °C in a spectrophotometer (M5 Molecular Devices) with >500 rpm orbital mixing. Calibration between OD600 measurement and actual cell density was determined by resolving the colony forming units (cfu) of each auxotrophic strain plated on solid media at different concentrations. All growth experiments were performed in M9-glucose media in the absence of antibiotics. Proportional strain abundance was determined via quantitative PCR. All population samples were frozen at -20 °C and assayed simultaneously to reduce run-to-run variations. Quantitative PCR were performed in 20ul reactions with 10 uL KAPA SYBR Fast Universal 2x MasterMix (KAPA KK4600), 4 uL of a 10X dilution of frozen cells, and 6 uL of primer pairs resulting in the following final primer concentrations: R,K,M,P,T,W,Y (200nM); G,I (150nM); H,S (100nM); C,L,F (50nM). PCR conditions were based on manufacture's recommendations (40 cycles of combined annealing extension at 60 °C for 20 s) and performed using a thermal cycler (Bio-Rad CTX96). The corresponding number of cycles was determined at a relative fluorescence unit of 150. The half-max values of each

qPCR curve was calibrated to actual cell density by serial dilution plating of each auxotrophic strains and determination of cfu on solid agar plates. The 13- and 14-member qPCR control populations were performed using equimolar mixtures of all auxotrophic strains. Relative proportion of each strain the population as determined by qPCR was further verified by plating using colorimetric assays. For 2-member and 3-member communities, each strain carried deletions Δ*malK*, Δ*lacZ*, or ΔmalK/ΔlacZ, which could be visually distinguished on MacConkey-maltose plates (BD Difco) supplemented with XGAL-IPTG (Growcells).

3.4.4 Dynamic model of 3-member consortia

The 3-member kinetic model was simulated using the Matlab ® environment. The model was initially seeded with 10^7 cells of each strain. Cooperativity coefficients were taken from pairwise cooperativity coefficients (c_{12}, c_{21}). Predicted fold growth was calculated by dividing time step t=5500 population values by the initial seeding value. The carrying capacity was set to 10^9 cells in all simulations. The β value was set to 1. Comparison between the predicted fold growth and observed fold growth is shown in Figure 14a. In general, we find high predictive power in a significant fraction of the consortia (Pearson coefficient r = 0.51, p-value = 5.2×10^{-25}). Differences between the predicted and observed fold growth highlights possible positive and negative epistatic interactions (Figure 14b).

3.4.5 Dynamic model of 14-member consortia

The 14-member kinetic model was simulated using the pairwise cooperativity coefficients (c_{12}, c_{21}). The population is seeded with equal amount of all 14 auxotrophs totaling 10^7 cells. The carrying capacity was set to 10^9 cells in all simulations. The β value was set to 1. The resulting steady state population abundance was determined as the percent of the total population dominated by each auxotroph type (Figure 16a). For each 13-member dropout simulations, each dropout auxotroph was set to an initial seeding value of 0 and all other aspects of the simulation were unchanged. The difference in population abundance for each auxotroph in the 13-member groups compared to the 14-member group is shown in Figure 16b.

4 Microbial Evolution in Amino Acid Cross-feeding Communities

4.1 Overview

As bacterial species grow, their genomes accumulate mutations that allow them to sample adjacent genotypes and phenotypes*(Kibota & Lynch, 1996)*. In part, the context of the surrounding environment determines the impact of a mutation the host strain's fitness and whether it is enriched in the population. Previous studies have identified the mutations that accumulate as an initially isogenic bacterial population undergoes divergent evolution to occupy various niches(Barrick et al., 2009; Chubiz, Lee, Delaney, & Marx, 2012; Fong, Joyce, & Palsson, 2005; Papadopoulos et al., 1999). In nature, bacteria often grow in the context of complex microbial ecosystems where individual species confront a dynamic metabolic milieu, chemical warfare in addition to competition for resources complicating the deconvolution and analysis of accumulated mutations(Cooper & Lenski, 2000; Czaran, Hoekstra, & Pagie, 2002). Here we propose synthetic bacterial ecosystems as a means to isolate amino acid auxotrophy and exchange from other selective pressures common to bacterial communities. A defined yet heterogeneous mixture of 14 amino acid auxotrophic strains was combined and passaged for ~650 generations in minimal media. Each strain derives its requisite nutrients from the rest of the population. We hypothesized that this partially syntrophic community would converge a strictly syntrophic community comprised of the two strains best able to complement each others' auxotrophy(Mee, Collins, Church, & Wang, 2014). Instead, we

observed maintenance of 4 auxotrophic lineages that accumulated mutations in reaction their distinct metabolic deficiencies. Fifty-eight isolates from two time points and separate replicates were sequenced and demonstrated common evolutionary strategies such as perturbation of biosynthesis, mobile element and frame shift disruptions of stress response pathways and tandem duplications of genomic regions enriched for amino acid importers. Interestingly, this environment enabled many of the initial auxotrophic lineages to sample mutations allowing them to overcome their engineered metabolic deficiencies. These results reflect the incredible robustness of bacterial genomes where loss of function can be regained through varied strategies.

4.2 Results

A synthetic bacterial consortium was constructed by pooling at equal densities 14 variants of *Escherichia coli (E. Coli)* each engineered to be auxotrophic for a single amino acid (designated by their respective initial engineered auxotrophies as ∆-C, F, G, H, I, K, L, M, P, R, S, T, W, and Y). As previously described, six replicates of this community were passaged daily in M9 minimal media where each member of this partially syntrophic community is dependent on the other members to exogenously provide the amino acid essential for it's growth(Mee et al., 2014). Expanding on previously discussed results, here we include additional replicates to further demonstrate the divergence of the community structure as mutations are accumulated. Figure 17

illustrates that by the 85th passage, the relative abundances of the remaining strains in each mixture bear little resemblance to each other or to the core group of 4 strains (R, K, M, and T) settled upon in the first 6 passages. We observe a drop in average doubling time of the mixtures from a high of 225±12 min on day 2 to a low of 123±21 min by day 5. We speculate that the mixture is settling on a structure optimal for efficient metabolic exchange. From day 5 to day 10 the doubling time increases on average by 5% per day reaching 157±25 min. This suggests that adaptations accumulated in an individual strain over this period while beneficial for it's own fitness may not directly improve the productivity of the community. Also of note is the emergence to varying degrees of the isoleucine strain from relative obscurity.

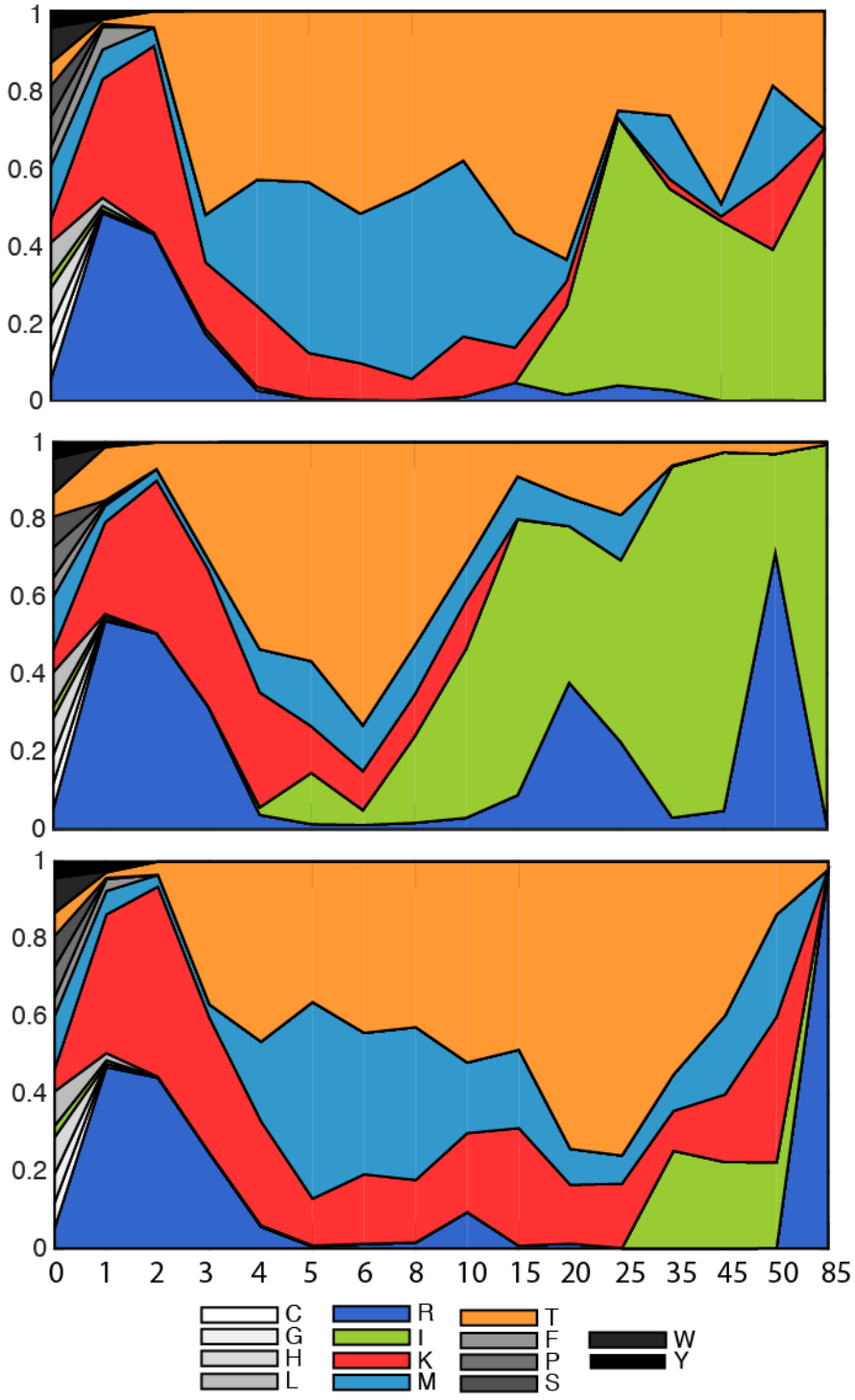

Figure 17: Long term dynamics of the synthetic microbial ecosystem for 3 of the 6 populations replicates

In an attempt to elucidate the underlying causes of community structure divergence, 32 strains from each mixture (192 total) at both day 50 and 85 (roughly 400 and 650 generations) were isolated by plating on LB-Lennox media and genotyped by allele specific PCR to confirm the engineered autotrophy. The growth rate in LB-Lennox for all isolates was determined using microtiter plate reader kinetic growth assays as previously described(Mee et al., 2014)(Figure 18a). The disparity between evolved and ancestral growth rates is indicative of fitness altering mutations we predicted were present. More specifically, the general increase in doubling times in rich media suggests that the mutations accumulated while grown in a cross feeding community render the strains less fit for growth in LB. To further investigate alterations to the phenotypes of these strains, maintenance of auxotrophy was assed by growing all isolated strains on M9 minimal media devoid of any supplemented amino acids. Surprisingly, many of these isolates demonstrated varying degrees of growth indicating that the selective environment of a partially syntrophic cross-feeding community allowed for reversion to prototrophy. In fact the over 50% of the assayed strains had reverted to prototrophy (here conservatively defined as having surpassed an OD600 value of 0.25) by time point 50 (Figure 18b – left panel) and over 75% of the strains were prototrophic by timepoint 85 (Figure 18b – right panel). Interestingly, the isoleucine strains had a disproportionate number of prototrophy reversions relative to the ΔK, ΔT, and ΔR strains assayed. Although many strains became prototrophic they did not overtake the population suggesting that the

causative mutations did not return the strain to wild type fitness. Thirty strains from day 50 and twenty-eight strains from day 85 were selected for whole genome sequencing in a distribution representative of the observed initial autotrophies and evolved growth phenotypes in hopes of identifying causative mutations. The phenotypes of these selected strains were further investigated by subjecting them to fitness competitions against their ancestral strains in M9 minimal media supplemented with the appropriate amino (methods)

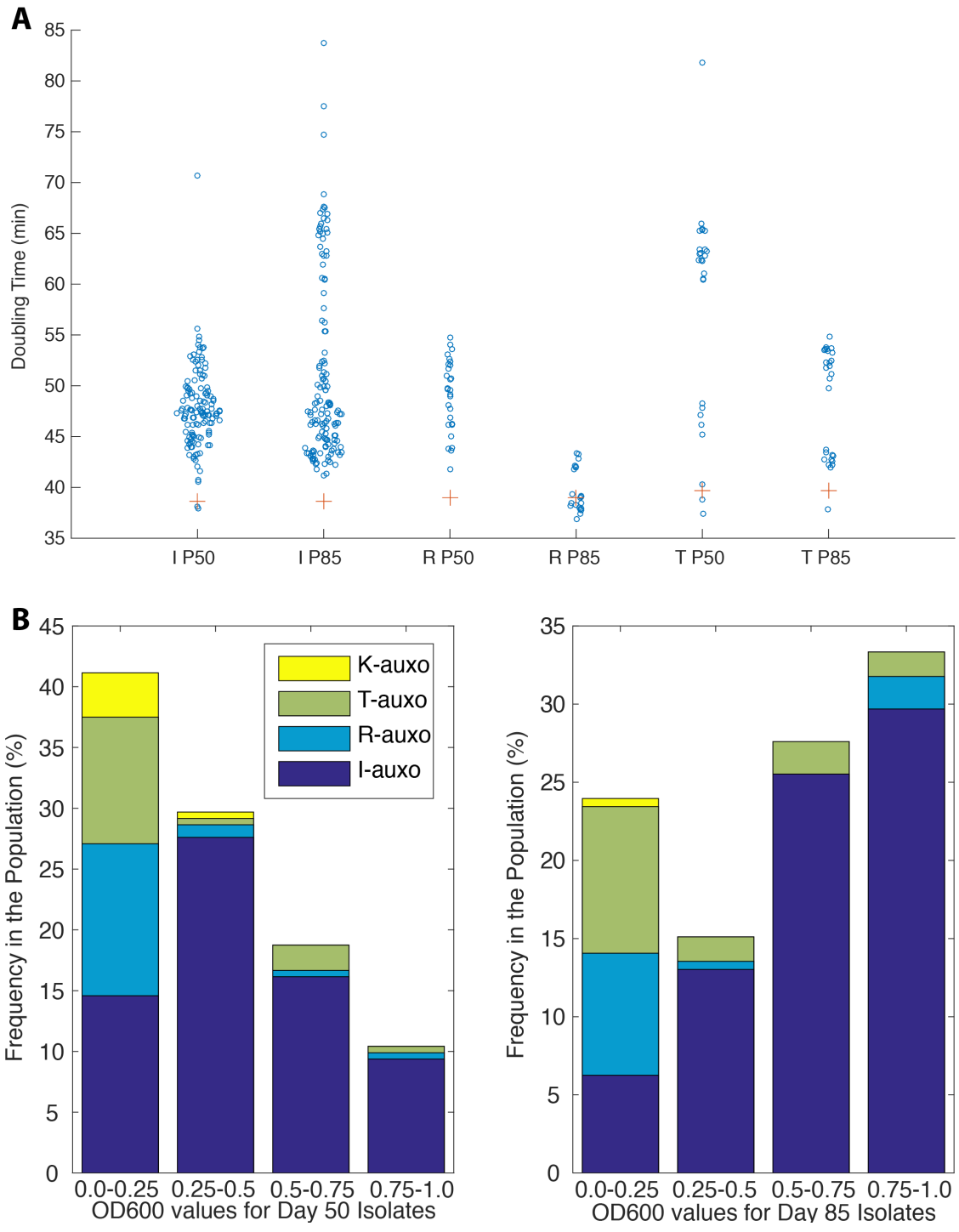

Figure 18: Evolving phenotypes of strains under cross-feeding growth. A) Doubling time in LB-L for all isolates from day 50 (P50) and day 85 (P85). Single letters designate the auxotrophic lineage of the assayed strain. B) Max OD reached in minimal media after 96 hours of growth. By day 85 many strains, especially those of the isoleucine lineage have regained a prototrophic phenotype.

4.2.1 Sequencing Results

The average coverage of the HiSeq run following alignment was 64.5x +/- 41.4x. Only a single strain, MM13ΔK, had average coverage below 20X. The number of mutations identified per isolate ranged from 6 to a high of 115. While no one specific mutation was the clear root of reversion to prototrophy for any given strain, there are a myriad of interesting candidates ripe for further investigation. Here, we highlight the most abundant and more interesting mutations found.

4.2.2 Mutations common across auxotrophies:

There are three mutations predominant throughout the sequenced strains with no enrichment bias in any one of the four observed auxotrophic lineages: ΔR, ΔK, ΔT, ΔI. The first consists of sequence junctions differing from the reference at the *fimA*/*fimE* locus found in 22 isolates. This structural variant has previously been reported and is responsible for phase variation whereby bacteria stochastically switch between fimbriated and non-fimbriated form resulting in a mixed population(Gally, Bogan, Eisenstein, & Blomfield, 1993). Stochastic switching of a population between these two expression states helps pathogenic bacterial strains evade a host's immune system(Goldberg, Fridman, Ronin, & Balaban, 2014). However, the fimbriated form has a significant fitness cost and is generally repressed outside of the host in conditions of high-nutrient exponential growth(Gally et al., 1993). Observation of phase variation in our

strains suggests that one of the several global regulators of phase variation is being affected by cross-feeding growth. Amino acid nutrient limitation is a likely culprit as it is know to trigger stress responses that modulate the global regulators in question(Blomfield, Calie, Eberhardt, McClain, & Eisenstein, 1993).

The second broadly disrupted region, mutated in roughly half of the isolates, is the *rph-pyrE* locus. The ancestral *E. coli* MG1655 strain has a 1-bp deletion in this locus resulting in low orotate phosphoribosyltransferase levels and suboptimal pyrimidine biosynthesis as a result of a disruption to the attenuation region upstream of *pyre*. This mutation reduces MG1655's growth rate in minimal media by 13%, making it an attractive target for evolutionary fitness gains(Conrad et al., 2009). An 82bp deletion was the most frequently identified mutation in this region. This seems to be a common evolutionary adaptation for alleviating the biosynthesis impairment as it has been identified in other laboratory evolution studies(Conrad et al., 2009). The critical nature of overcoming this metabolic flux limitation is reinforced by our observation of several additional mutational variants of this locus that may also be responsible for improving the organism's fitness including 1-bp deletions 53 & 55 bps downstream of *rph*, a premature stop codon at residue 213 of *rph*, and non-coding mutations 62/39 bases upstream/downstream of *rph* potentially further affecting the critical regulation of these *pyrE*. While additional work is needed to determine if any of these modification are hitchhiker mutations, taken together these observations reinforce the functional enrichment for perturbations to this

region.

The third common disruption is oddly depleted from strains in the isoleucine lineage. One third of all non-ΔI strains have mutations in or around *glnK*, a gene responsible for signal transduction under nitrogen starvation(Blauwkamp & Ninfa, 2002). The majority of these (6/7) are disruptive frameshifts resulting from 1-bp deletions but a mutation in the regulatory region of the gene, 4bp upstream of the start codon, is also observed. While a single isoleucine isolate also contained a mutation in the regulatory region (7 bp upstream of the start codon), there is a clear bias towards non-ΔI strains. We hypothesize that the increased rate of fixation of this disruption in non-ΔI lineages is due to an auxotrophy-dependent functional enhancement of this mutation. Modification to signal transduction for nitrogen starvation may have greater fitness benefits for the ΔR, ΔT and ΔK strains.

4.2.3 Auxotrophic Lineage Specific Mutations:

Enrichment for specific mutations within auxotrophic lineages is suggestive of a functional basis for their fixation. The mutations arising in these populations are likely adaptations to growth in the amino acid limiting conditions encountered under engineered cross-feeding. However, subsets of mutations are also likely to be mechanistically involved in the frequently observed reversion to prototrophy. To identify specific mutations associated with this striking prototrophy reversion phenotype, hierarchal clustering was performed (methods).

Sequenced strains whose mutation profiles cluster closely together but display distinct growth phenotypes are identified. Mutations specific to these prototrophic phenotypes are also identified and presented in the analysis below.

4.2.3.1 Mutations Enriched in Isoleucine Lineage:

As previously discussed, strains from the isoleucine linage revert to prototrophy more frequently than strains from the other lineages. The observed mutational strategies enabling these strains to mitigate the limitations imposed by isoleucine auxotrophy are grouped into functional classes and discussed below.

Regulator targets:

There were several mutations to regulator genes observed in isolates of this lineage. Arising independently in every population replicate, the most frequently observed mutation (present in all but 1) was an E45D substitution in *abgR*, the *lysR* type regulator of the *abg* operon. The frequency of this mutation suggests functional enrichment despite conserving the acidic nature of the affected residue. The regulated operon is responsible for the last step in the catabolism of diaminopimelic acid to lysine indicating that modulating metabolic flux through these affected reactions is likely critical to increasing the fitness of the isoleucine auxotroph(Hussein, Green, & Nichols, 1998). As this mutation resulted in acidic residue substitution and not a total loss of function, it may be possible that this residue substitution changed the DNA specificity of the

regulator, more broadly impacting the transcriptional state of the cell.

Another regulator, *dgsA,* is mutated in 48% of the sequenced ΔI isolates with the majority of the changes resulting in loss of function due to frameshifts or introduction of premature stop codons. This gene encodes a transcriptional repressor that controls expression of a number of genes in the phosphotranspherase system, phosephenyl pyruvate system, and genes involved in glucose uptake(Plumbridge, 2002). However, it is likely that the functional target of this mutation is the phosphotranspherase system as it is highly enriched for mutations in this lineage. Specifically, *ptsG* encoding one of the transporters responsible for beta-D-glucose import and also regulated by *dgsA*, has observed mutations in every ΔI isolate. Further supporting the hypothesis that disruption of phosphotranspherase system is highly beneficial to isoleucine auxotrophs is the fact that three separate classes of mutations arise independently in the *ptsG* locus: L425Q hydrophobic to polar residue substitution disrupting a hydrophobic region required for proper functionality(Nuoffer, Zanolari, & Erni, 1988); 1bp deletion mediated frameshifts; IS5 mobile element insertion 85bp upstream of the start codon. While it has been shown that inactivation of this gene results in a 20% reduction to growth rate(Tchieu, Norris, Edwards, & Saier, 2001), we find that disruptions to this gene are in fact correlate to reversion to isoleucine prototrophy (figure 20 prototrophic strains MM10, MM40 and MM48). The three prototrophic strains highlighted in figure 20 also have additional regulator system modifications in a potentially epistatic A272S

mutation to the allatonin transcriptional regulator *allS*. Further work is needed to quantify the relative contributions of these two regulatory modifications to the prototrophic phenotype. It is intriguing that modulation of sugar transport is uniquely advantageous to the strains auxotrophic for isoleucine biosynthesis as the carbon sources are equally available to all members of the community. The functionally disruptive nature of transporter mutations suggest affected cells may be shifting away from glucose as their primary carbon source. Modifications to the regulatory architecture of the cell likely enable such a shift and seem to be a common strategy to overcoming isoleucine auxotrophy.

Amino acid and TCA metabolism targets:

Genes catalyzing amino acid metabolism and tricarboxylic acid (TCA) cycle reactions of central metabolism were disproportionately affected by mutations in the sequenced isolates of the isoleucine lineage. This suggests an interesting functional archetype for adapting to isoleucine deficiencies through metabolic rerouting. Due to the linear nature of the isoleucine/threonine biosynthesis pathway, isoleucine auxotrophy-causing disruptions (such as the *ilvA* knockout used in this study) will likely increase in the intracellular threonine pool. Such shifts to intracellular amino acid pools can result in in greatly reducing the fitness of the strain (ref BCAA regulation). One observed strategy to mitigate fitness reductions from non-optimal metabolic equilibria seems to rely on mobile element mediated large-scale tandem genomic duplications. Two distinct regions

in separate strains had ~2-fold increases in sequencing coverage resulting in up-regulated expression levels for 2 groups of genes. The first duplication spans a 103,685bp region from *insJ* preceding *sokA* to *spoT* (Figure 19a). Critically, this region contains the alanine/valine transaminase encoded by *avtA*. It has been previously been reported that episomal over-expression of this gene rescues the growth of isoleucine auxotrophs in minimal medial(M. D. Wang, Liu, Wang, & Berg, 1987). The mechanism of this rescue is likely through shifting intracellular amino acid pools to a more desirable state. It is possible the increased abundance of the gene resulting from the genomic duplication increases activity enough to similarly rescue growth of this isolate. Both *tdh* and *yiaY* are also located in this duplicated region and encode threonine dehydrogenases, enzymes responsible for catabolizing the first step in the threonine catabolic utilization pathway(Boylan & Dekker, 1981). It is likely that these genes contribute to improving the fitness of affected strains through reducing the intracellular threonine pool. The increase to the combined action of these three genes is likely instrumental to the strong prototrophic growth phenotype we observed in the isolate with this duplication. A second tandem duplication spanning an 113,150bp region (from *nmpC* to the *gltI* adjacent instance of *insH)* is observed in seven isolates (figure 19b). The region is somewhat enriched for genes encoding catalytic activity(Mi, Muruganujan, & Thomas, 2013). The *cit* operon may be a specific target of enrichment due to its effect on the TCA cycle and potential to broadly modulate intracellular metabolic levels as previously

discussed. While further experimental validation is needed to determine the exact mechanistic effect of this duplication there is clearly a functional enrichment for this structural variant as it arises independently in 3 distinct population replicates.

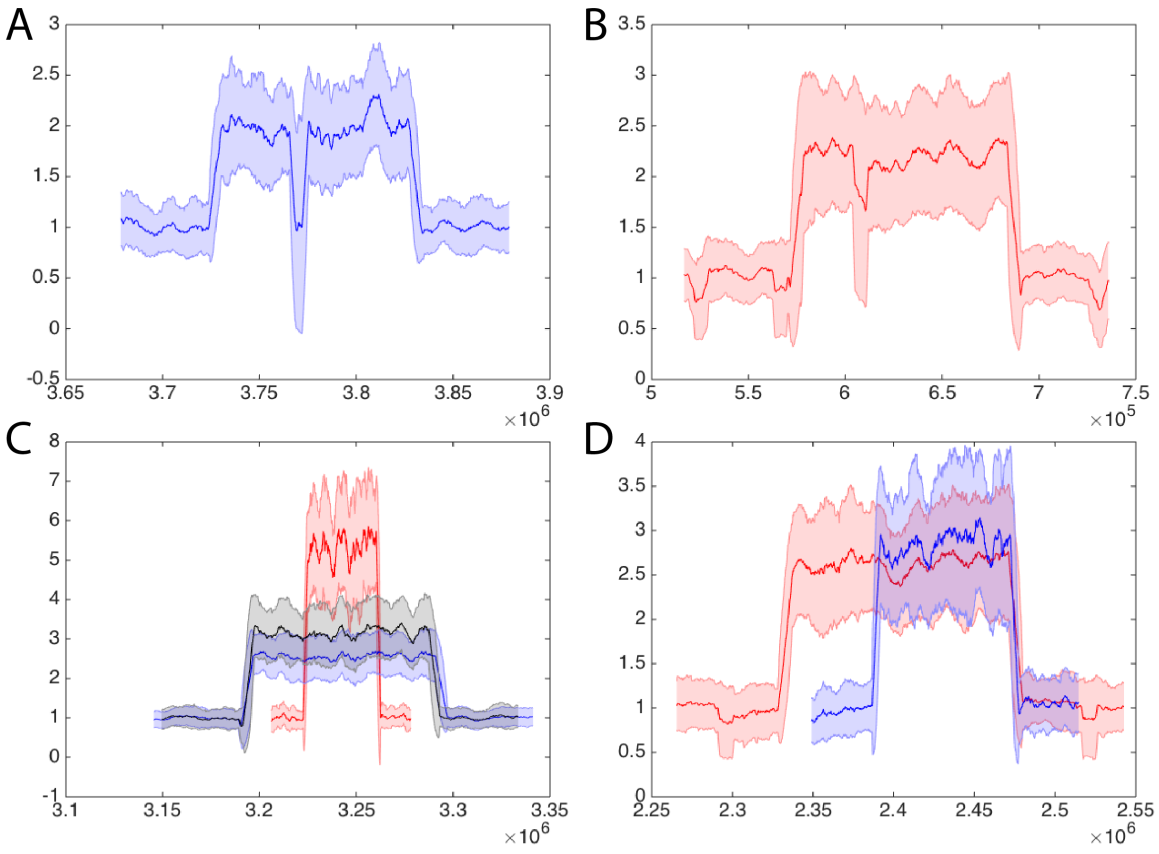

Figure 19: Normalized Coverage Enrichment of the Sequencing reads. Y-axis indicates coverage fold increase in coverage over average. X-axis indicates position in the reference MG1655 genome. Solid line indicates average coverage over a sliding 75bp window and the surrounding shaded area represents the standard deviation within the same window. A & B) Genomic duplications observed in the isoleucine lineage. C) 3 distinct genomic duplications observed within the Threonine lineage. D) 2 distinct duplications observed in the lysine lineage.

It has previously been reported that perturbations to branched chain amino acid biosynthesis result in increases to homocysteine toxicity(Tuite, Fraser, & O'Byrne C, 2005). A selective advantage may be endowed to strains that mutate to mitigate this toxicity. Three genes implicated in biosynthesis

reactions upstream of homocysteine synthesis (*cysE, metB and metQ*) have accumulated mutations in the isoleucine lineage. The mutations observed in *csyE* and *metB* (mutated in 78% and 32% of isoleucine strains respectively) are different across strains suggesting several unique changes are able to modulate the function of these proteins sufficiently to impact fitness. In *metQ* a 3bp in-frame deletion is observed midway through the ORF of the ABC methionine importer. Laws of classical biochemical reaction kinetics state increasing reactant metabolite concentrations result in increased flux through a reaction. Similarly, within a certain range, increasing the concentration of the catalyzing enzyme also increases reaction flux. It is likely that the well described rescue of isoleucine auxotrophy by *avtA* over-expression result from increasing the rate of a limiting reaction. If this is the case, the same result can be affected by modulating metabolite concentrations. We hypothesize that the disruption of this cysteine biosynthesis gene and of the genes responsible for methionine biosynthesis and transport enable the rebalancing the intracellular the amino acid levels towards improving the flux through reactions reducing homocysteine toxicity. As *metB* and *metQ* are disrupted in prototrophic strains, it is feasible that they enable growth without exogenous supply of isoleucine through changing intracellular metabolite pools.

RNA polymerase subunit targets:

Mutations to *rpoA* have been shown to enable broad reprograming of cellular regulation. It is theorized that *rpo* genes are plastic to a wide range of modifications enabling the escape from many different selective pressures(Conrad et al., 2010; Klein-Marcuschamer, Santos, Yu, & Stephanopoulos, 2009). We observed a duplication of the 6bp sequence 'AACATT' in the RNA polymerase alpha subunit in two strains. This mutation was one that differentiated these two strains with strong prototrophic growth, MM43 and MM51, from strains with similar mutational profiles that retained an auxotrophic phenotype (Figure 20 – left panel). These two prototrophic isolates also had disruptive residue change mutations (T322M) to *metB*. It is possible that the combined action of both mutations is necessary for the strong prototrophic growth. However, further experimentation is required to fully elucidate their relative impact of the isolates' phenotype.

Mutation Rate Increases:

An increase in abundance of mutations is observed in 10 of the ΔI isolates. This high mutation rate phenotype is mediated by the inactivation of *mutY* through a 31,084bp genomic deletion. All affected strains demonstrate a bias towards C to A or G to T mutations. This is to be expected as *mutS* encodes a protein responsible for a mismatch repair protein that specifically corrects G-A mispairs(Au, Clark, Miller, & Modrich, 1989). Among these high-mutator strains

there is a subset among which mutational profiles cluster closely together and yet demonstrate both extreme auxotrophic and prototrophic phenotypes (figure 20 – right panel). While there are 22 mutations that are unique to the prototrophic strain, given the sample size it is impossible to differentiate the genes as causative reversion candidates or hitchhiker mutations. Instead we simply highlight that these mutations are enriched functional groups of amino acid transport, regulation and biosynthesis functions. Additionally, the large genomic deletion causing the mutator phenotype is must have some function enrichment as it occurs independently in separate population replicates. To elucidate whether this deletion genotype is being fixed in the population due to the increased mutation rate or because of a fitness effect associated with the deletion of another gene in region (*yggM, ansB, yggN, yggL, trmI, mutY, yggX, mltC, nupG, speC, yqgA, pheV, yghD, yghE, yghF, yghG, pppA, yghJ, glcA, glcB, glcG, glcF, glcE, glcD, glcC and yghO*) requires further investigation.

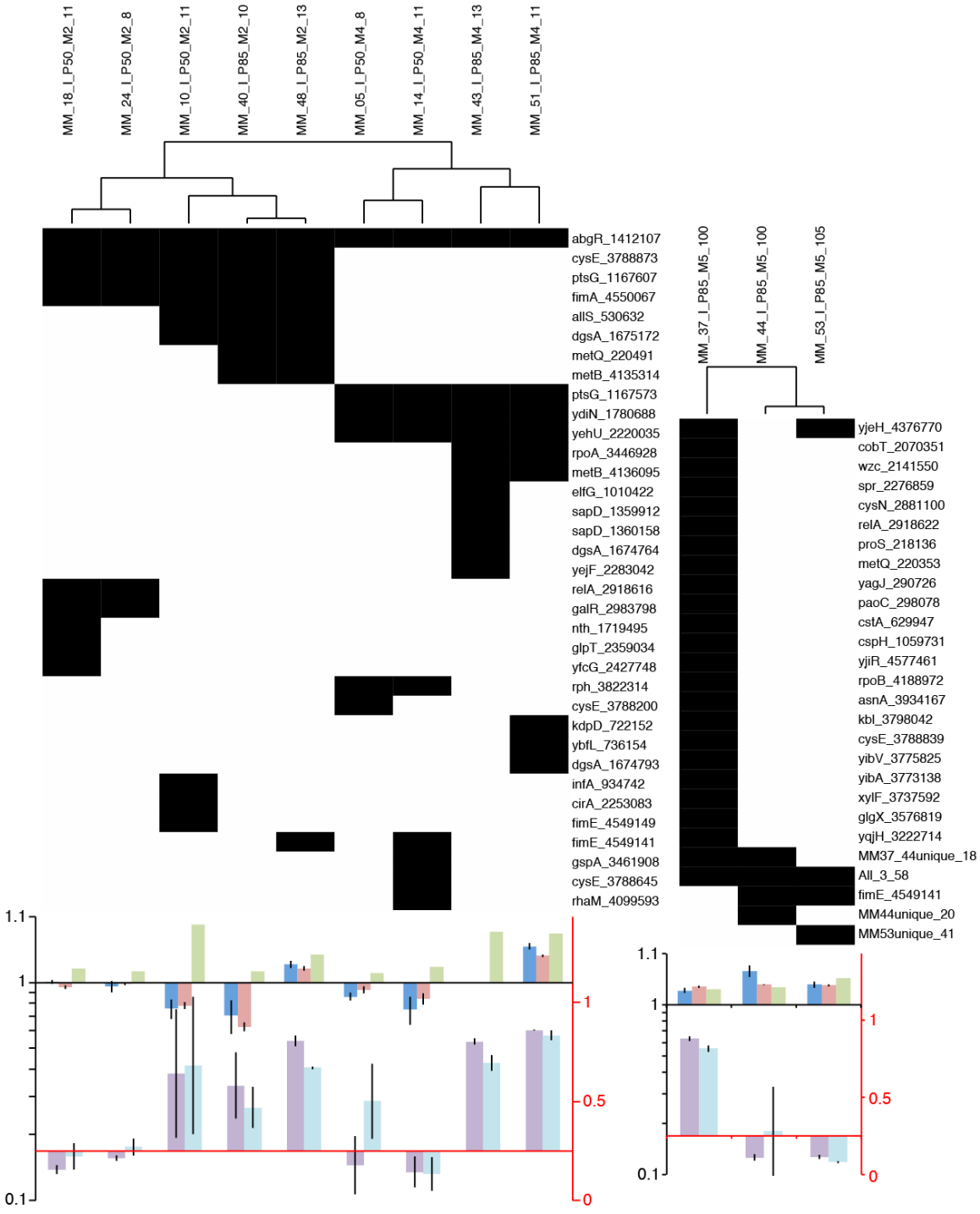

Figure 20: Compiled phenotype and gene mutation data for the Isoleucine strains with most similar mutation profile and distinct phenotypes. Right hand side has the high mutator phenotypes. Top Panel: Heat map of the mutated genes in each strain. Mutated genes present in 3 or more strains. Strains are clustered hierarchically using Euclidian distance and complete linkage. Bottom panel bar graphs: top graphs use left-hand black

axis: Fitness/Malthusian parameter for evolved strain from competition against the ancestral strain at 12hrs in blue and 21hrs in red; growth rate of evolved strain relative to ancestral auxotroph (evolved/WT) in green. Bottom bar graphs use right-hand red axis: OD600 value reached by isolates after 48/96hrs (purple/blue) of growth in M9 minimal media. Values above 0.25 are indicated in the upwards direction, values below 0.25 are indicated in the downward direction. s (perhaps make gene table).

4.3.2.2 Mutations Enriched in Threonine Lineage:

Theonine utilization and transport

The threonine operon (*thrLABC*) is in strongly regulated by its downstream metabolite (threonine) through attenuation in the *thrL* leader sequence. Low levels of threonine and isoleucine in the media result in transcription and translation of the entire operon whereas high levels of these metabolites reduce the translation rate(J. F. Gardner, 1979). To generate the threonine auxotrophic lineage, the *thrC* gene was removed rendering reactions catalyzed by these gene products futile. Under the limiting threonine conditions of the auxotroph there is likely to be high transcription and translation of this operon as the cell tries to no avail to increase the intracellular threonine concentration. This continued effort likely leads to a waste of precious cellular resources that if it could be avoided would be richly rewarded. In four of the sequenced threonine lineage isolates we observed an IS5-mediated insertion 71bp upstream from the transcription start site of *thrL* directly in the promoter region of this leader peptide. This structural variant is likely destroying the activity of this promoter region and therefor stopping any translation of the operon. This modification is likely to both be readily accessed via mutations and result in a sizeable fitness increase as discussed. This is supported by the fact that this modification is seen to arise

independently in multiple different population replicates making it a likely adaptation to growth as a threonine auxotroph.

Another commonly observed endpoint of threonine auxotrophs evolution under cross-feeding is tandem genome duplication mediated up regulation of threonine importers. Three distinct but over lapping genomic regions had higher coverage of mapped reads in 6 of the isolates from this lineage (figure 19c). The regions in increasing length are: ~5-fold coverage increase between *aer* and *yhaL* (37,800bp); ~3- coverage increase between *yqiH* and *agaB* (95,980bp); and ~2-fold coverage increase between *yqiG* and *yrhA* (101,097bp). All three of the regions contain the genes *ygjI* and *sstT* respectively encoding a putative amino acid uptake system(Riley et al., 2006) and a sodium ion coupled serine/threonine importer(Ogawa, Kim, Mizushima, & Tsuchiya, 1998). This indicates that increasing threonine uptake is understandably a critical function that is enriched in these threonine-limited strains.

The two largest regions also contain the *tdc* locus (comprised of *tdcABCDEFG*, *tdcR*, and *yhaO*) responsible for the transport and utilization of threonine and serine. The isolates containing these large duplications like reap even more benefits from further increasing their rate of threonine import. Both *tdcC* a known threonine transporter(Sumantran, Schweizer, & Datta, 1990) and *yhaO* predicted to be involved with threonine transport by sequence similarity to amino acid transporters and positioning with the operon(Riley et al., 2006) are likely contributing to these increased gains. While being a common strategy to

mitigate threonine limitations, these amplifications do have a clear relationship the isolates' grow phenotypes. This suggests that other mutations outside these regions are mostly likely important contributing factors to the strong prototrophic growth phenotypes observed. However within the mutational profile cluster containing to the most robust prototroph (MM36 reaches OD600 of >0.7 by 48hrs) the distinguishing mutations that are comprised of the previously discussed mutation upstream of *pyrE* and IS5 mediated insertion upstream of *thrL*, do not have mechanistic explanations for enabling prototrophic growth (figure 21b). It is possible that in this strain the robust prototrophic growth is enabled by epistatic interactions among the contained mutations.

Exporters and regulator targets:

Four of the 8 ΔT isolates have a mutation loss of function mutation in *fre*, a gene encoding a riboflavin reductase(Coves, Niviere, Eschenbrenner, & Fontecave, 1993). There seems to be a strong functional enrichment for the disruption of this gene as two separate types of disruptions are fixed in this lineage: a frame shift and an IS2-mediated insertion. Through it's flavin reductase activity in the electron transport chain oxidative phosphorylation, this gene is thought to be involved in stress-induced mutagenesis(Al Mamun et al., 2012). The enrichment of mutations in this gene suggests that this stress response pathway is active under threonine-limited conditions but the response may be detrimental to fitness in this contrived engineered genotype. Disrupting of this gene may be a means

to arrest signal propagation in this regulatory pathway and restrict transcriptional modulation to a number of genes,

In the mutation profile cluster comprising the threonine lineage isolates MM17, MM32 and MM54 there are 4 mutations that differentiate the prototroph (M17) the other two. Surprisingly, three of the mutations are associated with membrane proteins: 1bp deletion 182bp upstream of the porin *ompF*; a T6P mutation in the cholate/bile salt exporter *mdtM*; a predicted 67bp intergenic deletion in the *gsp* protein secretion operon. This strain also has an IS2 mediated insertion within the coding region of the predicted transcriptional regulator *yhcF*. There is no literature on these four genes that would suggest a mechanism for reversion to prototrophy therefor further experimentation is required to determine which of these if any is causative.

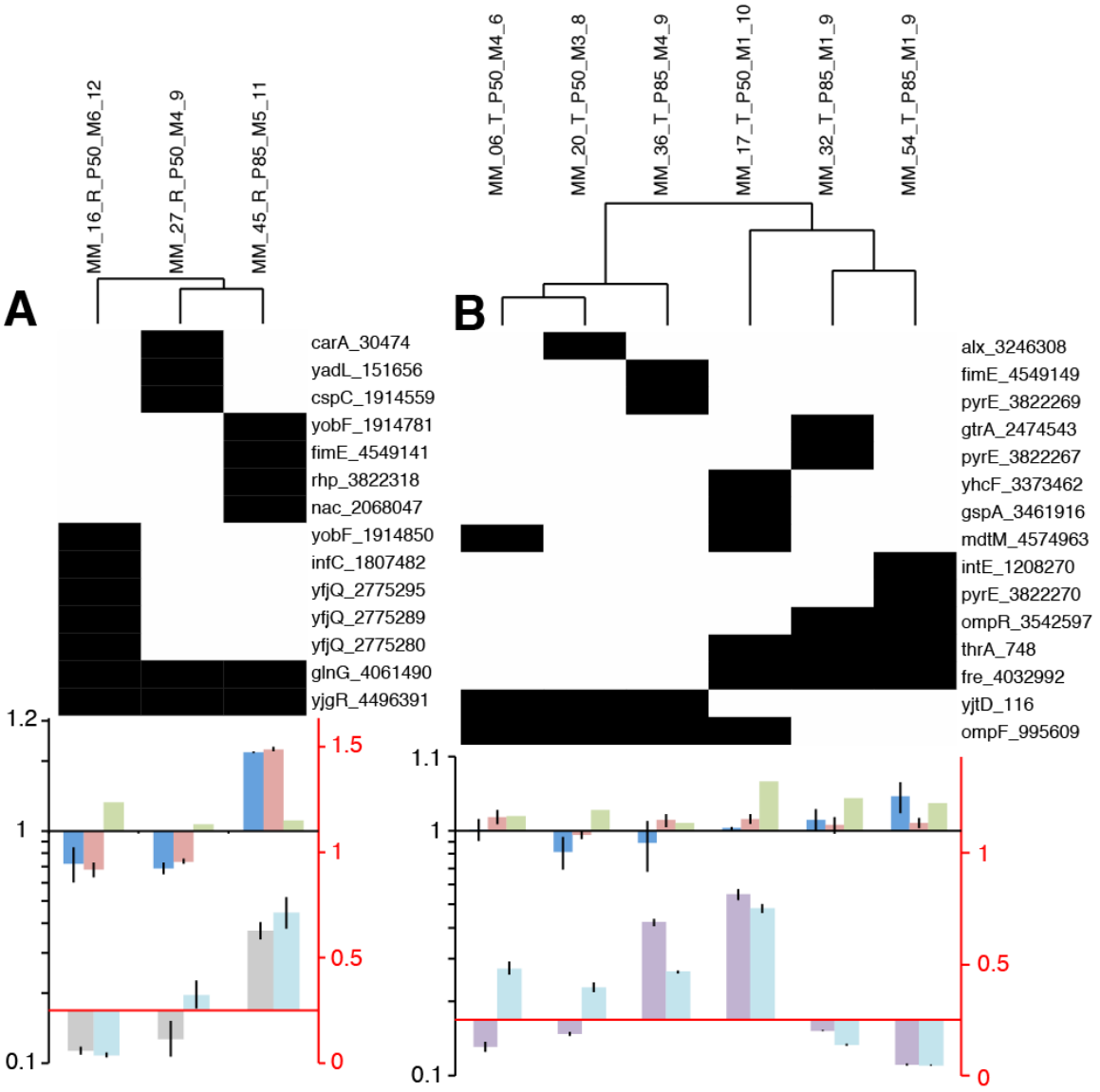

Figure 21: Compiled phenotype and gene mutation data for the arginine (A) and threonine (B) strains with most similar mutation profile and distinct phenotypes. Top Panel: Heat map of the mutated genes in each strain. Mutated genes present in 3 or more strains. Strains are clustered hierarchically using Euclidian distance and complete linkage. Bottom panel bar graphs: top graphs use left-hand black axis: Fitness/Malthusian parameter for evolved strain from competition against the ancestral strain at 12hrs in blue and 21hrs in red; growth rate of evolved strain relative to ancestral auxotroph (evolved/WT) in green. Bottom bar graphs use right-hand red axis: OD600 value reached by isolates after 48/96hrs (purple/blue) of growth in M9 minimal media. Values above 0.25 are indicated in the upwards direction, values below 0.25 are indicated in the downward direction.

4.2.3.3 Mutations Enriched in Arginine Lineage:

Regulator targets:

Six of eight ΔR isolates had IS-element insertions in or near *yobF*, an acid stress related small peptide(Hemm et al., 2010). The enrichment of these structural variants in the arginine auxotrophs leads us to speculate that this may be an adaptation to alleviate stress resulting from the deletion of *argA*. As IS elements are known to modulate expression of genes adjacent to the landing site it's important to consider the potential effects of changes to the expression of adjacent genes. In this case neither *cspC* (stress related antiterminator) nor *rlmA* (methyltransferase acting on 23S rRNA) are directly implicated in alleviating arginine limitation stress(Gustafsson & Persson, 1998; Phadtare & Inouye, 1999). In this case it is much more likely that the acid stress regulation modulation through disruption of *yobF* is what is being functionally enriched in this case.

There is a group of ΔR isolates with closely clustering mutational profiles and divergent phenotypes (figure 21a). The prototrophic strain in this grouping differs from the auxotrophic strain by only a handful of uniquely mutated genes (*rsp/pyrE, ypoB, fimE, nac*). Therefor it is likely that one of these genes enables strain MM45 to reach a high cell density in minimal media within 48 hours and drastically outcompetes the ancestral strain when grown together under arginine supplementation. Both the structural variants at the *rsp/pyre* and the *fimE* locus have been previously discussed and are unlikely to be causative of the reversion

to arginine prototrophy. Both the IS1-mediated disruption of the N-terminus of *ypoB* and the R115C residue substitution in *nac* however are potentially causative of this phenotype. As *ypoB* is not functionally annotated, it is difficult to predict the cause of this mutation. On the other hand, *nac (*Nitrogen Assimilation Control*)* is responsible in a sigma-70 dependent manner for the regulation of ~25 nitrogen assimilation genes through and is intimately linked to growth on nitrogen(Zimmer et al., 2000). *E. coli* is natively able to efficiently use arginine as a sole nitrogen source and has complex regulatory machinery to optimize this process. However, we hypothesize that this mutation to the *nac* gene is able to further modify nitrogen assimilation pathways such that it is able to more efficiently import and utilize arginine enabling it to outcompete the wild type strain as reported by other mutations to this gene(Muse & Bender, 1998). This broad modulation to nitrogen related pathways are also likely implicated in the reversion to prototrophy through metabolic rerouting and/or activation of cryptic arginine synthesis genes(Atkinson, Blauwkamp, Bondarenko, Studitsky, & Ninfa, 2002).

4.2.3.4 Mutations Enriched in Lysine Lineage:

Amino Acid Import Targets:

Distinct genomic amplifications are observed in two lysine isolates (Figure 19d). The larger of the two has ~2-fold higher coverage and spans 143,035bp between the *atoB* and *gtrB* genes. The smaller region, completely contained within the first, also has ~2-fold higher coverage and spans 85,545bp between the *yfbK* and

gtr genes. Both regions contain the lysine/arginine/ ornithine ABC transport system encoded by *hisP*, *hisMQ* and *argT* (respectively the ATP binding component, integral membrane domains and periplasmic binding protein) (ref). When these isolates are competed against their parental strain in minimal media under excess lysine supplementation, no fitness differential is observed. However, the limited lysine environment experienced in the cross-feeding community may pose a sufficient selection for improving lysine import through the duplication of this region.

4.3 Conclusions:

This study has identified several functional modifications that clearly reinforce the greedy nature of evolution despite the cooperative context. While growth in this cross feeding community is initially uniquely dependent on exchange within the community, the partially syntrophic nature of the exchange may allow greater exploration of selfish mutations. Genomic duplication seems to be one such greedy adaptive mutation that enables improved scavenging of the limited amino acid from the environment. These duplications may be an easily access mechanism for amino acid transport regulation and could potentially explain the early observation of divergence from reproducible dynamics of the community. Interestingly import augmentation mutations in the isoleucine lineage are noticeably absent. Instead these strain have a preponderance towards prototrophy reversion suggesting it may be an easier to access

genotype/phenotype. The multitude of mutations affecting amino acid synthesis and catabolism pathways suggest that some amount of rewiring of central metabolism is occurring in response to long terms growth in the cross-feeding environment. However, a lack of a clear signal for a causal mutation for prototrophy reversion suggests that it is likely to be a highly epistatic effect. In general these observation reinforce the fantastic genomic plasticity of bacterial species to functional disruptions.

4.4 Methods

4.4.1 General Methods:

Strain Construction, kinetic assays, fitness competitions, qPCR quantification, MASC-PCR and strain washing for minimal media inoculation are all described in methods sections in the previous chapter of this document.

4.4.2 Sequencing Library Prep and Analysis:

DNA of each isolate was purified using a genomic purification kit (GE Prokaryotic Genome kits). DNA shearing and addition of Illumina sequencing adapters and barcodes was performed enzymatically (Nextera). Samples were pooled in two groups of 30 and 29 strains (the parental strain, EcNR1 was also included) and each sequenced as single-end 100bp reads on a lane of an Illumina HiSeq. Genomic sequences were aligned to the reference genome of the parental strain EcNR1, a variant of *Escherichia coli* MG1655, and mutation calls were made using the breseq computational pipeline(Deatherage & Barrick, 2014). This

package is capable of performing split-read mapping enabling the identification of structural variants without using paired-end reads(Barrick et al., 2014). Hierarchal clustering using Euclidian distance and complete linkage was performed on the mutation matrix of the strains using Cluster3.0 and visualized with TreeView.

5 Engineered metabolite exchange in synthetic microbial ecosystems

5.1 Overview

Our ability to engineer bacterial genomes has matured to the point that we are able to engineer complex regulatory and bioproduction phenotypes. Scaling synthetic biology to the level of bacterial communities is a natural next step. Population level behaviors emerge from the combination of all involved individual strains. Maintaining robust community behaviors when transplanting to the milieu of complex microbial ecosystems remains a challenging task. Understanding how to best tune metabolitic transfer between community members will be critical to engineering community level behaviors.

Bacteria have evolved multiple strategies to control and direct mass/energy exchange across the membrane. High-throughput sequencing is continually discovering more trans-membrane transporters although in many cases the critical parameters of specificity and kinetics remain poorly annotated. Some bacteria are also able to develop direct cytosolic linkages, a costly appendage most likely paid for by mitigating diffusive losses of the exchanged metabolite. Cytochromes, enzyme secretion & nano-wires provide additional capabilities for bacteria to make use of the resources surrounding them(Benomar et al., 2015; Pande et al., 2015). Here we demonstrate that endowing synthetic microbial communities with amino acid secretion enables increased productivity. However, this strategy doesn’t come without risks. Increasing environmental availability of metabolites can result in the proliferation of cheaters. Therefor it is

critical that the to reap benefits of exchange that bacteria are able to direct exchange to strains that are able to return the favor.

5.2 Results:

5.2.1 Improving co-culture productivity through export:

It has been shown that auxotrophic bacteria unable to grow independently can grow when paired with a strain auxotrophic for a different metabolite(Mee & Wang, 2012; Wintermute & Silver, 2010a). For robust growth to occur, these mutualistic systems require exchange of the requisite metabolites. If rate of diffusion across the membrane is the main limitation for exchange of these metabolites, the degree of mutualistic growth would likely correlate to the hydrophobicity or membrane permeability of the involved metabolites. For the most part this holds and we see that highly hydrophobic metabolites demonstrate robust mutualistic growth. However, some highly hydrophilic amino acids (K & R) are efficiently exchanged despite physiochemical limitations to diffusion across the lipid membrane(Chakrabarti & Deamer, 1992; Monera, Sereda, Zhou, Kay, & Hodges, 1995). This suggests host expression of transmembrane transporters may play a significant role in increasing amino acid exchange. Engineering bacterial amino acid transport will likely have a profound effect on the productivity and dynamics of cross-feeding systems.

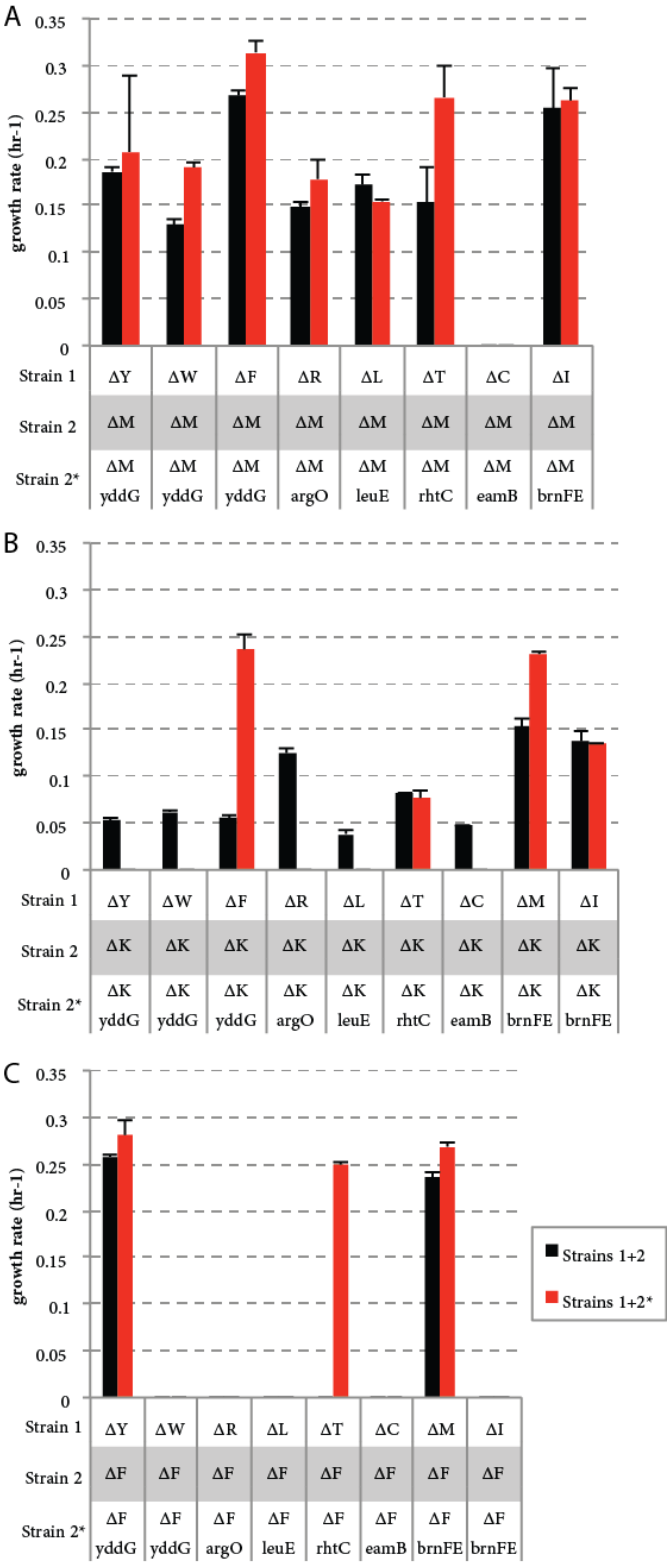

Figure 22: Observed syntrophic growth rate modulation with exports. Error bars represent standard deviation of 3 replicates.

Very few amino acid exporters are fully characterized relative to the abundance of importers. Through a literature search we identified 6 exporters that had putative amino acid export function for 9 amino acids (Table 3)(Kutukova et al., 2005; Nandineni & Gowrishankar, 2004; Trotschel, Deutenberg, Bathe, Burkovski, & Kramer, 2005; Yamada et al., 2006; Zakataeva et al., 1999). The genes were inserted into medium copy episomal vectors (p15A origin) under arabinose induction and cloned into each of the three auxotrophic strains with the highest syntrophic index (M,K and F)(Lutz & Bujard, 1997; Mee et al., 2014). The selection of the host strain's auxotrophies ensures an increased likelihood that any partnered strain will be able to provide the required amino acid. Each exporter-strain combination was inoculated together with a complimentary E. coli strain auxotrophic for the predicted amino acid exported. Pairings were grown in replicate in microtiter plates alongside a control pairing without any exporter. In order to equilibrate the fitness costs associated with induced protein overexpression, all non-exporting strains contained an identical episomal vector with the exporter swapped for a fluorescent marker (mCherry or GFP). Three of the exporters (*argO, rhtC* and *yddG*) significantly increased the productivity of the synthetic cross-feeding community in at least one of the tested pairings (figure 22). This initial screen identified export modulation as a viable mechanism for tuning the dynamics of these microbial communities. Analytical HPLC was performed to further confirm that engineered secretion resulted in increased extracellular amino acid concentrations (table 3 and figure 23). The

three exporters identified via the cross-feeding community growth screen were also the three exporters that demonstrated the greatest increases to extracellular concentration of their respective amino acids. Both *argO* and *rhtC* proved to be quite specific in that their induction didn't increase the presence of other amino acids in the media. Interestingly, *yddG* has a broad effect on changing extracellular amino acid concentrations relative to wild type. While it greatly increases the presence of tyrosine in the media (>9e6 fold over wild type), it also resulted in increases to the extracellular serine concentrations (>2e6 fold over wild type). Increases to the observed extracellular concentrations of phenylalanine and tryptophan were roughly 5-fold. Interestingly, despite the export bias of aromatic amino acid export towards tyrosine, *yddG* has the greatest impact on improving coculture growth when combined with phenylalanine auxotrophic strains (figure 22). This disparity suggests that import limitations may be equally impactful in the cross-feeding pairings of aromatic amino acid auxotrophs. The broad spectrum of increased export associated with *yddG* will also impact the metabolic cost of its induction.

Gene	Predicted Exported Amino Acid	Position in parental genome	Fold increase in export (HPLC)
argO	R	3,074,627...3,075,262	3.57E+06
brnFE	I M	*From Corynebacterium glutamicum	1.45E+00 1.7E+00
eamB	C	2,721,877…2,722,464	1.00E+00
leuE	L	1,887,353…1,887,991	1.00E+00
rhtC	T	4,014,212…4,014,832	5.69E+05
yddG	F W Y	1,553,520…1,554,401	4.88E+00 4.94E+00 9.16E+6

Table 3: Amino Acid Exporters Identified From Literature Search

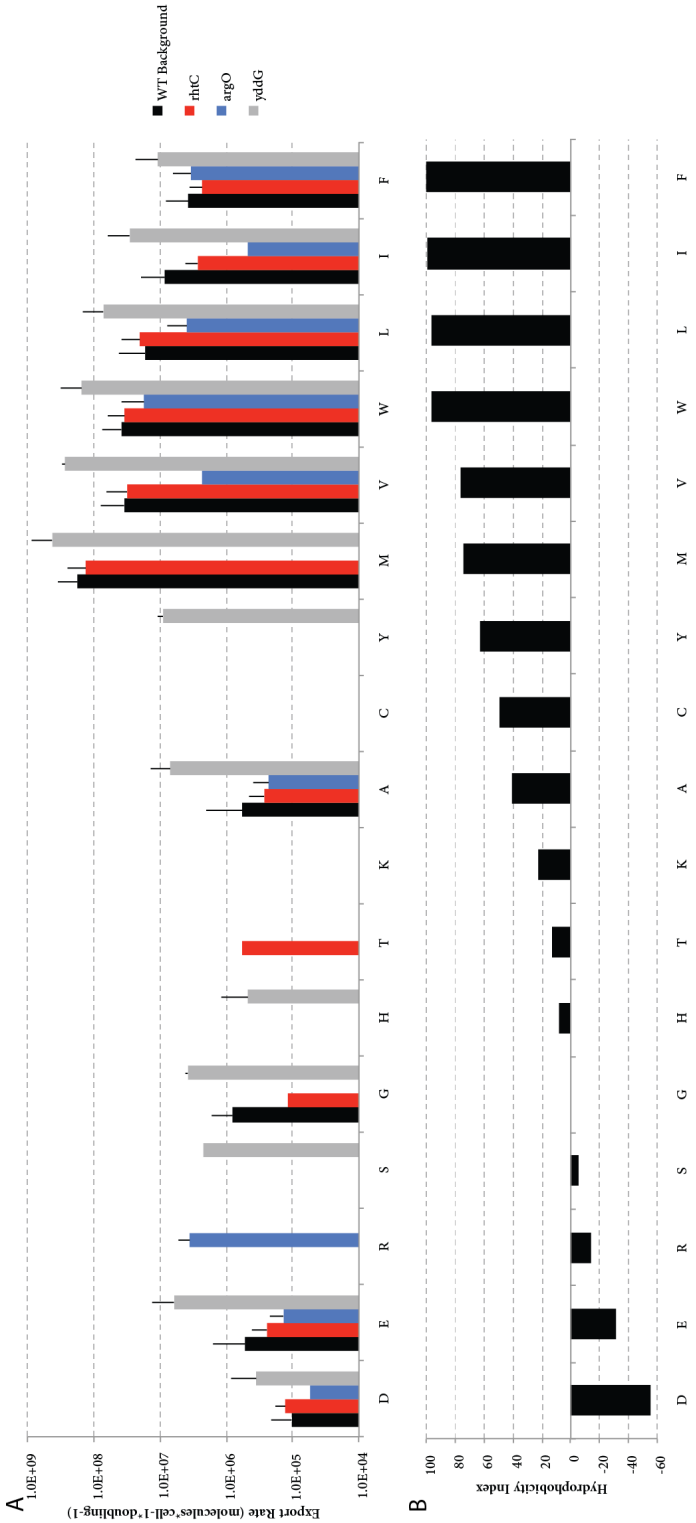

Figure 23: HPLC analysis of the top 3 exporters: argO, rhtC and yddG. A) Observed export rate calculated from the HPLC determined concentration, the total number of cells present

over the growth period and duration of exponential growth prior to analyzing the growth filtrate. B) Hydrophobicity index of all amino acids highlights high secretion from all cells for highly hydrophobic cell types and the critical nature of export induction for R and T.

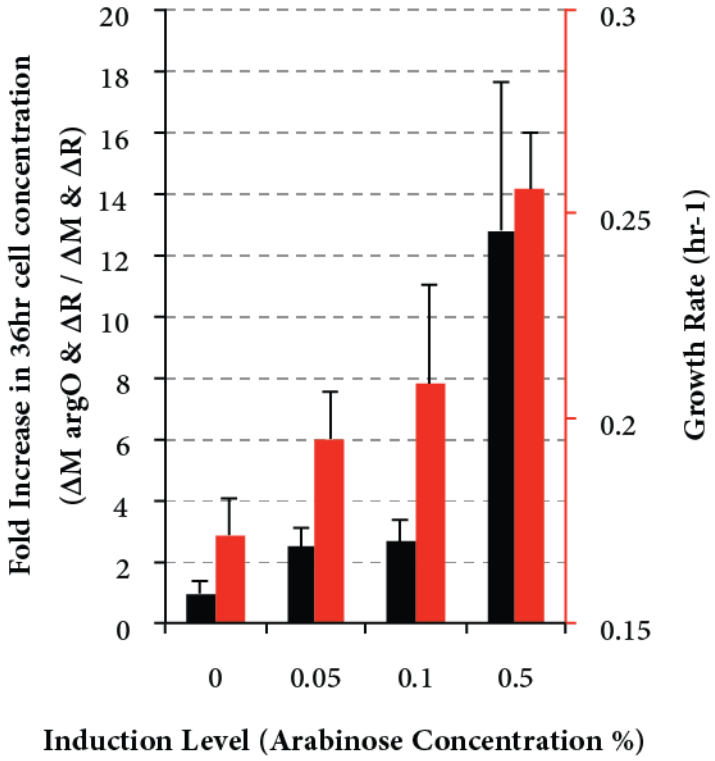

Figure 24: Increased Export Induction Leads to Increased Community Productivity. This assay is performed for the ΔM-ΔR pairing with argO induction. Error Bars Represent Standard Deviation of 3 replicates.

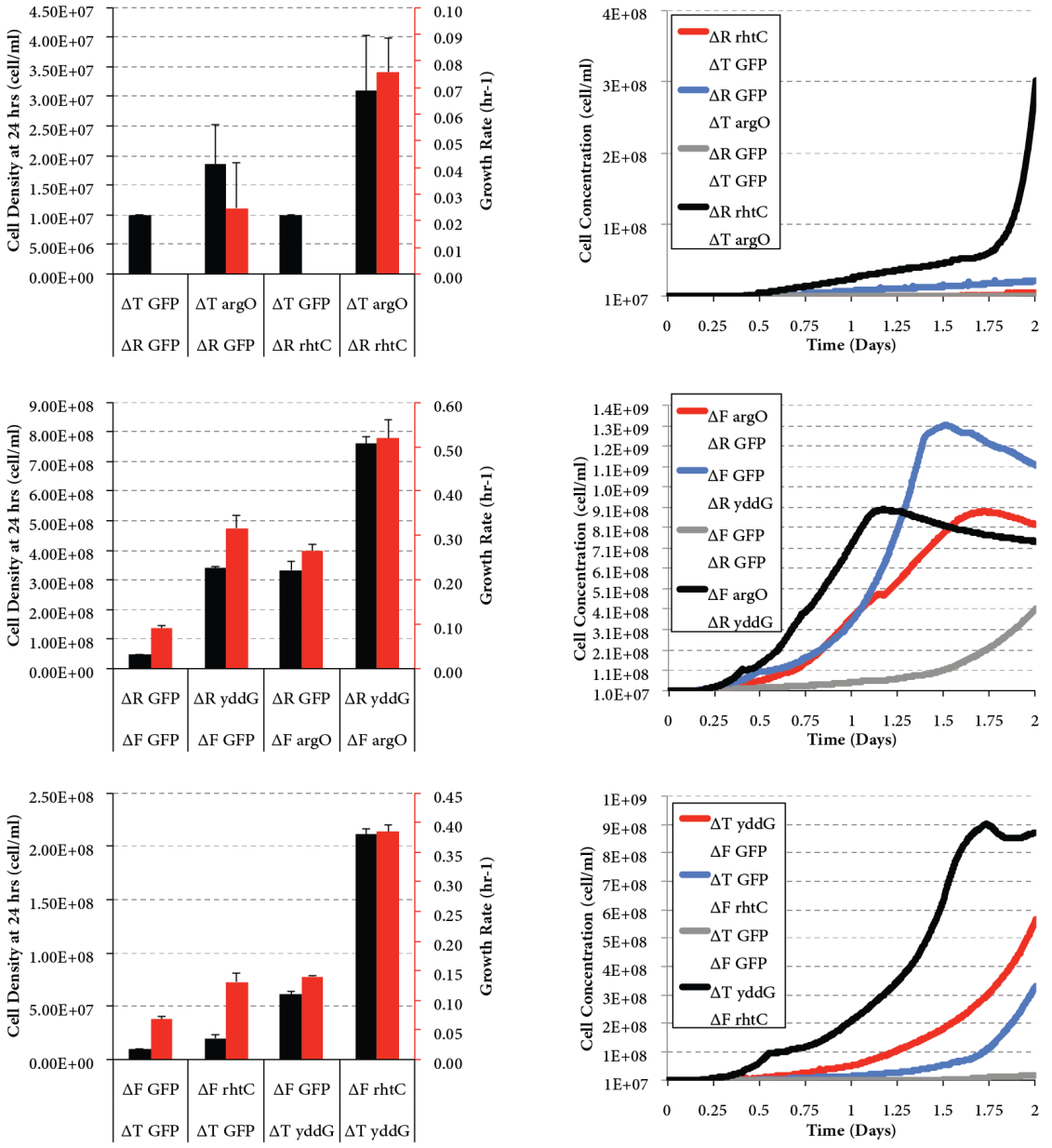

Figure 25: Synergistic Effect Of Combining Exporters In Syntrophic Pairs. Left Column – in black: Community density attained after 36 hours of growth plotted against the black, left hand vertical axis. – in red: Community growth rate plotted against the red, right hand vertical axis; Right Column: Representative growth curves for a single replicate of every synergistic pairing. Black line – Growth with 2 exporters; Red/Blue line – growth with 1 exporter; Grey line – growth with no exporters.

The top three amino acid exporters were further analyzed to understand the parameter space for expression and synergism where these exporters were

able to modulate the growth of these communities. Increased export induction resulted in increased growth rate of the synergistic pairing (figure 24). When comparing all three pairings of these exporters to non-exporting and single-exporting pairs a couple general trends emerged (figure 25). First, at a set expression level the *argO* and *yddG* exporters had a greater individual effect on the community productivity than *rhtC*. Second, combining complimentary exporting strains resulted in increases to growth rates and attained cell densities above that of the non-exporting and single exporting pairs. The magnitude of these increases however differed between strain combinations. In the ΔR-ΔT pairing, export of T alone had no effect on the productivity of the community however when combined with R export the synergistic effect was greater than two fold higher than the productivity improvement of R export alone. Combination of R and aromatic amino acid export in the R-F pairing resulted in an increase of productivity that was additive of the improvements resulting for each single exporter pairings. Combination of the *rhtC* and *yddG* exporters resulted in an effect size that was greater than a multiplicative combination of each individual exporter's effect on the ΔF-ΔT pairing. These results highlight the interesting result that on it's own *rhtC* has a relatively minimal effect on improving community productivity but when paired with a strain that also contains an exporter, that improvement is greatly magnified.

There are intriguing differences between community productivity levels under *rhtC* induction in a host dependent manner: in ΔM and ΔF *rhtC* causes

robust growth whereas in ΔR and ΔK there is very little improvement. This indicates that there is some sort of strain dependent multiplicative effect that is happening. In ΔM the improvement is likely due to the branched M/T biosynthesis pathway where the ΔM strain has an increased intracellular T pool that the exporter can act on. However, the same reasoning doesn't apply to the other discrepancies. Some hypotheses on the reasons behind this include: higher absolute levels of T requirements relative to marginal increases from *rhtC* induction; auxotrophy specify regulation changes resulting shifting intracellular metabolite concentrations – but further testing would be required to conclusively determine the root cause of this behavior.

5.2.2 Exchange costs and the potential for cheaters:

Within the three selected exporters, the effect size of the improvement in syntrophic pairings with a single exporter single exporter correlates with the biosynthetic cost of the amino acid involved. As previously shown this translates to an inverse relationship to the quantity of amino acid required for growth of these auxotrophic strains. So increasing the extracellular pool for an amino acid that is required in relatively lower abundance results in a much greater increase in productivity of the community. This increase in co-culture productivity occurs despite the tradeoff that exists in exporting strains where pumping out increasingly biosynthetically expensive metabolites results in lower fitness. To quantify this fitness cost for auxotrophic strains expressing the *argO*, *rhtC* and

yddG exporters competition assays were performed. The metric used to quantify the relative fitness of strains is the ratio of the number of doublings of the analyzed strains. Auxotrophic exporting strains were competed against non-exporting strains of identical auxotrophy in saturated amino acid supplemented minimal media (figure 26). As predicted, export of more expensive amino acids resulted in a greater fitness reduction with a fitness decrease rank of: rhtC < argO < yddG. If exporters existed for every amino acid this exchange/fitness tradeoff in proportion to the biosynthetic cost would likely hold. Flux balance analysis (FBA) modeling is used to extend the investigation of this relationship to amino acids without known exporters. Monitoring the effect on predicted flux to biomass while changing the constraints on the secretion of each amino acid reveals that the relationship of growth rate reduction to increasing secretion rate holds *in silico*. Biosynthetically expensive amino acids have a greater modeled reduction of flux to biomass than less costly ones (figure 27). The *in vitro* engineered syntrophic communities ensure the recapture the cost of cooperative exchange by design through cross-feeding with a mutualistic partner. However, the high fitness cost associated with the exchange of these metabolites introduces the potential for strains to evolve into or be outcompeted by competing “cheater” strains(Gore et al., 2009). In this context, cheaters are individuals that reduce their contribution of expensive amino acids to the environment or media while continuing to reap the benefits of the provided amino acids by the exporting the amino acid for which they are auxotrophic.

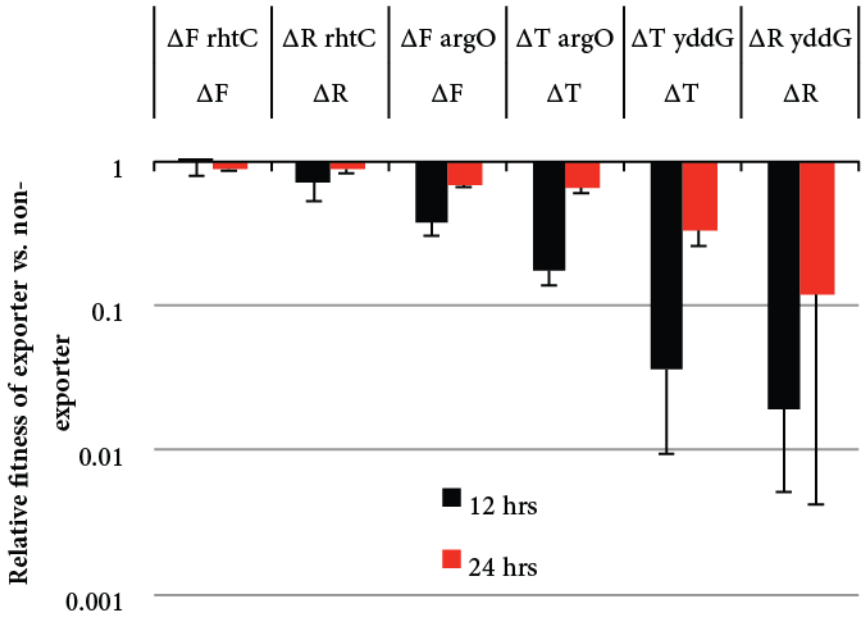

Figure 26: Fitness Impact of Amino Acid Import. Relative fitness is calculated as the ratio of the number of doublings of the exporter vs. the non-exporter.

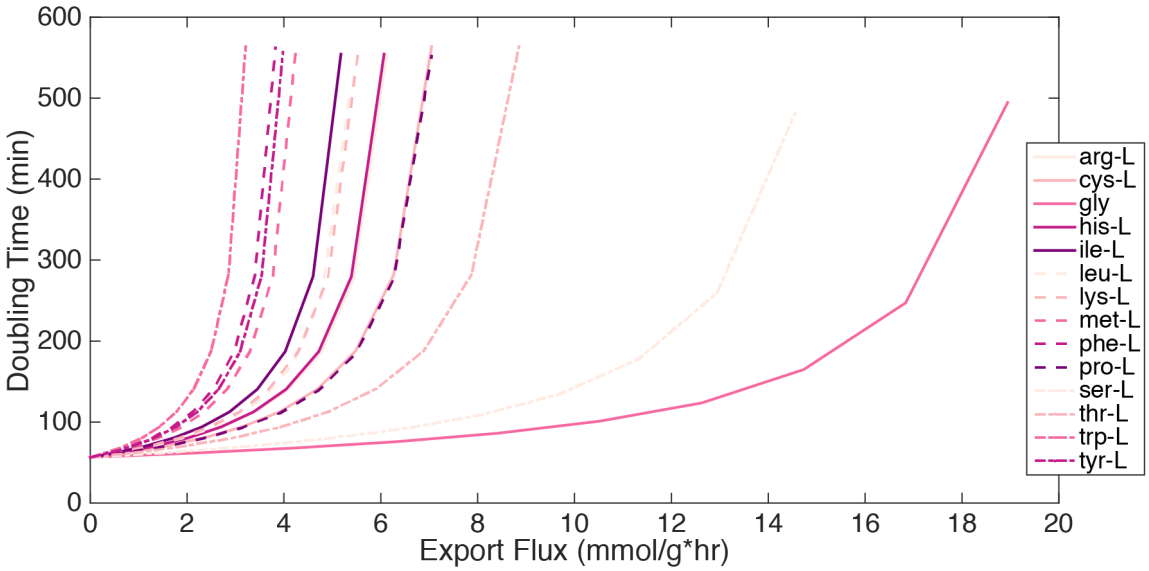

Figure 27: *In silico* determination of cost of export. FBA is used to determine the tradeoff that must occur between secreting amino acid at a set rate (x-axis) and flux of metabolites towards biomass production, here represented as doubling time (y-axis). The greater the more quickly the derivative of the line increases, the more biosynthetically costly the metabolite. Tryptophan is predicted to be the most expensive and glycine the lease.

Many transport processes result in the expense of energy beyond that associated with the loss of the secreted metabolite through dependence on ATP or changing chemical gradients. We sought to quantify whether induction of exporters resulted in changes in the efficiency of conversion of glucose to biomass. As previously discussed, comparing growth saturation levels over a titration of a metabolite concentration allows the cellular requirement for that metabolite to be quantified. Here, glucose concentrations were titrated and growth saturation levels for exporting and non-exporting strains were examined (figure 28). While there was no observable difference between induced and uninduced expression of the *argO* and *rhtC* exporters, there was a significant shift in the glucose requirements under *yddG* expression. Cells growing with induction of *yddG* require roughly 3 times more glucose molecules to produce a given cell. This decrease in conversion efficiency increases both the benefit an individual would derive from adapting to become a cheater (reduce the amino acid secretion caused by *yddG* expression) and likelihood such strategies would arise in the population.

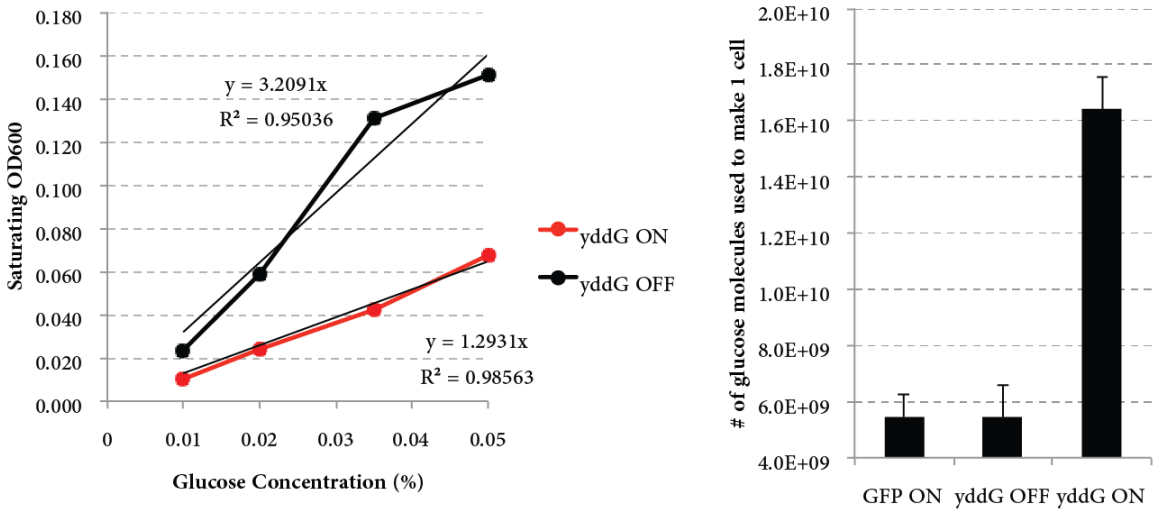

Figure 28: Impact of yddG induction on the efficiency of glucose utilization. Left panel: Shift in saturating OD levels as glucose concentration is titrated. The slope is used as previously discussed to calculate the number of glucose molecules required to produce a given cell. Right panel: yddG induction increases the glucose requirements by ~3 fold. GFP induction, and rhtC and argO (data not shown but similar to GFP on) do not increase the glucose requirements.

5.2.3 Competing cooperative vs. cheating strategies in mutualistic communities:

To investigate the ability of cooperative exchange phenotypes to persist when confronted with cheating phenotypes, a triplet community was assembled consisting of two exporting cross-feeding strains (A and B) and a third cheater strain (A* or B*) with one of the two comprised auxotrophies but without any export. In this complex three-way interaction, strain A (exporter) and A' (cheater) both derive resources pumped into the environment by strain B (exporter). Strain B acts as a dynamic buffer to the community, providing a distinct environment to that encountered under the previously discussed competition in saturating levels of amino acid. In this context, competition is likely to be occurring in very low amino acids concentrations as the strains consume the amino acid as soon as it

is secreted into the environment. The rate of amino acid consumption by A and A' is proportional to their relative abundance in the population. We hypothesize that feedback exists in this system in that as A is out-competed by the more fit cheater strain A', B will receive less and less of it's required metabolite. This will result in reduced growth of B and a subsequent reduction in the secretion of the metabolite required by A and A'. Whether this feedback is sufficient to maintain A in the population is dependent on relative benefits/requirements of the exchanged metabolites and the ability of A'-B to grow robustly without A. Measuring the relative abundance of the strains in these triplet communities over time allowed us to quantify the strength of this feedback and it's ability to maintain cooperative phenotypes in the community (methods). (figure 29 & figure 30)

When cross-feeding with a dynamic buffer strain, cooperator strains are outcompeted to a lesser degree than when competing under saturated amino acid conditions in the cases of the *yddG* and *argO* exporters. Induction of *rhtC* seems to have a greater fitness impact in the context of cross-feeding than under saturating levels of F or R. In most cases competing under cross-feeding conditions only changes the degree or amount by which the cooperator is outcompeted by the cheater strain and not which strain becomes dominant. Except for the case of expression of *argO* in the ΔF strain cross-feeding with the ΔR *yddG* strain. In this case the exporting strain is of more or less equally fit with the non-exporting strain. The order of the fitness costs associated with exporting

T, R and F in the cross feeding context is the same as in saturated amino acid conditions. However, under cross-feeding the fitness cost associated with expressing *rhtC* and *argO* is much more similar than under saturated conditions. This indicates that there is likely complex strains and exporter dependent amino acid limitation effect on the fitness cost of export induction.

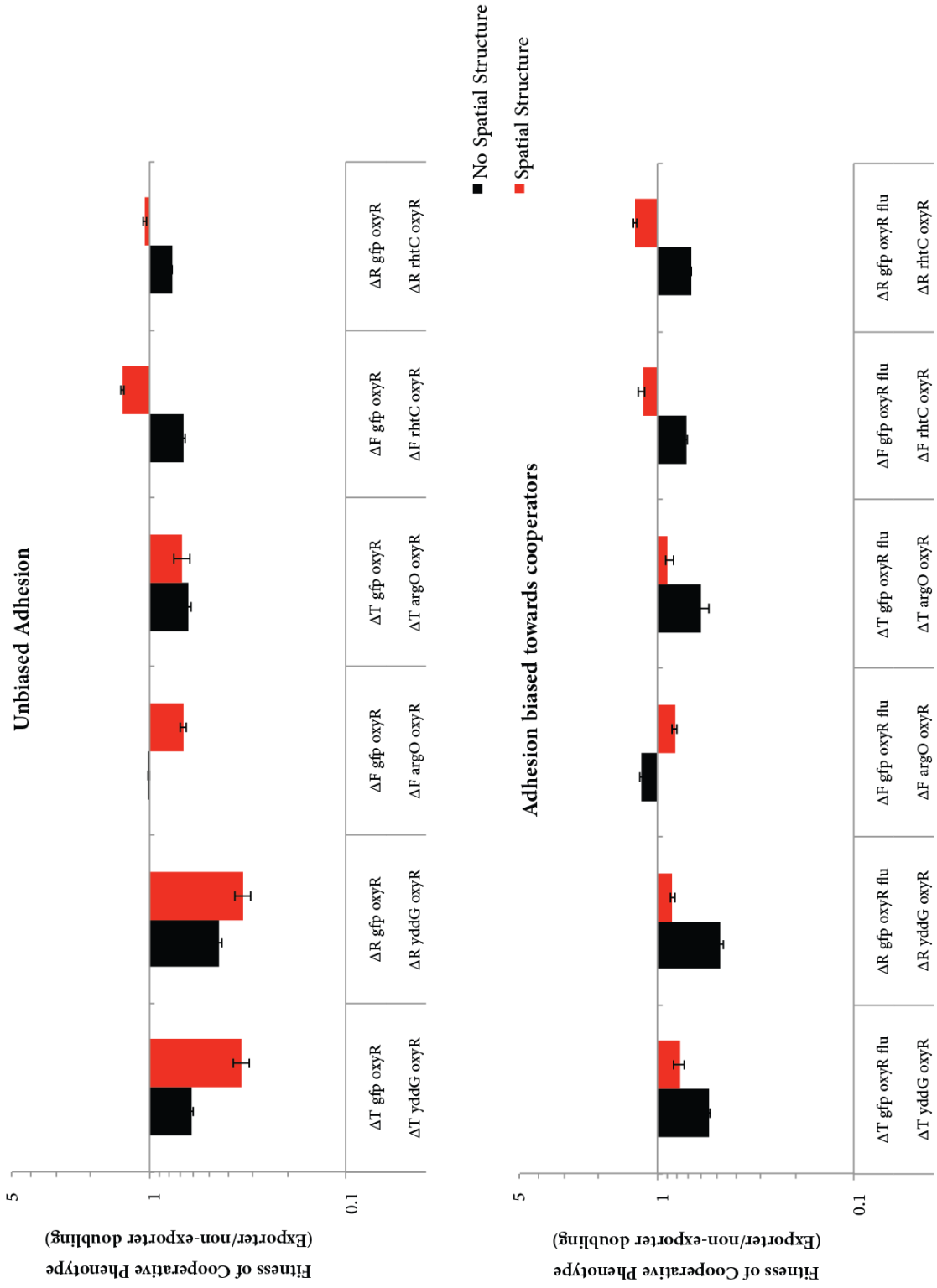

Figure 29: The effect of spatial structure on the relative fitness of exporters vs. non-exporters. Top Panel: Mixtures of strain where all members had oxyR disabled were

analyzed in both static cultures and well mixed cultures. In these conditions all strains are equally likely to form aggregates. Bottom Panel: Adhesion is biased towards the cooperative phenotypes by disabling the flu gene in the cheater strain thereby making it less likely to flocculate.

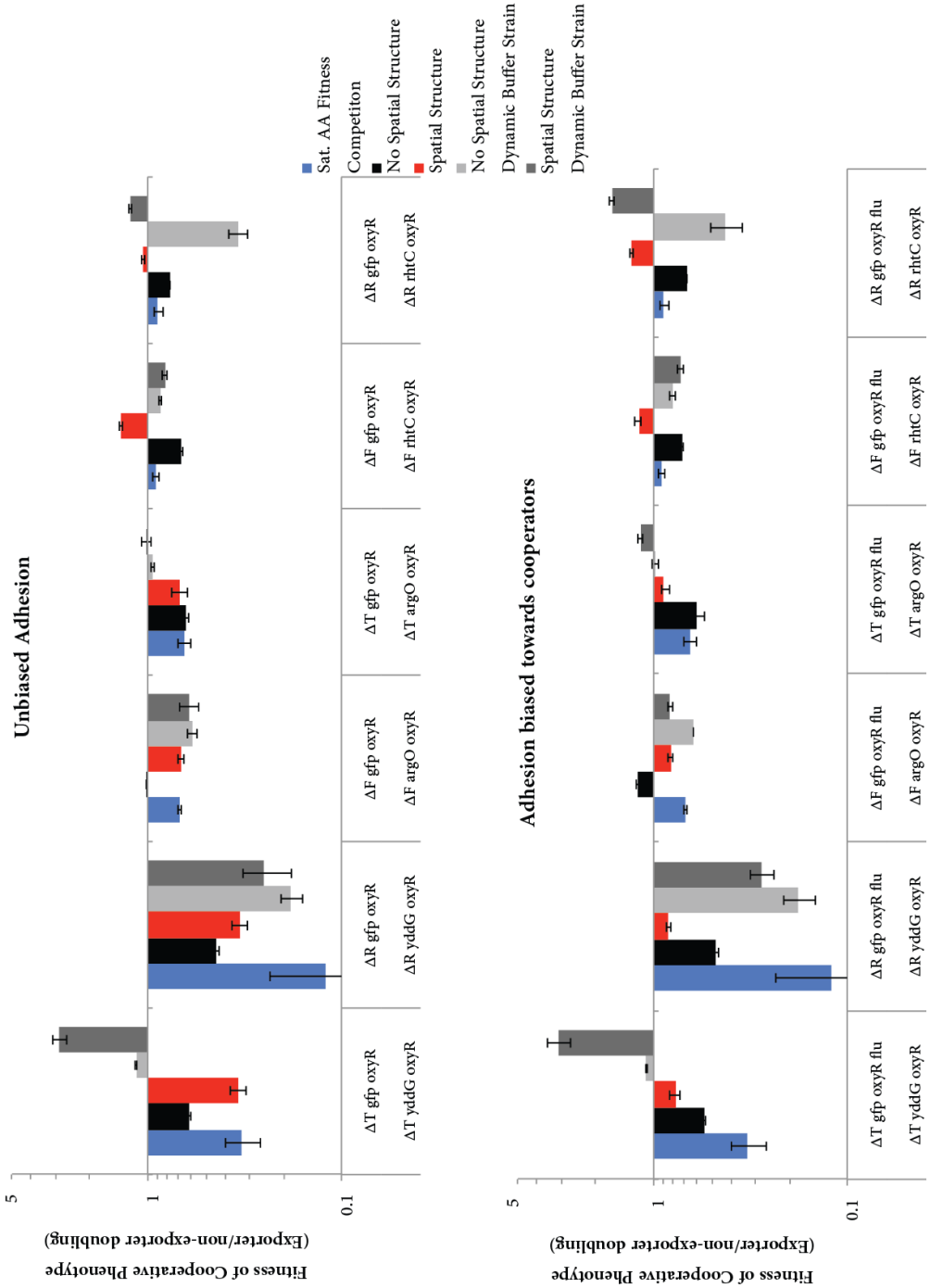

Figure 30: Relative abundance of the dynamic buffer strain in biased and unbiased adhesion fitness competitions.

5.2.4 Engineered community spatial structure encourages cooperativity

While microbial species often exist in a planktonic form, even in the vast oceans spatial structure of is observed in microbial communities as they colonize nutrient substrates. These structures impact competition between cooperative and cheating phenotypes(Cordero, Ventouras, DeLong, & Polz, 2012). The evolution and emergence of cooperative behaviors in nature is a highly studied area with multiple theories on which parameters are most critical(Nowak, 2006). Spatial structure is also commonly observed within microbial ecosystems in digestive tracks of animals and other biofilms. The maintenance of this spatial structure is commonly cited as a critical as strategy to direct the benefits of cooperativity towards you own kin and allow for the evolution of the many cooperative behavior evident in our biosphere(Chuang, Rivoire, & Leibler, 2009; Nadell et al., 2010).

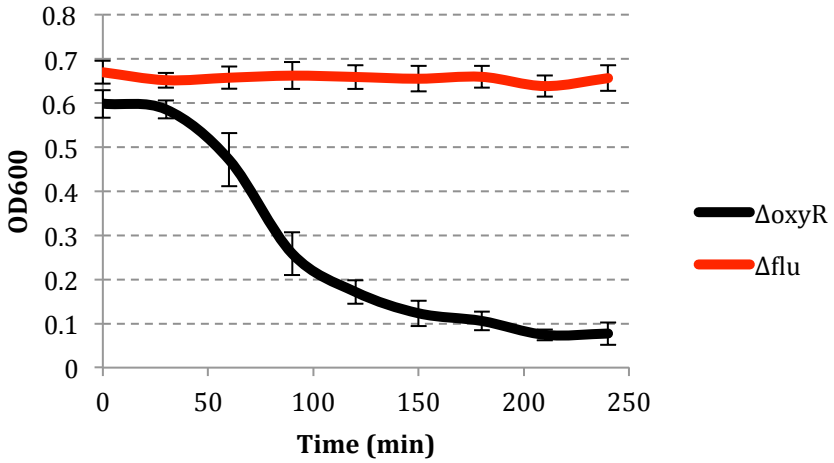

Figure 31: Impact of inhibiting the repression AG43 through deletion of oxyR. Here cells were grown to late log phase in 3ml of supplemented minimal media and then left undisturbed to monitor the rate of flocculation. 200ul samples were take from the top ~0.5 cm of growth tube every 30 minutes.

Inspired by these natural strategies, we sought to investigate the ability of spatial structure to bias the benefits of exporter-improved exchange towards the strains incurring the exporting cost. The repressor of the surface adhesin ag43: *oxyR* was knocked out in all involved strains resulting in constitutive expression of the adhesion causing surface protein(Hasman, Chakraborty, & Klemm, 1999; Johnson, Clabots, & Rosen, 2006). This genetic perturbation has previously been show to increase the rate of flocculation and settling of bacterial strains. This effect of the *oxyR* knockout was experimentally confirmed (figure 31). We additionally generated cheater strains lacking the ability to produce the adhesin ag43 by disabling the gene responsible for it's production, *flu*. With these engineered strains in hand we were able to further probe the effect of spatial structure on the fitness costs of exporting amino acids in cross-feeding systems. Specifically, the relative abundance of strains in two combinations is compared: when all were equally able to flocculate and when only the cooperative strains were endowed with increased aggregation. This growth was then performed under two conditions, mixed and still, to respectively disrupt or allow the formation of spatially structured bacterial aggregates. As expected, exporter and cost of secreted metabolite stratified the effect of spatial structure on the competition between cooperative and cheating phenotypes.

yddG

Interestingly, in the triplet aggregation combinations involving yddG the increased aggregation resulted in enrichment of the cheater strain relative to the

well-mixed environment. Specifically in the ΔT-ΔF pairing, increased aggregation also resulted in a significant increase the proportion of the population comprised of dynamic buffer ΔF-rhtC strain relative to that present in the well-mixed environments (figure 29). Excluding the cheating strains from the aggregates in communities involving *yddG* mediated exchange resulted in a substantial improvement to the performance of the ΔT/ΔR-yddG strains. However it wasn't enough of an improvement to enable the cooperative strain to be more abundant in the population than the cheater.

argO

Whether spatial structure provided fitness benefits to the cooperating *argO* strain depended on the strain it was in. In ΔF it decreased the fitness of the cooperator whereas in ΔT there was a marginal improvement of the cooperative strain's fitness. In the arginine (*argO*) cooperative/cheating systems, precluding the cheater from the aggregate improved the performance of the cooperator relative to when the cheater was able to participate in spatially structured growth. The benefit for the cooperator is greatest in the context of the ΔT strain when paired with the ΔR *rhtC* dynamic buffer.

rhtC

The expression of *rhtC* generally has the lowest fitness impact on the expressing strain. When provided spatial structure to capture the benefits of cooperative behavior, the *rhtC* expressing strains were the only ones that displayed enrichment over the complimentary cheating strain. It is likely the low

fitness cost starting point was critical for the success of these strains in competition with the cheater meaning the marginal fitness cost is easiest to overcome. This however isn't always the case as the ΔF *argO* strain also displays a low fitness cost of sharing arginine yet when provided spatial structure isn't able to overcome competition from a cheater. The ΔR-ΔT pairing also displays a great shift in the relative abundance of the dynamic buffer strain between the disrupted and spatially stabilized growth suggesting that perhaps spatial structure is more impactful on the buffer ΔT-argO.

Combined conclusions:

The intermediate fitness cost of argO results in highly contextually dependent effects on the enrichment of cooperativity. The low fitness cost associated with rhtC results in improvements for the cooperator with spatial structure in a very non-specific manner in that it occurs regardless of the presence or lack of the cheater strain in the aggregate. The high costs associated with yddG expression can be to a great degree recaptured within a mutualistic community if any cheaters can be excluded. This doesn't bode well for the exclusion of spontaneous mutation of cheaters in these systems. Further work should be performed on the R- and F-systems to understand the intricacies of what is occurring and why directed exchange can't allow for the cooperative phenotype to fully overtake the cheaters.

5.2.5 A question of degree? – Mutualism along cooperativity gradients:

The previous experiments investigated competition between the two extremes of cooperativity: high exporter and non-exporter. In nature while mobile element insertions, frameshifts, and other debilitating SNPs can totally ablate metabolite secretion ability, it is also possible to have members of the community that exhibit varying degrees of cooperativity. Such intermediate levels can be achieved by genomic changes affecting the kinetics of export such as mutations resulting in altered regulation of copy number of a gene or the less likely case of changes to the protein that modulate kinetics while maintaining proper function. As previously discussed, modulation of export as well as introduction of auxotrophies can change the efficiency for biosynthesis for certain metabolites from glucose. Such differences are likely to hold for various growth substrates. This metabolic exchange is essential to allow species to specialize their biosynthetic capabilities. Therefor, it is possible that within natural microbial communities the production efficiencies and trade of resources is delicately balanced to ensure that as the economic theory of comparative advantage predicts, trade between cells with varying production efficiency of metabolites can lead to improved productivity of the community as a whole(Enyeart, Simpson, & Ellington, 2015). Therefor, it is likely that neither of the extremes were in fact optimal for the productivity of the system.

There is a critical tradeoff between an individual strains growth and metabolic secretion and the resulting productivity of these mutualistic systems. At

very low export levels, increasing export will result in increased environmental metabolites with relatively low impact on the strains growth rate. This increase in environmental availability will result in improving the growth rate of the partnered strain dependent on that metabolite resulting in increased overall productivity of the community. However, at high export levels a similar increase in secretion could decrease the growth rate of the host strain enough that it will become less abundant in the population, provide fewer amino acids to the environment and subsequently by feeding it's partner strain less and less of the metabolite, result in a decrease to the community growth rate. This tradeoff between a species growth/export and the relative abundance in the population is critical to robust growth of the community and highlight the fact that some intermediate "Goldilocks" level of secretion is likely to be optimal in these systems. This reinforces the notion that fine-grained control of secretion levels is critical and an evolutionarily tuned parameter of these systems. Natural questions following these realizations include: Which levels of cooperativity are optimal for exchange and productivity of the system? Is there a single attractor or point of intersection along the secretion gradients that is optimal? Does the fitness landscape of such a community have multiple wells or stable points? What is the peakedness of the fitness landscape? These questions are addressed through i*n vitro* experiments using the presented mutualistic systems. The degree of cooperativity is genetically encoding by modulating exporter expression levels through tweaking the ribosome binding site (RBS and promoter of the respective regulator(Kosuri

et al., 2013).

To study the fine-grained effects of modulating degree of cooperativity, the ΔR-*yddG* ΔF-*argO* pairing was selected. We propose that the higher fitness costs identified to be associated with *argO yddG* expression would provide a strong selective pressure to the optimization of their expression. Episomal expression of exporters was modulated through genetic engineering of the RBS-promoter combination controlling the expression of these genes. The pTet promoter was selected as it is know to be a relatively strong promoter with stringent on/off induction behavior(Lutz & Bujard, 1997). The vector was based on the same backbone with a p15A medium copy origin of replication. Thirteen RBS-promoter pairs were selected from Goodman *et al* to represent a diversity of expression levels. It is critical to have these expression levels genetically encoded and not modulated through chemical induction systems to independently tune the expression of both exporters in the same environment and to maintain the selected diversity of expression levels throughout the experiment. To confirm the amplitude and distribution of expression levels resulting from the selected promoter/RBS combinations, GFP was inserted downstream of the engineered regulator region. Each population was grown to mid-log phase and the expression levels were assayed on a cell-by-cell basis using Flow Cytometry (methods; figure 32). While expression of the exporters themselves may differ slightly from the level determined with GFP, the levels are likely to be related(Kosuri et al., 2013). The library demonstrates an evenly distributed

degree of expression levels spanning two logs of normalized fluorescent units. Interestingly, many fluorescent distributions for individual RBS/promoter pairings had a long low-expression tail. While unlikely that this tail will significant skew the behavior of the system, it is important to keep in mind.

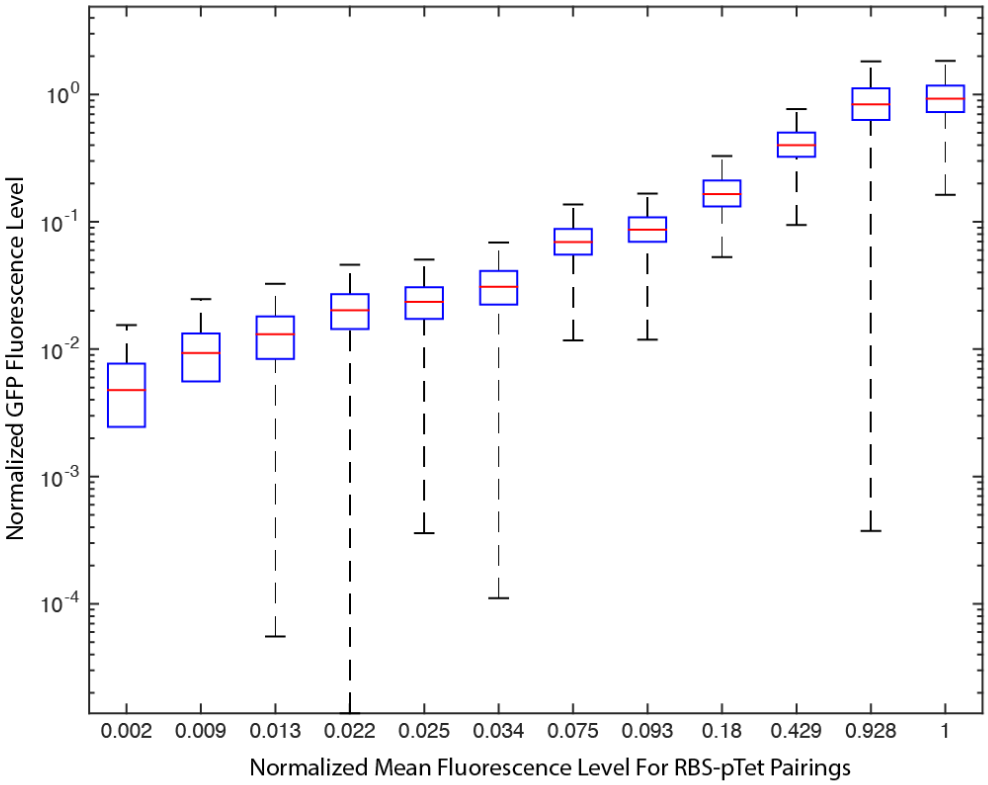

Figure 32: Distributions and means for GFP expression levels of the constructed RBS-promoter pairs.

Interactions along the generated cooperativity gradients are first examined in a pair-wise manner to map out the dynamics of the interaction space. ΔR auxotrophs with 10 *yddG* expression levels and ΔR auxotrophs with 6 *argO* expression levels representative of the constructed library's expression distribution are selected. Each ΔR-ΔF combination is inoculated at equal density and monitored continuously for growth rate and total productivity (OD attained)

and discretely for relative abundance (Figure 33).

Along the *yddG* expression gradient, we are seeing the predicted tradeoff behavior between *yddG* expression levels and community growth. High levels of expression of the exporter result in very low levels of community growth likely due to the observed reduction of relative abundance of the *yddG* expressing strains in those cases (small light blue circles in figure 33). Similarly, as *yddG* expression reaches the lowest levels assayed, we begin to observe a reduction in growth rate and final OD reached by the co-cultures. Generally the relative abundance of the *yddG* expressing strain increases as its expression levels decreases. However, there are exceptions to this, medium *yddG* expression levels result in a dip in the abundance of the ΔR-*yddG* strain. Similarly, the two lowest *yddG* expression levels have a decrease in that strain's abundance.

The predicted secretion/community growth tradeoff is less apparent along the *argO* gradient. At high *yddG* levels, all *argO* expression levels result in similar community productivities and the relative abundance of the strains is biased towards the arginine exporting strain. This indicates that changes to the expression of *argO* in a large population of *argO* containing strains results in an insubstantial marginal change to the extracellular R concentrations which isn't reflected by growth of the community. At medium *yddG* levels the productivity (specifically at *yddG* level 0.1) of the community more or less comparable across all *argO* expression levels with the relative abundance becoming increasingly biased towards the ΔR-*yddG* strain with increasing *argO* expression. This

expression regime suggests that multiple abundance/expression combinations may have equivalent fitness relative to our ability to quantify fitness differences and may pose multiple selective points or solutions for this community. This broad solution space represents a flat fitness landscape for the community. In such a regime, community feed back for cooperativity is likely to be limiting and selection on the individual is likely to dominate. There for natural selection in this regime would likely favor the lower *argO* expressing ΔF strains. At the lower *yddG* expression levels the community productivity/secretion tradeoff becomes more apparent especially at the extremes where community growth quickly drops off at very high and very low levels of *argO* expression.

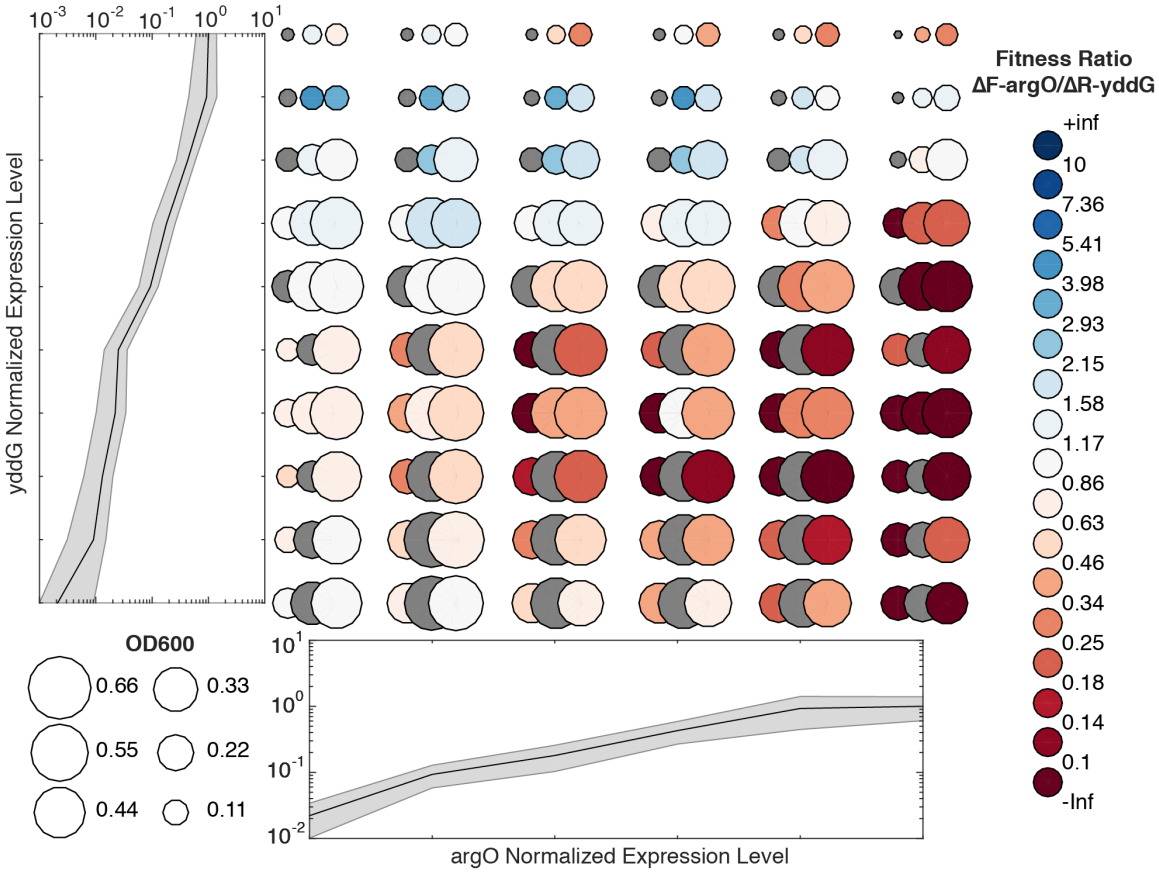

Figure 33: Pairwise growth and relative abundance of cross-feeding strains across a diverse matrix of cooperativity pairings.

These results also emphasize that it we may be missing the tradeoffs occurring at even lower *argO* expression levels. This may be due to the fact that the *argO* exporter had higher export levels of R than *yddG* did for F at equivalent expression levels (HPLC). To further investigate the regime with limiting R availability, *yddG* expression in the ΔR strain was decoupled from induction by deleting the *tetR* repressor and concentration of the inducer, anhydrous-tetracycline (ATC), was titrated to investigate lower *argO* expression levels. To further investigate the lowest possible *argO* expression levels the null RBS construct was also included in this experiment (figure 34). Over these lower

argO expression levels, the exchange/community productivity tradeoff comes into focus. Intermediate levels of *argO* secretion result in the highest community growth rates. When *argO* is increased or decreased beyond these levels, the growth rate of the community decreases. Similar to the case for *yddG* presented above, increasing *argO* expression levels results in a shift towards relative abundance dominated by the ΔR-*yddG* strain. This results in increases to community growth with increasing R secretion at low argO expression levels. At higher expression levels, the marginal cost of further increasing secretion of R is costly to the ΔF strains and results in a reduction in relative abundance. With this resulting lower relative abundance the ΔF strain's ability to provide R to the high abundance of the partnered ΔR strain is reduced and the overall community productivity takes a hit. Therefor the community growth/ secretion and relative abundance tradeoff s exist for the *argO* exporter as well but the effect is only visible over a much large range than the *yddG* exporter.

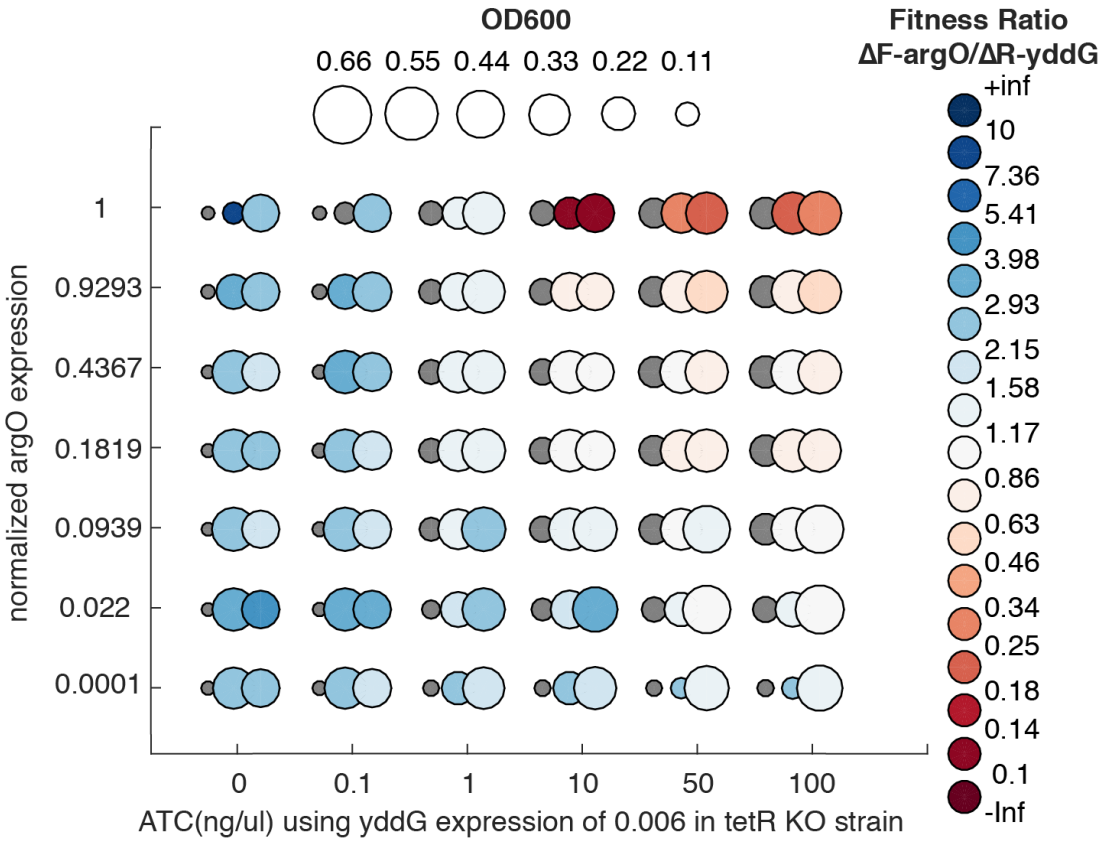

Figure 34: Cross-feeding community productivity and strain bias at low levels of *argO* expression

This lower range of *argO* expression also highlights in interesting tradeoff between relative abundance and absolute abundance of the ΔF-*argO* strain. The increase in expression of the exporter from ~0 to 0.1 results in a decrease in relative abundance from 73% to 59%. However, the increase in export results in great community productivity and an increase in absolute abundance of the ΔF-*argO* strain at 24 hours from 1.4e8 cells to 3.8e8 cells. This behavior reinforces the complex tradeoffs and feedback that exists in these syntrophic communities where even if the partnered strain captures most of the benefit of an increase in export, it may still be more beneficial to exchange than not exchange at all (figure 35).

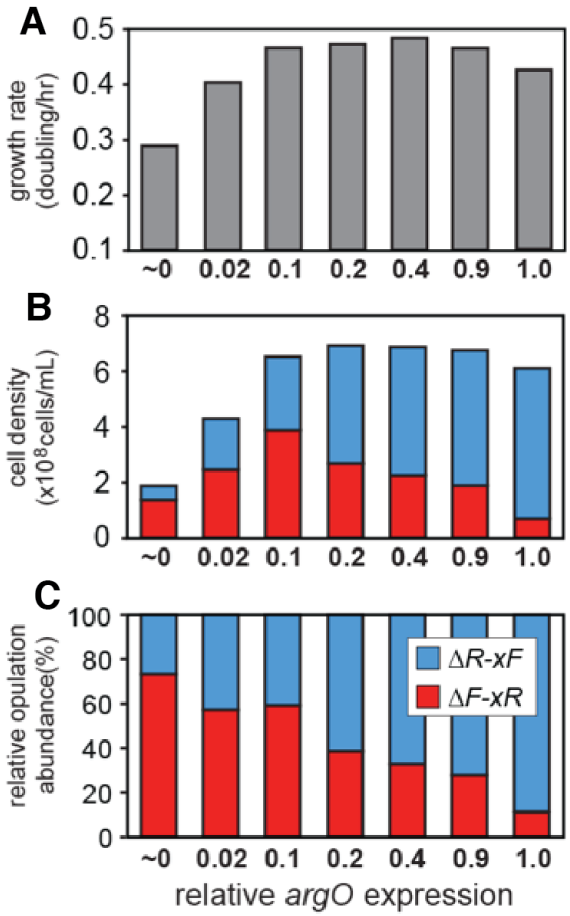

Figure 35: Experimental measurement of growth-abundance tradeoffs. (a) Growth rate (b) 24-hr cell density, and (c) population ratio of co-cultures of ΔR-xF (yddG) and ΔF-xR (argO) where argO expression is increased on a relative scale of 0 to 1.

5.2.6 Cooperativity dynamics in complex communities:

To better control the dynamics and structure of these cooperative behaviors in the context of natural communities it is critical to understand how the tradeoffs elucidated in the pairwise exchange interactions scale to more complex communities comprised of varying degrees of cooperativity. Similarly, to control genomic modulation of export traits better selections for cooperative behaviors are critical. An effective genomic engineering strategy involves increasing

genotypic diversity followed by selection of the optimal phenotype of choice(H. H. Wang et al., 2009). To utilize such a strategies for parameters affecting cooperative behaviors it is critical to determine whether mixtures of cells containing diverse phenotypes can be stratified by a selection or enrichment. Ten secretion levels of both the ΔF-*argO* and ΔR-*yddG* strains were combined in equal levels based on OD6000 measurements and grown with and without spatial structure. This mixture inoculated in well-mixed liquid media represents a lack of spatial structure. Such an environment is hypothesized to bias selection towards secretion levels optimal for an individual cell, i.e. lower degrees of secretion will be enriched. To provide spatial structure, the mixture is spread on minimal media agar plates supplemented with appropriate carbon sources and inducers for the amino acid exporters. Cells were seeded on the plates at various densities to examine the effect of starting intercellular distance on modulating which strain/export strength combination propagates the most. To quantify the structural dynamics of the community, cells were harvested at 0, 12 and 24 hours for the liquid growth and 0, 36 and 72 hours for growth on plates. Episomal DNA is isolated from the cells and the regulator region was amplified with barcoded primers for pooled amplicon next generation sequencing (methods).

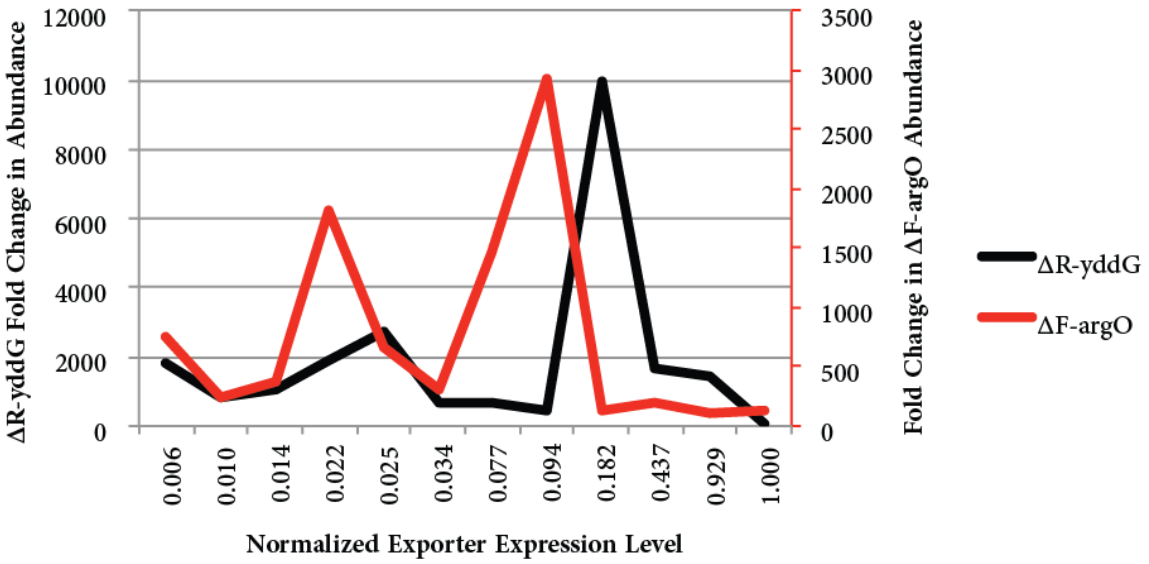

Figure 36: Monitoring competition between gradients of cooperative phenotypes. Here the fold enrichment after 48hrs of growth is presented for strains seeded at low density on agar plates.

At the low seeding levels we found that as in the majority of the pair-wise experiments, the ΔR-yddG strain was more enriched in roughly a ratio of 5:1 relative to the ΔF-argO strain. Both the R secreting strain and the F secreting strain seemed to enrich two distinct levels of export suggesting more than one combination of cooperativity is optimal and being selected in this context (figure 36). The most enriched level of cooperativity in these communities was at about 20% and 10% of the maximum level tested for yddG and argO respectively excluding many less cooperative phenotypes. The exclusion of lower level of cooperativity from the population suggest that solid surface growth may be a viable method for enriching populations of cell with genetically encoded diversity of cooperative phenotypes for an optimal level. However, the exclusion of the most cooperative phenotype tested indicates that it is unlikely that such a

strategy would be able to pull out of the population exporters with greatly increased kinetics.

5.3 Conclusion:

In these studies we've demonstrated the potential for syntrophic exchange to screen for functional export. Critically, the observed productivity of this cross-feeding growth is sensitive to the expression level of the exporter and therefor any implementations of high-throughput bioprospecting for exporters should use multiple expression levels. Cooperative phenotypes associated with export can be maintained in mutualistic communities despite competition with cheaters. However, the degree of feedback from the dynamic buffer strain is dependent on it's auxotrophy and exporter and combinations must be carefully selected depending on the application. The caveats associated with the partnered strain are less critical if it is possible to use spatial structure to direct exchange between cooperative strains and exclude cheaters from the population.

Tradeoffs between an individual's relative abundance and community growth are consistently observed along gradients of the cooperative export phenotype. At the lowest levels of export, benefits of increases in export can be disproportionately captured by partnered cross-feeding strains yet exporting is still more beneficial than not exporting in terms of absolute abundance of the species in question. This observation is critical to the evolution of cooperativity and division of labor in natural communities. This finding is also critical to

designing system to evade encroachment from cheaters. It suggests that there is a low expression regime where maintenance of some amount of export is favored. These tradeoffs also have implications for automated genome engineering: null solutions or sequences resulting in broken exporters are likely to be excluded from the population but so are cases where the kinetics of secretion is greatly increased. However optimal levels optimal for community growth are enriched and such a selection system could be useful for tuning and ensuring robust growth of more complex engineered communities.

6 Beyond 20 Amino Acids: Metabolic Auxotrophies for Biocontainment

6.1 Overview

We've demonstrated that amino acids auxotrophies have great potential to modulate the dynamics and structure of bacterial communities. However, the bacterial strains we use also show great genomic plasticity, allowing them to circumvent engineered metabolic dependencies. As we move towards engineering microbial communities in open environments, it is essential to have tools that enable orthogonal control of our synthetic system. Such tools should ensure robust and evolution resistant modulation of the community not only to ensure optimal functionality but also to address the risks of unintended proliferation of GMOs in natural ecosystems through biocontainment. As we've discussed in previous sections, current state of the art metabolic dependencies impose evolutionary pressure on the organism that can easily be circumvented by environmentally available compounds, or overcome by horizontal gene transfer (HGT) or mutation. Here we investigate the ability of redesigned essential enzymes in the first organism possessing an altered genetic code to confer metabolic dependence on nonstandard amino acids for survival. The resulting GMOs cannot metabolically circumvent their biocontainment mechanisms using environmentally available compounds, and they exhibit unprecedented resistance to evolutionary escape *via* mutagenesis and HGT. This work provides a foundation for safer GMOs that are isolated from natural ecosystems by reliance on synthetic metabolites.

Current strategies for containment of GMO's rely on integrating toxin/antitoxin "kill switches"(Q. Li & Wu, 2009; Molin et al., 1987), establishing auxotrophies for essential compounds(Curtiss, 1978), or both(Ronchel & Ramos, 2001; Wright, Delmans, Stan, & Ellis, 2015). Toxin/antitoxin systems suffer from selective pressure to improve fitness through deactivation of the toxic product(Knudsen et al., 1995; Pasotti, Zucca, Lupotto, Cusella De Angelis, & Magni, 2011), while metabolic auxotrophies can be circumvented by scavenging essential metabolites from nearby decayed cells or cross-feeding from established ecological niches. Effective biocontainment strategies must protect against three possible escape mechanisms: mutagenic drift, environmental supplementation and horizontal gene transfer (HGT). Here we introduce "synthetic auxotrophy" for non-natural compounds as a means to biological containment that is robust against all three mechanisms. Using the first genomically recoded organism (GRO)(Lajoie et al., 2013) we assigned the UAG stop codon to incorporate a nonstandard amino acid (NSAA) and computationally redesigned the cores of essential enzymes to require the NSAA for proper translation, folding and function. Combining multiple redesigned enzymes resulted in GROs that exhibit dramatically reduced escape frequencies and readily succumb to competition by unmodified organisms in nonpermissive conditions.

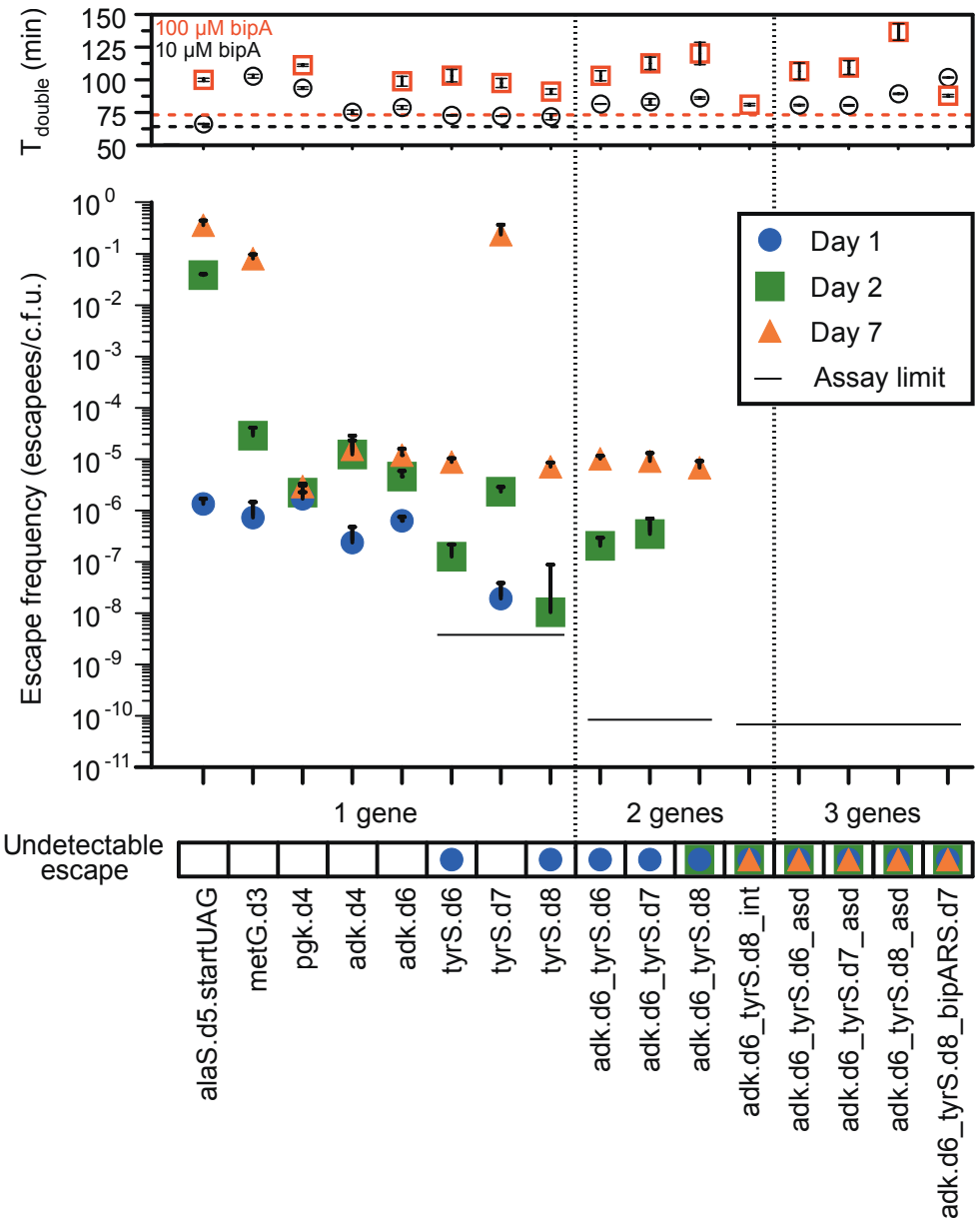

Figure 37: Escape frequencies and doubling times of auxotrophic strains. Escape frequencies for engineered auxotrophic strains calculated as colonies observed per colony forming unit (c.f.u.) plated over 3 technical replicates on solid media lacking arabinose and bipA. Assay limit is calculated as 1/(total c.f.u. plated) for the most conservative detection limit of a cohort, with a single-enzyme auxotroph limit of 3.5 x 10^{-9} escapees/c.f.u., a double-enzyme auxotroph limit of 8.3 x 10^{-11} escapees/c.f.u. and a triple-enzyme auxotroph limit of 6.41 x 10^{-11} escapees/c.f.u. Positive error bars are standard error of the mean (s.e.m.) of the escape frequency over three technical replicates (Methods). The top panel presents the doubling times for each strain in the presence of 10 μM or 100 μM bipA, with the parental strain doubling times represented by the dashed horizontal lines. MetG.d3 growth was undetectable in 100 μM bipA. Positive and negative error bars are s.e.m.

6.2 Results

6.2.1 Susceptibility of Natural Auxotrophies to Escape and Supplementation

To compare synthetic auxotroph strains to current biocontainment practices we generated natural metaboblic auxotrophs by knocking out *asd* and *thyA* genes from an MG1655-derived *E. coli* strain (EcNR1). The *asd* knockout renders the strain dependent on diaminopimelic acid (DAP) for cell wall biosynthesis(Curtiss, 1978), while the *thyA* knockout deprives the cell of thymine, an essential nucleobase(Steidler et al., 2003). These well-studied auxotrophies are commonly incorporated into biocontainment strategies(Curtiss, 1978; Wright et al., 2015). In agreement with previous studies, the *asd* knockout shows strong dependence on its requisite metabolite, with a 7 day escape frequency of 8.97 x 10^{-9} escapees/c.f.u.. Knocking out *thyA* from this strain to produce a double-enzyme auxotroph did not reduce the 7 day escape frequency (8.79 x 10^{-9} escapees/c.f.u.) (table 4). Nevertheless, metabolic strategies could complement synthetic auxotrophies to improve escape frequencies in defined ecological niches. To test this principle we knocked out *asd* from the double-enzyme synthetic auxotrophs of *adk* and *tyrS* resulting in three triple-enzyme auxotrophs (adk.d6_tyrS.d6_asd, adk.d6_tyrS.d7_asd and adk.d6_tyrS.d8_asd) that grow robustly in permissive conditions but show undetectable escape after 7 days on media lacking bipA and DAP (figure 37, detection limit 6.4 x 10^{-11} escapees/c.f.u.).

Strain	Escape assay plate media	Day 1 esc. frequency	Day 2 esc. frequency	Day 7 esc. frequency	Doubling time in LB^L + DAP + thymidine
EcNR1ΔthyA	LB^L	< 3.51E-11	2.70E-09 ± 1.20E-10	7.40E-08 ± 5.14E-09	43.6 ± 0.2
EcNR1Δasd	LB^L	< 5.98E-10	5.98E-10 ± 5.38E-10	8.97E-09 ± 5.46E-09	47.7 ± 0.3
EcNR1ΔasdΔthyA	LB^L	< 1.83E-10	< 1.83E-10	8.79E-09 ± 4.81E-09	57.5 ± 0.2
EcNR1ΔasdΔthyA	LB^L + thymidine	< 5.49E-10	1.10E-09 ± 5.65E-10	2.86E-08 ± 4.40E-09	n.a.

Table 4: Escape and growth rates of natural metabolic auxotrophs. Doubling times in minutes. Errors reported are s.e.m. of the rate (Methods) < Indicates below given detection limit. n.a. = not applicable

While bacterial growth assays are often carried out in variations of media enriched with yeast extract, GMOs are increasingly deployed among a diversity of ecosystems that may provide opportunities for scavenging or cross-feeding essential metabolites. To compare metabolic and synthetic auxotroph strategies in an environment mimicking endogenous bacterial communities we grew engineered variants of both natural and synthetic auxotrophs in LB^L containing *E. coli* lysate (Methods). We hypothesized that since DAP is an essential component of the bacterial cell wall, the Δ*asd* strains may scavenge sufficient DAP from *E. coli* lysate to complement the auxotrophy. As anticipated, metabolic auxotrophs obtained sufficient nutrients from the yeast/tryptone (LB^L) and the bacterial remnants (lysate) to support exponential growth (figure 38a-d), while the synthetic auxotrophs failed to circumvent their dependencies. These results highlight the importance of establishing auxotrophies for compounds that are not

environmentally available, and of ensuring the metabolic essentiality of enzymes intended to confer dependence.

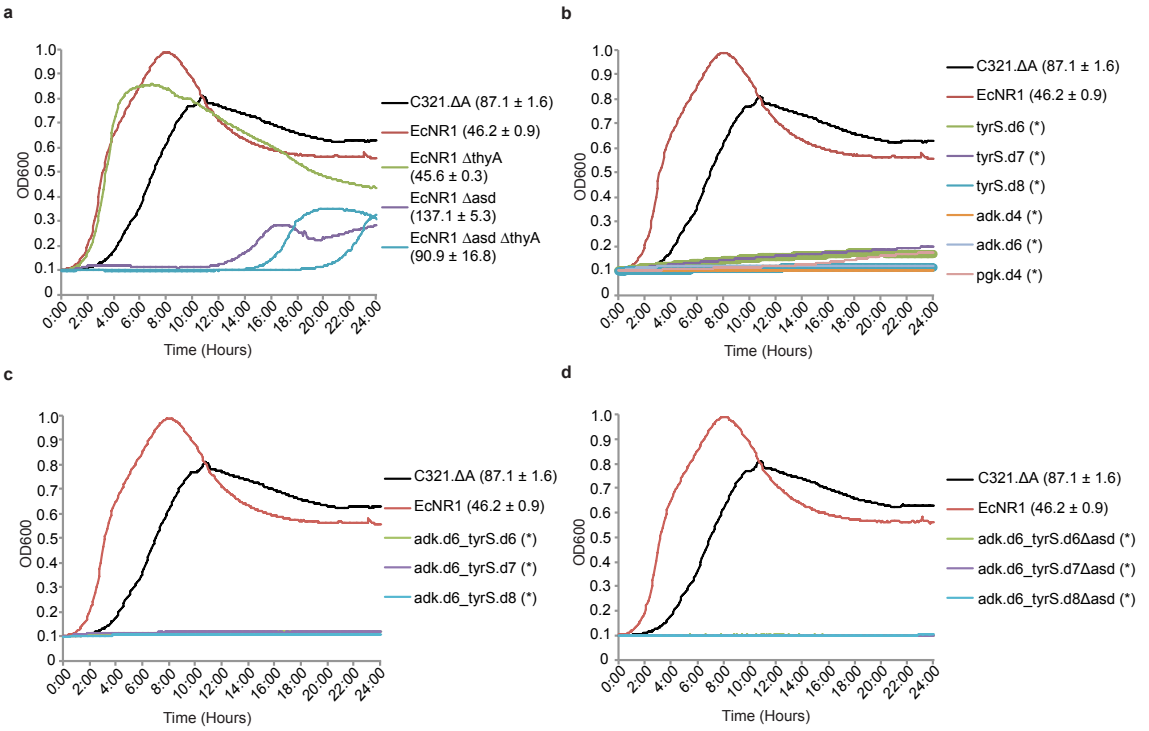

Figure 38: Natural metabolites can circumvent auxotrophies. a, The synthetic auxotroph parental strain (C321.ΔA), a second prototrophic MG1655-derived strain (EcNR1), and three natural auxotroph derivatives of EcNR1 were grown in LB^L supplemented with 166.66 ml/L bacterial lysate (Teknova). Growth curves are shown with doubling times ± one standard deviation of three technical replicates next to the labels. The conditions fully complement the metabolic auxotrophy of EcNR1.ΔthyA, which doubles as robustly as prototrophic EcNR1. Strains lacking the *asd* gene (EcNR1.Δasd and the EcNR1.ΔasdΔthyA double knockout) show more impairment but enter exponential growth with doubling times of 91 to 137 minutes, respectively. b, (single-) and c, (double-)enzyme synthetic auxotrophies are not complemented by natural products in rich media or bacterial lysate. d, When the Δasd auxotrophy is combined with double-enzyme synthetic auxotrophies the natural products are no longer sufficient to support growth. No growth is indicated by * in b-d.

6.2.2 Resistance to horizontal gene transfer

Horizontal gene transfer is an important mechanism of evolution in any genetically rich environment(Smillie et al., 2011). We developed a conjugation escape assay to assess how DNA transfer within an ecosystem enables a GMO to escape biocontainment. Whereas any recombination event that replaces an inactivated gene could overcome metabolic auxotrophies(Wollman, Jacob, & Hayes, 1956), we hypothesized that conjugal escape would be disfavored in GROs because donor DNA replacing bipA-dependent genes would also overwrite crucial genetic elements involved in genetic code reassignment (figure 39). For example, reintroducing UAG stop codons into essential genes without restoring RF1-mediated translational termination could be deleterious(Lajoie et al., 2013) or lethal(Mukai et al., 2010). Furthermore, reintroducing RF1 would result in competition between bipA incorporation and translational termination, undermining the recoded functions of the GRO.

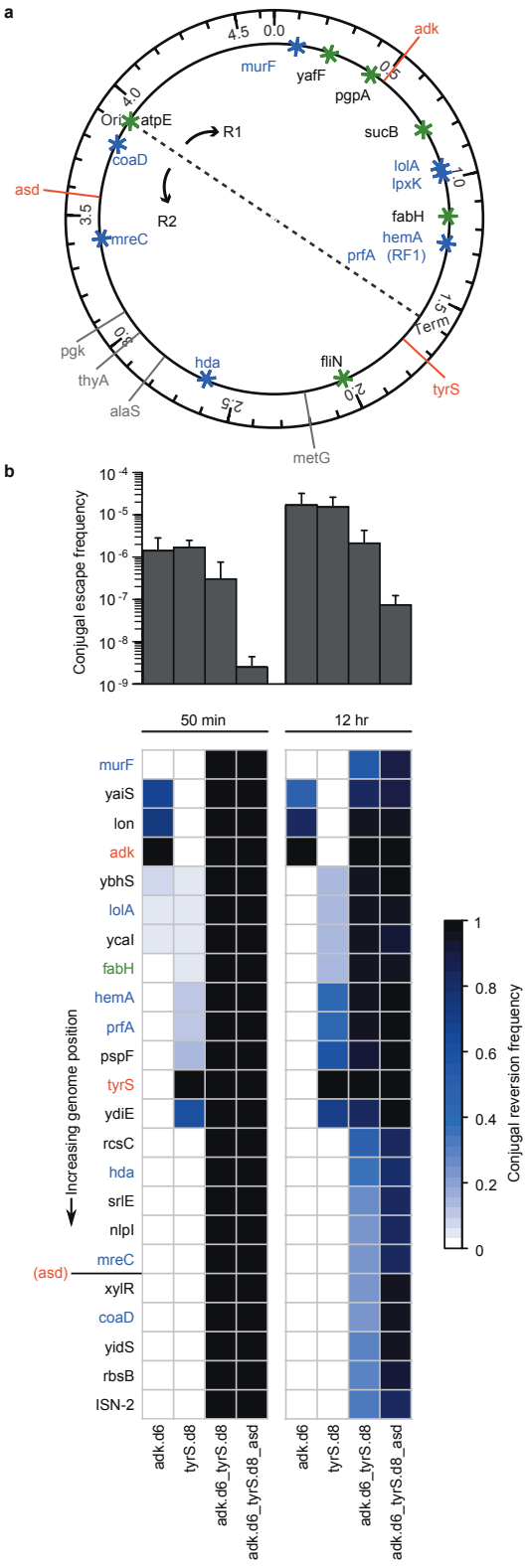

Figure 39: Synthetic auxotrophy and genomic recoding reduce HGT-mediated escape. a,

The positions of key alleles are plotted to scale on the genome schematic. Red lines indicate auxotrophies used in the multi-enzyme auxotrophs and gray lines indicate other auxotrophies that were not included in this assay. Asterisks indicate important alleles associated with the reassignment of UAG translation function (blue are essential genes and green are potentially important genes(Lajoie et al., 2013)). Conjugation-mediated reversion of the UAA codons back to the wild-type UAG is expected to be deleterious unless the natural UAG translational termination function is reverted. R1 and R2 denote replicores 1 and 2, respectively. b, Combining multiple synthetic auxotrophies in a single genome requires a large portion of the genome to be overwritten by wild-type donor DNA, reducing the frequency of conjugal escape (top panel) and increasing the likelihood of overwriting the portions of the genome (bottom panel) that provide expanded biological function (e.g., *prfA* encodes RF1, which mediates translational termination at UAG codons). Positive error bars indicate standard deviation.

In order to simulate a worst-case scenario in ecosystems containing a rich source of conjugal donors, we used Tn5 transposition to integrate an origin of transfer (*oriT*) into a population of *E. coli* MG1655 conjugal donor strains. We isolated a population of ~450 independent clones (one *oriT* for every ~10kb portion of the 4.6 megabase pair genome) and sequenced the flanking genomic regions of 96 donor colonies to confirm that *oriT* integration was well-distributed throughout the population. We then conjugated this donor population into our auxotrophic strains at a ratio of 1 donor to 100 recipients to increase the probability that conjugal transfer will initiate from one *oriT* position per recipient. Conjugation was performed for durations of 50 minutes and 12 hours (average conjugation times predicted to transfer 0.5 and 7.2 genomes) to simulate a single conjugal interaction and an ecological worst-case scenario, respectively. Conjugal escapees were selected on nonpermissive media, and 23 alleles distributed throughout the genome (figure 39a) were screened using multiplex allele-specific colony PCR (mascPCR) to assess how much of the recoded genome is replaced by wild-type donor DNA.

Conjugal escape frequency decreases as the number of auxotrophic gene variants increases (figure 39b, top panel; figure 40), consistent with larger portions of the genome that must be overwritten for conjugal escape of the multi-enzyme auxotrophs (figure 39b, bottom panel). The 12 hour conjugations effect higher escape frequencies than do the 50 minute conjugations, and the 12 hour conjugations produce a larger diversity of conjugal escape genotypes, consistent with an increased opportunity to initiate new conjugal transfers during the mating period. Encouragingly, all 50-minute conjugal escapees from multi-enzyme auxotrophs exhibit the wild-type donor sequence at all 23 assayed alleles (figure 39b, bottom panel), resulting in the reintroduction of release factor 1 and its UAG-mediated translational termination function. This collateral replacement of recombinant genomic DNA could be extended to other recombinant payloads such as antibiotic resistance genes, recombinases, catabolic enzymes, toxins, and orthogonal aaRS/tRNA pairs used for NSAA incorporation.

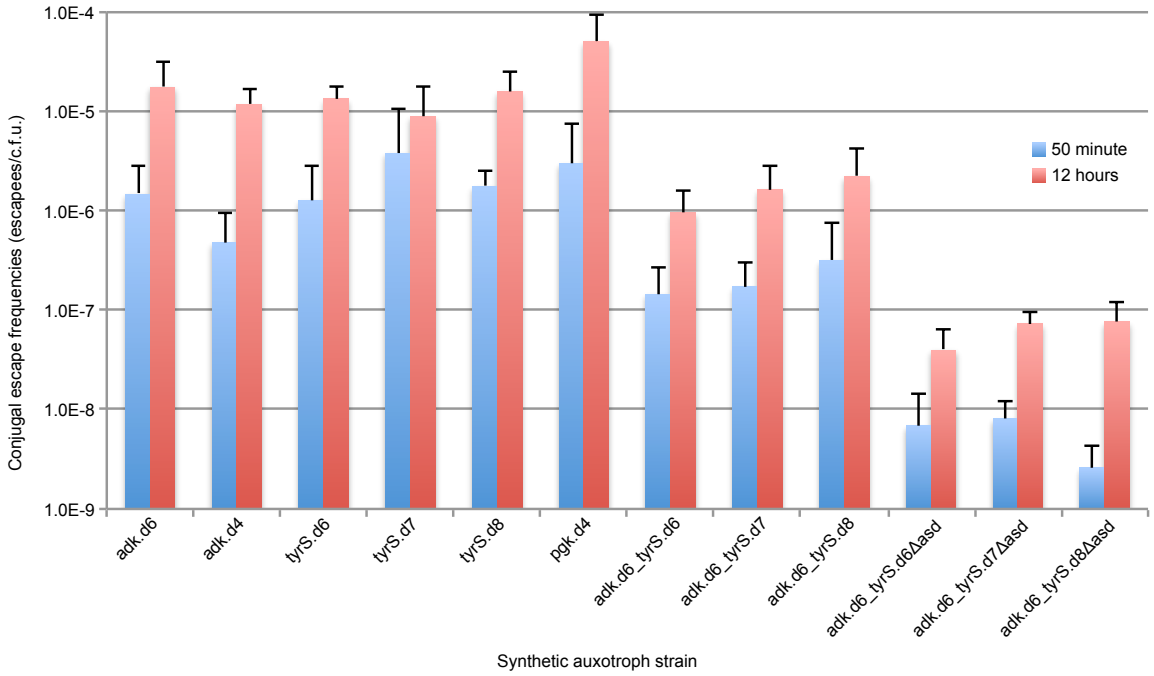

Figure 40: Conjugal escape frequencies of synthetic auxotrophs. Single, double, and triple-enzyme auxotrophs were assayed to determine the frequency of escape by horizontal genetic transfer and recombination from a prototrophic donor as described in the methods. These results highlight the benefit of having multiple auxotrophies distributed throughout the genome. Notably, scaling from a single synthetic auxotroph to three distributed auxotrophies results in a reduction of conjugal escape by at least two orders of magnitude.

6.2.3 Competition between synthetic auxotroph escapees and prototrophs.

Any biocontainment mechanism – however robust – is vulnerable to evolutionary pathways through the fitness landscape that lead to survival in nonpermissive conditions. Although escape pathways may be exceedingly rare, it is critical that genotypes along these pathways sufficiently decrease fitness so that escapees are outcompeted in natural ecosystems. In this respect toxin/antitoxin systems are disadvantaged because the primary escape mechanism – ejecting the toxic gene – typically improves fitness. In contrast,

escapees of our synthetic auxotrophs are highly impaired under nonpermissive growth conditions. We quantified the ability of escaped synthetic auxotrophs to compete in an ecosystem using a flow cytometry-based competition assay (figure 41).

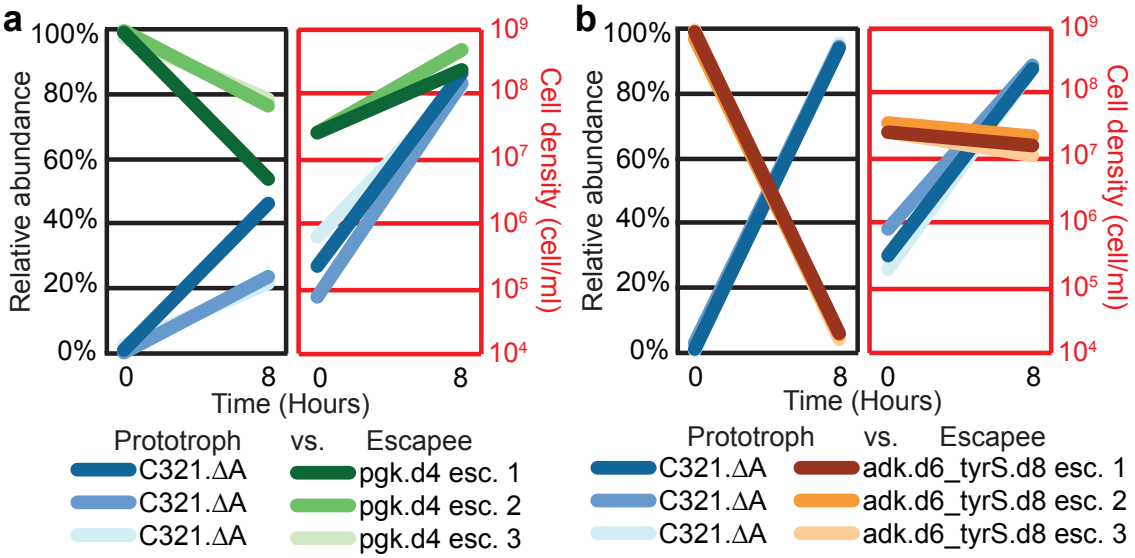

Figure 41: Competition between synthetic auxotroph escapees and prototrophic *E. coli*. C321.ΔA was competed in the absence of bipA against escapees from a single-enzyme bipA auxotroph (pgk.d4, moderate NSAA dependence), or from a double-enzyme bipA auxotroph (adk.d6_tyrS.d8, strong bipA-dependence). Populations were seeded with 100-fold excess escapees and grown for 8 hours in nonpermissive conditions. The populations were evaluated using flow cytometry for episomally expressed fluorescent proteins at t = 0 and t = 8 hours. Results from separate competition experiments against 3 different escapees are shown for each synthetic auxotroph. a, Pgk.d4 escapees continue to expand in a mixed population with C321.ΔA after 8 hours. b, Adk.d6_tyrS.d8 escapees are rapidly outcompeted by C321.ΔA, which overtakes the population after 8 hours.

We clonally isolated escapee strains that emerged both from single-enzyme (pgk.d4) and double-enzyme (adk.d6_tyrS.d8) synthetic auxotrophs and competed them against C321.ΔA, used here as a proxy for prototrophic environmental competitors. From an initial seeding density 100-fold higher than the prototrophs, the single-enzyme pgk.d4 escapees maintain a significant presence (>50% of the total population) as resources become limiting after 8

hours of growth (figure 41a). The continued presence of pgk.d4 in the 8 hour population arises from escapees growing in nonpermissive conditions, even though fitness is impaired in comparison to the prototroph. In contrast, under identical seeding conditions, the prototrophic strain completely overtakes the adk.d6_tyrS.d8 escapees, inverting their relative abundance within 8 hours (figure 41b). This extreme effect is largely due to the severe fitness impairment of the adk.d6_tyrS.d8 escapees in nonpermissive conditions. Further, these results emphasize that cross-feeding from prototrophic co-cultures cannot circumvent synthetic auxotrophies. Thus, while GMOs biocontained by multiple synthetic auxotrophies may explore mutations conferring nominal viability under nonpermissive conditions, the associated fitness impairment renders them readily outcompeted by prototrophic microbial competitors.

6.3 Conclusions

The work presented here highlights the importance of considering environmental context when defining a metabolism-catalyzing gene as essential or not. We've previously shown that amino acid dependencies are commonly circumvented by cross-feeding with partners in the community. Here we demonstrate that nutrients such as DAP can be scavenged from microbial communities and their use as biocontainment strategies is unjustified. Alternatively, non-standard amino acids that are only produced through non-synthetic chemical synthesis present an excellent opportunity for engineering dependencies that cannot be circumvented by individual strains, nor by the

efforts of microbial communities. Engineering a dependency on a synthetic amino acid has show to decrease the rate of mutagenic escape orders of magnitude below the current suggested level. This strong dependence also makes these types of metabolites a fantastic orthogonal control mechanism for modulating microbial communities in nature.

6.4 Methods

6.4.1 Culture and selection conditions.

Growth media consisted of LB-Lennox (LB^L, 10 g/L bacto tryptone, 5 g/L sodium chloride, 5 g/L yeast extract). Permissive growth media for bipA-dependent auxotrophs was LB^L supplemented with sodium dodecyl sulfate (SDS), chloramphenicol, bipA, and arabinose. Nonpermissive media lacked bipA and arabinose. The following selective agents, nutrients, and inducers were used when indicated: chloramphenicol (20 mg/ml), kanamycin (30 mg/ml), spectinomycin (95 mg/ml), SDS (0.005% w/v), bipA (10 μM), glucose (0.2% w/v), arabinose (0.2% w/v). Permissive media for metabolic auxotrophs is LB^L supplemented with 75 mg/ml DAP and 100 mg/ml thymidine.

6.4.2 Strain engineering.

The tyrS.d6, tyrS.d7, and tyrS.d8 gene variants were constructed by PCR amplification of the *E. coli* MG1655 *tyrS* gene with mutagenic primers, followed by full-length Gibson assembly(Gibson et al., 2009) and recombination onto the genome using λ Red recombineering(Datsenko & Wanner, 2000; Yu et al.,

2000). Strains tyrS.d6, tyrS.d7, tyrS.d8, adk.d6_tyrS.d6, adk.d6_tyrS.d7, and adk.d6_tyrS.d8 were produced by (1) deleting the endogenous *tdk* gene from C321.ΔA, (2) replacing the endogenous *adk* and *tyrS* genes with their codon-shuffled variants (*adk*(recode)-*tdk* and *tyrS*(recode)-*tdk*, Supplementary Table 3) transcriptionally fused to *tdk*, and (3) replacing the fusion cassettes with the adk.d6, tyrS.d6, tyrS.d7, or tyrS.d8 variants. Variants of adk.d6, tyrS.d7 and tyrS.d8 containing leucine and tryptophan at the bipA position were constructed by MAGE with oligos containing the appropriate mutations and clonal populations were isolated on LB^L plates lacking bipA and arabinose. Triple-enzyme auxotrophs were created by replacing *asd* with a *Δasd::spec*^{R} cassette. All genotypes (Supplementary Table 3) were confirmed using mascPCR(Isaacs et al., 2011) and Sanger sequencing using primers from Supplementary Table 1.

6.4.3 Strain doubling time analysis.

Strain doubling times were calculated as previously described(Lajoie et al., 2013).

6.4.4 Solid media escape assays for natural metabolic and synthetic auxotrophs.

All strains were grown in permissive conditions and harvested in late exponential phase. Cells were washed twice in LB^L and resuspended in LB^L. Viable c.f.u. were calculated from the mean and standard error of the mean

(s.e.m.) of 3 technical replicates of 10-fold serial dilutions on permissive media. Three technical replicates were plated on nonpermissive media and monitored for seven days. The order of magnitude of cells plated ranged from 10^2 to 10^9 depending on the escape frequency of the strain. Synthetic auxotrophs were plated on two different nonpermissive media conditions: “SC”, LB^L with SDS and chloramphenicol, and “SCA”, LB^L with SDS, chloramphenicol, and 0.2% arabinose. Metabolic auxotrophs were plated on LB^L for nonpermissive conditions (Supplementary Table 11). If synthetic auxotrophs exhibited escape frequencies above detection limit (lawns) on SC at day 1, 2 or 7 (alaS.d5, metG.d3, tyrS.d7), escape frequencies for those days were calculated from additional platings at lower density. Additional platings at higher density were also used to obtain day 1 and day 2 escape frequencies for pgk.d4 on SC. The s.e.m. $S_{\bar{X}}$ across technical replicates of the cumulative escape frequency ν scored for a given day was calculated as: $S_{\bar{X}} = \nu\sqrt{\left(\frac{S_{\bar{X}}\tau}{\tau}\right)^2 + \left(\frac{S_{\bar{X}}n}{n}\right)^2}$, where τ is the mean number of c.f.u. plated, $S_{\bar{X}}\tau$ is the s.e.m. of c.f.u. plated, n is the mean cumulative colony count up to the given day, and $S_{\bar{X}}n$ is the s.e.m. of the cumulative colony count up to the given day. If synthetic auxotroph escapees emerged on SC, three clones were isolated, their growth rates were calculated as described above, and the doubling time of the fastest escapee was recorded.

6.4.5 Growth competition assays.

The assayed single- and double-enzyme synthetic auxotroph escape strains (pgk.d4 esc. 1, 2, and 3; adk.d6_tyrS.d8 esc. 1, 2, and 3) were

transformed with a pZE21 vector(Lutz & Bujard, 1997) bearing mCFP under anhydrotetracycline (aTc) inducible control in the multiple cloning site. The parental prototrophic C321.ΔA strain was similarly transformed with an identical vector except that the fluorophore is YFP. Strains were grown up to late exponential phase in LB^L supplemented with antibiotics (SDS, chloramphenicol, kanamycin), inducers (0.1% L-arabinose, 100 ng/ml aTc), and bipA. Cells were washed twice in M9 salts and adjusted to a cell concentration of roughly 1 x 10^9 cells/ml. Biological replicates of synthetic auxotroph escapees were mixed with the C321.ΔA strain at a ratio of 100:1 and diluted to a seeding concentration of roughly 2.5 x 10^7 cells/ml in nonpermissive media (LB^L supplemented with SDS, chloramphenicol, kanamycin and aTc). Growth kinetics of the competition mixture were assayed in 200 µl sample volumes on microtiter plates incubated in a Biotek Synergy microplate reader at 34 ºC. Cell mixtures were fixed in PBS with 1% paraformaldehyde at time 0 and at 8 hours. Fixed cells were run on a BD LSRFortessa and populations were binned based on YFP expression level. CFP was not used for species discrimination but rather to maintain consistent fitness costs associated with episomal DNA maintenance and fluorophore expression.

6.4.6 Bacterial lysate growth assays.

All strains were grown up in permissive conditions and harvested in late exponential phase. Cells were washed twice in M9 salts (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl) by centrifugation at 17,900 x g and then diluted 100-fold into LB^L supplemented with 166.66 ml/L trypsin digested *E. coli*

extract (Teknova Cat. No. 3T3900). Growth kinetics were assayed in 200 ul sample volumes on microtiter plates as described above. Three biological replicates were performed by splitting a single well-mixed initial seeding population.

6.4.7 Conjugal escape assays.

The conjugal donor population was produced using the Epicentre EZ-Tn5 Custom Transposome kit to insert a mosaic-end-flanked *kan*^{R}-oriT cassette into random positions of the *E. coli* MG1655 genome. The population of integrants was plated on LB^{L} agar plates supplemented with kanamycin. Approximately 450 clones were lifted from the plate and pooled, which corresponds to one *kan*^{R}-oriT per ~10 kb region of the genome, assuming an equal distribution of transposition across the 4.6 megabase *E. coli* MG1655 genome. The pRK24 conjugal plasmid was conjugated(Isaacs et al., 2011) from *E. coli* strain 1100-2(Tolonen, Chilaka, & Church, 2009) into the *kan*^{R}-oriT donor population. The *kan*^{R}-oriT insertion sites were confirmed to be well-distributed. Briefly, the donor population was sheared on a Covaris E210, end repaired, and ligated to Illumina adapters as described by Rohland and Reich(Rohland & Reich, 2012). Genomic sequences flanking the insertion site were amplified using the Sol-P5-PCR primer and a series of nested primers (Supplementary Table 1) that hybridize within the *kan*^{R} gene. PCR products corresponding to ~1 kb were gel purified from the smear and TOPO cloned (Invitrogen pCR™-Blunt II-TOPO®). Flanking genomic

sequences were then identified by Sanger sequencing 96 TOPO clones. Conjugal escape assays were performed as described previously(Isaacs et al., 2011) with 50 minute and 12 hour conjugal duration and a donor:auxotroph ratio of 1:100. Three technical replicates of two biological replicates were performed for all conjugation assays with the exception of the double-enzyme synthetic auxotroph experiments, which were performed with three biological replicates (3 technical replicates each) to produce enough escapees for mascPCR screening. To determine the proportion of the genome overwritten by donor DNA the following numbers of colonies were scored for the 50 minute/12 hour time points: adk.d6 n=51/6; tyrS.d8 n=44/7; adk.d6_tyrS.d8 n=8/59; adk.d6_tyrS.d8_asd:specR n=5/38. This set omits a small collection of clones that could not be scored due to polyclonality.

7 Conclusions

7.1 Summary of Results

In natural ecological settings, metabolite exchange between cooperative organisms is often difficult to identify, the interactions are hard to quantitatively measure, and the resulting significance are challenging to interpret. However, using comparative genomic analysis across thousands of sequenced genomes we can make can make the interesting observation that there are widespread trends of metabolic cross-feeding based on phylogenetic distributions of metabolic capabilities. The metabolic interrelatedness of organisms and ecosystems is a likely driver of the disproportionate loss of expensive biosynthesis capabilities across the studied metabolome of bacteria. This is further supported by our observations that robust growth can be driven by amino acid exchange and highlights the possibility that these metabolic transfers can enrich consortium-level associations and contribute significantly to the development and persistence of auxotrophic phenotypes in the biosphere.

Further, we developed an ODE model based on simplistic pairwise interactions that is able to scale to higher order complex 3-member syntrophic exchange systems. Both this model and the experimental results highlight methionine, lysine and phenylalanine as strong candidates for modulating robust coculture growth. Work with exporter overexpression also showed that the transmembrane flux of threonine can be greatly improved especially in the context of the methione auxotroph which disrupts and shunts the involved

branched biosynthesis pathway towards threonine production. Exporters overexpression represents a strong cooperative phenotype that can be readily outcompeted by cheaters in head-to-head completion. However, our work highlights that there is feedback within cross-feeding communities that provides some benefit to this behavior. This work has further elucidated a critical tradeoff between export levels, abundance in the population and absolute abundance in the environment. These tradeoffs likely play a critical role in the evolution of cooperatively as these experiments have shown that very low levels secretion can result in greater community productivity and absolute abundance of a cooperative phenotype despite loosing out in relative abundance to surrounding strains.

Long-term evolution of cross-feeding communities has illustrated that complex partially syntrophic exchange networks can be more beneficial to ecosystem productivity than reducing to a minimal binary exchange system. The more complex networks also proved to be fruitful grounds for evolution that unlock varied mitigation mechanism for dealing with stresses of amino acid starvation. The observed genomic plasticity is at once both an inspiration for new community control strategies (e.g. importer over-expression, amino acid pool redistribution, and perturbation of nitrogen starvation regulation) and a reminder of the difficulty to robustly engineer functional behaviors in the face of evolutionary pressures. To this end this story is capped by efforts leveraging an organism with a reassigned genetic code towards mitigating the risks associated

with events that can disrupt an engineered phenotype (genetic drift & mutation, horizontal gene transfer and environmental metabolic cross talk). We demonstrate that an organism with an engineered dependency on a non-standard amino acid is metabolically orthogonal to a multitude of commonly occurring natural environments and demonstrates great robustness to genetic disruptions via mutation and conjugative horizontal gene transfer. Given these characteristics, strains with engineered non-standard amino acid dependences will likely play a leading role in future efforts to perturb and control natural microbial communities.

7.2 Concluding Remarks

The prospect is bright for synthetic biologists to build ecosystem that reproducibly exhibit complex behavior. Yet there remain many challenges ahead that reflect our incomplete understanding of the many governing principles that underlie microbial physiology, ecology, and evolution. A better working knowledge of the different parameters that drive social interaction in cell populations will be needed. As most intercellular interactions exhibit non-linear relationships based on spatial, temporal, thermodynamic, and energetic constraints, we expect that new theoretical frameworks need to be developed to describe these complex, dynamic, and heterogeneous ecosystems. New techniques that facilitate massively parallel synthesis, engineering, and analysis of microbial consortia at single-cell resolution will be critical for predictive

programming of synthetic communities. As we progress toward engineering biological systems of ever-increasing sophistication, social and ethical concerns surrounding the creation of non-natural life forms and ecosystems will require open dialogue between researchers and the public on the risks and rewards of these activities in the post-Darwinian era of biology.

BIBLIOGRAPHY

Achtman, M., & Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. *Nature Reviews. Microbiology, 6*(6), 431–440.

Akashi, H., & Gojobori, T. (2002). Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America, 99*(6), 3695–3700.

Al Mamun, A. A., Lombardo, M. J., Shee, C., Lisewski, A. M., Gonzalez, C., Lin, D., Rosenberg, S. M. (2012). Identity and function of a large gene network underlying mutagenic repair of DNA breaks. *Science, 338*(6112), 1344–1348.

Alper, H., & Stephanopoulos, G. (2009). Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nature Reviews. Microbiology, 7*(10), 715–723.

Atkinson, M. R., Blauwkamp, T. A., Bondarenko, V., Studitsky, V., & Ninfa, A. J. (2002). Activation of the glnA, glnK, and nac promoters as Escherichia coli undergoes the transition from nitrogen excess growth to nitrogen starvation. *Journal of Bacteriology, 184*(19), 5358–5363.

Au, K. G., Clark, S., Miller, J. H., & Modrich, P. (1989). Escherichia coli mutY gene encodes an adenine glycosylase active on G-A mispairs. *Proceedings of the National Academy of Sciences of the United States of America, 86*(22), 8877–8881.

Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L., & Leibler, S. (2004). Bacterial persistence as a phenotypic switch. *Science, 305*(5690), 1622–1625.

Balagadde, F. K., Song, H., Ozaki, J., Collins, C. H., Barnet, M., Arnold, F. H., . . You, L. (2008). A synthetic Escherichia coli predator-prey ecosystem. *Molecular Systems Biology, 4*, 187.

Balagadde, F. K., You, L., Hansen, C. L., Arnold, F. H., & Quake, S. R. (2005). Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. *Science, 309*(5731), 137–140.

Barrick, J. E., Colburn, G., Deatherage, D. E., Traverse, C. C., Strand, M. D., Borges, J. J., . . . Meyer, A. G. (2014). Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. *BMC Genomics, 15*, 1039.

Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., . . . Kim, J. F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature, 461*(7268), 1243–1247.

Bassler, B. L., & Losick, R. (2006). Bacterially speaking. *Cell, 125*(2), 237–246.

Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H., & Weiss, R. (2005). A synthetic multicellular system for programmed pattern formation. *Nature, 434*(7037), 1130–1134.

Bayer, T. S., Widmaier, D. M., Temme, K., Mirsky, E. A., Santi, D. V., & Voigt, C. A. (2009). Synthesis of methyl halides from biomass using engineered microbes. *Journal of the American Chemical Society, 131*(18), 6508–6515.

Benomar, S., Ranava, D., Cardenas, M. L., Trably, E., Rafrafi, Y., Ducret, A., . . . Giudici-Orticoni, M. T. (2015). Nutritional stress induces exchange of cell material and energetic coupling between bacterial species. *Nature Communications, 6*, 6283.

Biers, E. J., Sun, S., & Howard, E. C. (2009). Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. *Applied and Environmental Microbiology, 75*(7), 2221–2229.

Bizukojc, M., Dietz, D., Sun, J., & Zeng, A. P. (2010). Metabolic modelling of syntrophic-like growth of a 1,3-propanediol producer, Clostridium butyricum, and a methanogenic archeon, Methanosarcina mazei, under anaerobic conditions. *Bioprocess and Biosystems Engineering, 33*(4), 507–523.

Blauwkamp, T. A., & Ninfa, A. J. (2002). Physiological role of the GlnK signal transduction protein of Escherichia coli: survival of nitrogen starvation. *Molecular Microbiology, 46*(1), 203–214.

Blomfield, I. C., Calie, P. J., Eberhardt, K. J., McClain, M. S., & Eisenstein, B. I. (1993). Lrp stimulates phase variation of type 1 fimbriation in Escherichia coli K-12. *J Bacteriol, 175*(1), 27–36.

Bochner, B. R., Gadzinski, P., & Panomitros, E. (2001). Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Research, 11*(7), 1246–1255.

Borths, E. L., Poolman, B., Hvorup, R. N., Locher, K. P., & Rees, D. C. (2005). In vitro functional characterization of BtuCD-F, the Escherichia coli ABC transporter for vitamin B12 uptake. *Biochemistry, 44*(49), 16301–16309.

Boylan, S. A., & Dekker, E. E. (1981). L-threonine dehydrogenase. Purification and properties of the homogeneous enzyme from Escherichia coli K12. *Journal of Biological Chemistry, 256*(4), 1809–1815.

Brenner, K., & Arnold, F. H. (2011). Self-organization, layered structure, and aggregation enhance persistence of a synthetic biofilm consortium. *PLoS One, 6*(2), e16791.

Brenner, K., Karig, D. K., Weiss, R., & Arnold, F. H. (2007). Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. *Proceedings of the National Academy of Sciences of the United States of America, 104*(44), 17300–17304.

Brenner, K., You, L., & Arnold, F. H. (2008). Engineering microbial consortia: a new frontier in synthetic biology. *Trends in Biotechnology, 26*(9), 483–489.

Bull, J. J., & Harcombe, W. R. (2009). Population dynamics constrain the cooperative evolution of cross-feeding. *PLoS One, 4*(1), e4115.

Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., & Thomas, T. (2011). Bacterial community assembly based on functional genes rather than species. *Proceedings of the National Academy of Sciences of the United States of America, 108*(34), 14288–14293.

Burkovski, A., & Kramer, R. (2002). Bacterial amino acid transport proteins: occurrence, functions, and significance for biotechnological applications. *Applied and Environmental Microbiology, 58*(3), 265–274.

Burrus, V., & Waldor, M. K. (2004). Shaping bacterial genomes with integrative and conjugative elements. *Research in Microbiology, 155*(5), 376–386.

Canton, B., Labno, A., & Endy, D. (2008). Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology, 26*(7), 787–793.

Carr, P. A., & Church, G. M. (2009). Genome engineering. *Nature Biotechnology, 27*(12), 1151–1162.

Caspi, R., Altman, T., Dreher, K., Fulcher, C. A., Subhraveti, P., Keseler, I. M., . . . Karp, P. D. (2012). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research, 40*(Database issue), D742–753.

Chakrabarti, A. C. (1994). Permeability of membranes to amino acids and modified amino acids: mechanisms involved in translocation. *Amino Acids, 6*, 213–229.

Chakrabarti, A. C., & Deamer, D. W. (1992). Permeability of lipid bilayers to amino acids and phosphate. *Biochimica et Biophysica Acta, 1111*(2), 171–177.

Chen, W. W., Niepel, M., & Sorger, P. K. (2010). Classic and contemporary approaches to modeling biochemical reactions. *Genes and Development, 24*(17), 1861–1875.

Chou, H. H., Chiu, H. C., Delaney, N. F., Segre, D., & Marx, C. J. (2011). Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science, 332*(6034), 1190–1192.

Chow, S. S., Wilke, C. O., Ofria, C., Lenski, R. E., & Adami, C. (2004). Adaptive radiation from resource competition in digital organisms. *Science, 305*(5680), 84–86.

Chuang, J. S., Rivoire, O., & Leibler, S. (2009). Simpson's paradox in a synthetic microbial system. *Science, 323*(5911), 272–275.

Chubiz, L. M., Lee, M. C., Delaney, N. F., & Marx, C. J. (2012). FREQ-Seq: a rapid, cost-effective, sequencing-based method to determine allele frequencies directly from mixed populations. *PLoS One, 7*(10), e47959.

Church, M. J., Jenkins, B. D., Karl, D. M., & Zehr, J. P. (2005). Vertical distributions of nitrogen-fixing phylotypes at Stn ALOHA in the oligotrophic North Pacific Ocean. *Aquatic Microbial Ecology, 38*(1), 3–14.

Conrad, T. M., Frazier, M., Joyce, A. R., Cho, B. K., Knight, E. M., Lewis, N. E., . . . Palsson, B. O. (2010). RNA polymerase mutants found through adaptive evolution reprogram Escherichia coli for optimal growth in minimal media. *Proceedings of the National Academy of Sciences of the United States of America, 107*(47), 20500–20505.

Conrad, T. M., Joyce, A. R., Applebee, M. K., Barrett, C. L., Xie, B., Gao, Y., & Palsson, B. O. (2009). Whole-genome resequencing of Escherichia coli K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. *Genome Biology, 10*(10), R118.

Conrad, T. M., Lewis, N. E., & Palsson, B. O. (2011). Microbial laboratory evolution in the era of genome-scale science. *Molecular Systems Biology, 7*, 509.

Cooper, V. S., & Lenski, R. E. (2000). The population genetics of ecological specialization in evolving Escherichia coli populations. *Nature, 407*(6805), 736–739.

Cordero, O. X., Ventouras, L. A., DeLong, E. F., & Polz, M. F. (2012). Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proceedings of the National Academy of Sciences of the United States of America, 109*(49), 20059–20064.

Coves, J., Niviere, V., Eschenbrenner, M., & Fontecave, M. (1993). NADPH-sulfite reductase from Escherichia coli. A flavin reductase participating in the generation of the free radical of ribonucleotide reductase. *Journal of Biological Chemistry, 268*(25), 18604–18609.

Cruz-Ramos, H., Cook, G. M., Wu, G., Cleeter, M. W., & Poole, R. K. (2004). Membrane topology and mutational analysis of Escherichia coli CydDC, an ABC-type cysteine exporter required for cytochrome assembly. *Microbiology, 150*(Pt 10), 3415–3427.

Curtiss, R., 3rd. (1978). Biological containment and cloning vector transmissibility. *Journal of Infectious Diseases, 137*(5), 668–675.

Czaran, T. L., Hoekstra, R. F., & Pagie, L. (2002). Chemical warfare between microbes promotes biodiversity. *Proceedings of the National Academy of Sciences of the United States of America, 99*(2), 786–790.

Danino, T., Mondragon-Palomino, O., Tsimring, L., & Hasty, J. (2010). A synchronized quorum of genetic clocks. *Nature, 463*(7279), 326–330.

Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America, 97*(12), 6640–6645.

Datta, S., Costantino, N., Zhou, X., & Court, D. L. (2008). Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proceedings of the National Academy of Sciences of the United States of America, 105*(5), 1626–1631.

Deatherage, D. E., & Barrick, J. E. (2014). Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods in Molecular Biology, 1151*, 165–188.

Dethlefsen, L., McFall-Ngai, M., & Relman, D. A. (2007). An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature, 449*(7164), 811–818.

Doroshenko, V., Airich, L., Vitushkina, M., Kolokolova, A., Livshits, V., & Mashko, S. (2007). YddG from Escherichia coli promotes export of aromatic amino acids. *FEMS Microbiology Letters, 275*(2), 312–318.

Duan, F., & March, J. C. (2010). Engineered bacterial communication prevents Vibrio cholerae virulence in an infant mouse model. *Proceedings of the National Academy of Sciences of the United States of America, 107*(25), 11260–11264.

Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., . . . Palsson, B. O. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences of the United States of America, 104*(6), 1777–1782.

Duarte, N. C., Herrgard, M. J., & Palsson, B. O. (2004). Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research, 14*(7), 1298–1309.

Dubey, G. P., & Ben-Yehuda, S. (2011). Intercellular nanotubes mediate bacterial communication. *Cell, 144*(4), 590–600.

Dyall, S. D., Brown, M. T., & Johnson, P. J. (2004). Ancient invasions: from endosymbionts to organelles. *Science, 304*(5668), 253–257.

Eggeling, L., & Sahm, H. (2003). New ubiquitous translocators: amino acid export by Corynebacterium glutamicum and Escherichia coli. *Archives of Microbiology, 180*(3), 155–160.

Eiteman, M. A., Lee, S. A., & Altman, E. (2008). A co-fermentation strategy to consume sugar mixtures effectively. *Journal of Biological Engineering, 2*, 3.

Ellis, H. M., Yu, D., DiTizio, T., & Court, D. L. (2001). High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-

stranded oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America, 98*(12), 6742–6746.

Endy, D. (2005). Foundations for engineering biology. *Nature, 438*(7067), 449–453.

Enyeart, P. J., Simpson, Z. B., & Ellington, A. D. (2015). A microbial model of economic trading and comparative advantage. *Journal of Theoretical Biology, 364*, 326–343.

Estrela, S., & Gudelj, I. (2010). Evolution of cooperative cross-feeding could be less challenging than originally thought. *PLoS One, 5*(11), e14121.

Faith, J. J., McNulty, N. P., Rey, F. E., & Gordon, J. I. (2011). Predicting a human gut microbiota's response to diet in gnotobiotic mice. *Science, 333*(6038), 101–104.

Faith, J. J., Rey, F. E., O'Donnell, D., Karlsson, M., McNulty, N. P., Kallstrom, G., .Gordon, J. I. (2010). Creating and characterizing communities of human gut microbes in gnotobiotic mice. *ISME Journal, 4*(9), 1094–1098.

Falkowski, P. G., Fenchel, T., & Delong, E. F. (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science, 320*(5879), 1034–1039.

Feist, A. M., Herrgard, M. J., Thiele, I., Reed, J. L., & Palsson, B. O. (2009). Reconstruction of biochemical networks in microorganisms. *Nature Reviews. Microbiology, 7*(2), 129–143.

Fernandez, A., Huang, S., Seston, S., Xing, J., Hickey, R., Criddle, C., & Tiedje, J. (1999). How stable is stable? Function versus community composition. *Applied and Environmental Microbiology, 65*(8), 3697–3704.

Fong, S. S., Joyce, A. R., & Palsson, B. O. (2005). Parallel adaptive evolution cultures of Escherichia coli lead to convergent growth phenotypes with different gene expression states. *Genome Research, 15*(10), 1365–1372.

Foster, K. R. (2011). The sociobiology of molecular systems. *Nature Reviews. Genetics, 12*(3), 193–203.

Franke, I., Resch, A., Dassler, T., Maier, T., & Bock, A. (2003). YfiK from Escherichia coli promotes export of O-acetylserine and cysteine. *Journal of Bacteriology, 185*(4), 1161–1166.

Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G., & Hanage, W. P. (2009). The bacterial species challenge: making sense of genetic and ecological diversity. *Science, 323*(5915), 741–746.

Frost, L. S., Leplae, R., Summers, A. O., & Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nature Reviews. Microbiology, 3*(9), 722–732.

Fux, C. A., Costerton, J. W., Stewart, P. S., & Stoodley, P. (2005). Survival strategies of infectious biofilms. *Trends in Microbiology, 13*(1), 34–40.

Gally, D. L., Bogan, J. A., Eisenstein, B. I., & Blomfield, I. C. (1993). Environmental regulation of the fim switch controlling type 1 fimbrial phase variation in Escherichia coli K-12: effects of temperature and media. *Journal of Bacteriology, 175*(19), 6186–6193.

Gardner, J. F. (1979). Regulation of the threonine operon: tandem threonine and isoleucine codons in the control region and translational control of transcription termination. *Proceedings of the National Academy of Sciences of the United States of America, 76*(4), 1706–1710.

Gardner, M. (1970). The fantastic combinations of John Conway's new solitare game "life". *Scientific American* (223), 120–123.

Gawronski, J. D., Wong, S. M., Giannoukos, G., Ward, D. V., & Akerley, B. J. (2009). Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proceedings of the National Academy of Sciences of the United States of America, 106*(38), 16422–16427.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods, 6*(5), 343–345.

Giongo, A., Gano, K. A., Crabb, D. B., Mukherjee, N., Novelo, L. L., Casella, G., . . . Triplett, E. W. (2011). Toward defining the autoimmune microbiome for type 1 diabetes. *ISME Journal, 5*(1), 82–91.

Gogarten, J. P., & Townsend, J. P. (2005). Horizontal gene transfer, genome innovation and evolution. *Nature Reviews. Microbiology, 3*(9), 679–687.

Goh, C. S., Gianoulis, T. A., Liu, Y., Li, J., Paccanaro, A., Lussier, Y. A., & Gerstein, M. (2006). Integration of curated databases to identify genotype-phenotype associations. *BMC Genomics, 7*, 257.

Goldberg, A., Fridman, O., Ronin, I., & Balaban, N. Q. (2014). Systematic identification and quantification of phase variation in commensal and pathogenic Escherichia coli. *Genome Medicine, 6*(11), 112.

Goodarzi, H., Hottes, A. K., & Tavazoie, S. (2009). Global discovery of adaptive mutations. *Nature Methods, 6*(8), 581–583.

Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., Mitra, R. D., Lozupone, C. A., .Gordon, J. I. (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host & Microbe, 6*(3), 279–289.

Gore, J., Youk, H., & van Oudenaarden, A. (2009). Snowdrift game dynamics and facultative cheating in yeast. *Nature, 459*(7244), 253–256.

Goyal, G., Tsai, S. L., Madan, B., DaSilva, N. A., & Chen, W. (2011). Simultaneous cell growth and ethanol production from cellulose by an engineered yeast consortium displaying a functional mini-cellulosome. *Microbial Cell Factories, 10*, 89.

Gustafsson, C., & Persson, B. C. (1998). Identification of the rrmA gene encoding the 23S rRNA m1G745 methyltransferase in Escherichia coli and characterization of an m1G745-deficient mutant. *Journal of Bacteriology, 180*(2), 359–365.

Hall-Stoodley, L., Costerton, J. W., & Stoodley, P. (2004). Bacterial biofilms: from the natural environment to infectious diseases. *Nature Reviews. Microbiology, 2*(2), 95–108.

Hall-Stoodley, L., & Stoodley, P. (2002). Developmental regulation of microbial biofilms. *Current Opinion in Biotechnology, 13*(3), 228–233.

Hamady, M., & Knight, R. (2009). Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Research, 19*(7), 1141–1152.

Hansen, A. K., & Moran, N. A. (2011). Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proceedings of the National Academy of Sciences of the United States of America, 108*(7), 2849–2854.

Harcombe, W. (2010). Novel cooperation experimentally evolved between species. *Evolution, 64*(7), 2166–2172.

Hasman, H., Chakraborty, T., & Klemm, P. (1999). Antigen-43-mediated autoaggregation of Escherichia coli is blocked by fimbriation. *Journal of Bacteriology, 181*(16), 4834–4841.

Hauert, C., Michor, F., Nowak, M. A., & Doebeli, M. (2006). Synergy and discounting of cooperation in social dilemmas. *Journal of Theoretical Biology, 239*(2), 195–202.

Hayes, C. S., Aoki, S. K., & Low, D. A. (2010). Bacterial contact-dependent delivery systems. *Annual Review of Genetics, 44*, 71–90.

Hayes, F. (2003). Transposon-based strategies for microbial functional genomics and proteomics. *Annual Review of Genetics, 37*, 3–29.

He, T., Venema, K., Priebe, M. G., Welling, G. W., Brummer, R. J., & Vonk, R. J. (2008). The role of colonic metabolism in lactose intolerance. *European Journal of Clinical Investigation, 38*(8), 541–547.

Hemm, M. R., Paul, B. J., Miranda-Rios, J., Zhang, A., Soltanzad, N., & Storz, G. (2010). Small stress response proteins in Escherichia coli: proteins missed by classical proteomic studies. *Journal of Bacteriology, 192*(1), 46–58.

Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews. Microbiology, 8*(1), 15–25.

Hori, H., Yoneyama, H., Tobe, R., Ando, T., Isogai, E., & Katsumata, R. (2011). Inducible L-alanine exporter encoded by the novel gene ygaW (alaE) in Escherichia coli. *Applied and Environmental Microbiology, 77*(12), 4027–4034.

Hu, B., Du, J., Zou, R. Y., & Yuan, Y. J. (2010). An environment-sensitive synthetic microbial ecosystem. *PLoS One, 5*(5), e10619.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., . . . Yong, S. Y. (2011). InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Research, 40*(Database issue), D306–312.

Hussein, M. J., Green, J. M., & Nichols, B. P. (1998). Characterization of mutations that allow p-aminobenzoyl-glutamate utilization by Escherichia coli. *Journal of Bacteriology, 180*(23), 6260–6268.

Isaacs, F. J., Carr, P. A., Wang, H. H., Lajoie, M. J., Sterling, B., Kraal, L., . . . Church, G. M. (2011). Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science, 333*(6040), 348–353.

Ispolatov, I., Ackermann, M., & Doebeli, M. (2012). Division of labour and the evolution of multicellularity. *Proceedings. Biological Sciences / The Royal Society, 279*(1734), 1768–1776.

Johnson, J. R., Clabots, C., & Rosen, H. (2006). Effect of inactivation of the global oxidative stress regulator oxyR on the colonization ability of Escherichia coli O1:K1:H7 in a mouse model of ascending urinary tract infection. *Infection and Immunity, 74*(1), 461–468.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research, 40*(Database issue), D109–114.

Katashkina, J. I., Hara, Y., Golubeva, L. I., Andreeva, I. G., Kuvaeva, T. M., & Mashko, S. V. (2009). Use of the lambda Red-recombineering method for genetic engineering of Pantoea ananatis. *BMC Molecular Biology, 10*, 34.

Kato, S., Haruta, S., Cui, Z. J., Ishii, M., & Igarashi, Y. (2005). Stable coexistence of five bacterial strains as a cellulose-degrading community. *Applied and Environmental Microbiology, 71*(11), 7099–7106.

Kato, S., Haruta, S., Cui, Z. J., Ishii, M., & Igarashi, Y. (2008). Network relationships of bacteria in a stable mixed culture. *Microbial Ecology, 56*(3), 403–411.

Kellermayer, R., Dowd, S. E., Harris, R. A., Balasa, A., Schaible, T. D., Wolcott, R. D., . . . Smith, C. W. (2011). Colonic mucosal DNA methylation, immune response, and microbiome patterns in Toll-like receptor 2-knockout mice. *FASEB Journal, 25*(5), 1449–1460.

Kennerknecht, N., Sahm, H., Yen, M. R., Patek, M., Saier Jr, M. H., Jr., & Eggeling, L. (2002). Export of L-isoleucine from Corynebacterium glutamicum: a two-gene-encoded member of a new translocator family. *Journal of Bacteriology, 184*(14), 3947–3956.

Kerner, A., Park, J., Williams, A., & Lin, X. N. (2012). A programmable Escherichia coli consortium via tunable symbiosis. *PLoS One, 7*(3), e34032.

Khalil, A. S., & Collins, J. J. (2010). Synthetic biology: applications come of age. *Nature Reviews. Genetics, 11*(5), 367–379.

Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E., & Cooper, T. F. (2011). Negative epistasis between beneficial mutations in an evolving bacterial population. *Science, 332*(6034), 1193–1196.

Kibota, T. T., & Lynch, M. (1996). Estimate of the genomic mutation rate deleterious to overall fitness in E. coli. *Nature, 381*(6584), 694–696.

Kiers, E. T., Duhamel, M., Beesetty, Y., Mensah, J. A., Franken, O., Verbruggen, E., . . . Bucking, H. (2011). Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science, 333*(6044), 880–882.

Kim, H. J., Boedicker, J. Q., Choi, J. W., & Ismagilov, R. F. (2008). Defined spatial structure stabilizes a synthetic multispecies bacterial community. *Proceedings of the National Academy of Sciences of the United States of America, 105*(47), 18188–18193.

Kirkup, B. C., & Riley, M. A. (2004). Antibiotic-mediated antagonism leads to a bacterial game of rock-paper-scissors in vivo. *Nature, 428*(6981), 412–414.

Klein-Marcuschamer, D., Santos, C. N., Yu, H., & Stephanopoulos, G. (2009). Mutagenesis of the bacterial RNA polymerase alpha subunit for improvement of complex phenotypes. *Applied and Environmental Microbiology, 75*(9), 2705–2711. doi:10.1128/AEM.01888-08

Klitgord, N., & Segre, D. (2010). Environments that induce synthetic microbial ecosystems. *PLoS Computational Biology, 6*(11), e1001002.

Klitgord, N., & Segre, D. (2011). Ecosystems biology of microbial metabolism. *Current Opininon in Biotechnology, 22*(4), 541–546.

Knudsen, S., Saadbye, P., Hansen, L. H., Collier, A., Jacobsen, B. L., Schlundt, J., & Karlstrom, O. H. (1995). Development and testing of improved suicide functions for biological containment of bacteria. *Applied and Environmental Microbiology, 61*(3), 985–991.

Koide, T., Pang, W. L., & Baliga, N. S. (2009). The role of predictive modelling in rationally re-engineering biological systems. *Nature Reviews. Microbiology, 7*(4), 297–305.

Kosuri, S., Goodman, D. B., Cambray, G., Mutalik, V. K., Gao, Y., Arkin, A. P., . . . Church, G. M. (2013). Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America, 110*(34), 14024–14029.

Kow, Y. W., Bao, G., Reeves, J. W., Jinks-Robertson, S., & Crouse, G. F. (2007). Oligonucleotide transformation of yeast reveals mismatch repair complexes to be differentially active on DNA replication strands. *Proceedings of the National Academy of Sciences of the United States of America, 104*(27), 11352–11357.

Kutukova, E. A., Livshits, V. A., Altman, I. P., Ptitsyn, L. R., Zyiatdinov, M. H., Tokmakova, I. L., & Zakataeva, N. P. (2005). The yeaS (leuE) gene of Escherichia coli encodes an exporter of leucine, and the Lrp protein regulates its expression. *FEBS Letters, 579*(21), 4629–4634.

Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H. R., Haimovich, A. D., Kuznetsov, G., . . . Isaacs, F. J. (2013). Genomically recoded organisms expand biological functions. *Science, 342*(6156), 357–360.

Langridge, G. C., Phan, M. D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., Turner, A. K. (2009). Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Research, 19*(12), 2308–2316.

Lenski, R. E., Ofria, C., Pennock, R. T., & Adami, C. (2003). The evolutionary origin of complex features. *Nature, 423*(6936), 139–144.

Letunic, I., & Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research, 39*(Web Server issue), W475–478.

Lewis, N. E., Nagarajan, H., & Palsson, B. O. (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature Reviews. Microbiology, 10*(4), 291–305.

Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America, 102*(31), 11070–11075.

Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Microbial ecology: human gut microbes associated with obesity. *Nature, 444*(7122), 1022–1023.

Li, L., Yang, C., Lan, W., Xie, S., Qiao, C., & Liu, J. (2008). Removal of methyl parathion from artificial off-gas using a bioreactor containing a constructed microbial consortium. *Environmental Science and Technology, 42*(6), 2136–2141.

Li, Q., & Wu, Y. J. (2009). A fluorescent, genetically engineered microorganism that degrades organophosphates and commits suicide when required. *Applied Microbiology and Biotechnology, 82*(4), 749–756.

Li, X. Z., & Nikaido, H. (2009). Efflux-mediated drug resistance in bacteria: an update. *Drugs, 69*(12), 1555–1623.

Link, A. J., Jeong, K. J., & Georgiou, G. (2007). Beyond toothpicks: new methods for isolating mutant bacteria. *Nature Reviews. Microbiology, 5*(9), 680–688.

Lu, T. K., Khalil, A. S., & Collins, J. J. (2009). Next-generation synthetic gene networks. *Nature Biotechnology, 27*(12), 1139–1150.

Lutz, R., & Bujard, H. (1997). Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research, 25*(6), 1203–1210.

Mackelprang, R., Waldrop, M. P., DeAngelis, K. M., David, M. M., Chavarria, K. L., Blazewicz, S. J., . . . Jansson, J. K. (2011). Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature, 480*(7377), 368–371.

MacLean, D., Jones, J. D., & Studholme, D. J. (2009). Application of 'next-generation' sequencing technologies to microbial genetics. *Nature Reviews. Microbiology, 7*(4), 287–296.

Mahadevan, R., Edwards, J. S., & Doyle, F. J., 3rd. (2002). Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophysical Journal, 83*(3), 1331–1340.

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Dore, J. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut, 55*(2), 205–211.

Margulis, L. (1971). The origin of plant and animal cells. *American Scientist, 59*(2), 230–235.

Markowitz, V. M., Chen, I. M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Kyrpides, N. C. (2012). IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Research, 40*(Database issue), D115–122.

Mazumdar, V., Amar, S., & Segre, D. (2013). Metabolic proximity in the order of colonization of a microbial community. *PLoS One, 8*(10), e77617.

McCutcheon, J. P., McDonald, B. R., & Moran, N. A. (2009). Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proceedings of the National Academy of Sciences of the United States of America, 106*(36), 15394–15399.

McCutcheon, J. P., & Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nature Reviews. Microbiology, 10*(1), 13–26.

McCutcheon, J. P., & von Dohlen, C. D. (2011). An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Current Biology, 21*(16), 1366–1372.

Medema, M. H., van Raaphorst, R., Takano, E., & Breitling, R. (2012). Computational tools for the synthetic design of biochemical pathways. *Nature Reviews. Microbiology, 10*(3), 191–202.

Medini, D., Serruto, D., Parkhill, J., Relman, D. A., Donati, C., Moxon, R., . . . Rappuoli, R. (2008). Microbiology in the post-genomic era. *Nature Reviews. Microbiology, 6*(6), 419–430.

Mee, M. T., Collins, J. J., Church, G. M., & Wang, H. H. (2014). Syntrophic exchange in synthetic microbial communities. *Proceedings of the National Academy of Sciences of the United States of America, 111*(20), E2149–2156.

Mee, M. T., & Wang, H. H. (2012). Engineering ecosystems and synthetic ecologies. *Molecular bioSystems, 8*(10), 2470–2483.

Mi, H., Muruganujan, A., & Thomas, P. D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research, 41*(Database issue), D377–386.

Minty, J. J., Singer, M. E., Scholz, S. A., Bae, C. H., Ahn, J. H., Foster, C. E., . . . Lin, X. N. (2013). Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. *Proceedings of the National Academy of Sciences of the United States of America, 110*(36), 14592–14597.

Moffitt, J. R., Lee, J. B., & Cluzel, P. (2012). The single-cell chemostat: an agarose-based, microfluidic device for high-throughput, single-cell studies of bacteria and bacterial communities. *Lab on a Chip, 12*(8), 1487–1494.

Molin, S., Klemm, P., Poulsen, L. K., Biehl, H., Gerdes, K., & Andersson, P. (1987). Conditional Suicide System for Containment of Bacteria and Plasmids. *Nature Biotechnology, 5*, 1315 – 1318.

Monera, O. D., Sereda, T. J., Zhou, N. E., Kay, C. M., & Hodges, R. S. (1995). Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. *Journal of Peptide Science, 1*(5), 319–329.

Morris, J. J., Lenski, R. E., & Zinser, E. R. (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *mBio, 3*(2).

Mukai, T., Hayashi, A., Iraha, F., Sato, A., Ohtake, K., Yokoyama, S., & Sakamoto, K. (2010). Codon reassignment in the Escherichia coli genetic code. *Nucleic Acids Research, 38*(22), 8188–8195.

Muller, K. M., & Arndt, K. M. (2012). Standardization in synthetic biology. *Methods in Molecular Biology, 813*, 23–43.

Muse, W. B., & Bender, R. A. (1998). The nac (nitrogen assimilation control) gene from Escherichia coli. *Journal of Bacteriology, 180*(5), 1166–1173.

Nadell, C. D., Foster, K. R., & Xavier, J. B. (2010). Emergence of spatial structure in cell groups and the evolution of cooperation. *PLoS Computational Biology, 6*(3), e1000716.

Nandineni, M. R., & Gowrishankar, J. (2004). Evidence for an arginine exporter encoded by yggA (argO) that is regulated by the LysR-type transcriptional regulator ArgP in Escherichia coli. *Journal of Bacteriology, 186*(11), 3539–3546.

Nell, S., Suerbaum, S., & Josenhans, C. (2010). The impact of the microbiota on the pathogenesis of IBD: lessons from mouse infection models. *Nature Reviews. Microbiology, 8*(8), 564–577.

Nicholson, J. K., Holmes, E., & Wilson, I. D. (2005). Gut microorganisms, mammalian metabolism and personalized health care. *Nature Reviews. Microbiology, 3*(5), 431–438.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science, 314*(5805), 1560–1563.

Nuoffer, C., Zanolari, B., & Erni, B. (1988). Glucose permease of Escherichia coli. The effect of cysteine to serine mutations on the function, stability,

and regulation of transport and phosphorylation. *Journal of Biological Chemistry, 263*(14), 6647–6655.

Ofria, C., Brown CT, Adami C. (1998). The Avida Technical Manual. *Introduction to Artificial Life*, 297–350.

Ogawa, W., Kim, Y. M., Mizushima, T., & Tsuchiya, T. (1998). Cloning and expression of the gene for the Na+-coupled serine transporter from Escherichia coli and characteristics of the transporter. *Journal of Bacteriology, 180*(24), 6749–6752.

Oh, Y. K., Palsson, B. O., Park, S. M., Schilling, C. H., & Mahadevan, R. (2007). Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. *Journal of Biological Chemistry, 282*(39), 28791–28799.

Pande, S., Merker, H., Bohl, K., Reichelt, M., Schuster, S., de Figueiredo, L. F., . . . Kost, C. (2014). Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *ISME Journal, 8*(5), 953-62. doi: 10.1038/ismej.2013.211.

Pande, S., Shitut, S., Freund, L., Westermann, M., Bertels, F., Colesie, C., . . . Kost, C. (2015). Metabolic cross-feeding via intercellular nanotubes among bacteria. *Nature Communications, 6*, 6238.

Papadopoulos, D., Schneider, D., Meier-Eiss, J., Arber, W., Lenski, R. E., & Blot, M. (1999). Genomic evolution during a 10,000-generation experiment with bacteria. *Proceedings of the National Academy of Sciences of the United States of America, 96*(7), 3807–3812.

Papin, J. A., Stelling, J., Price, N. D., Klamt, S., Schuster, S., & Palsson, B. O. (2004). Comparison of network-based pathway analysis methods. *Trends in Biotechnology, 22*(8), 400–405.

Park, J., Kerner, A., Burns, M. A., & Lin, X. N. (2011). Microdroplet-enabled highly parallel co-cultivation of microbial communities. *PLoS One, 6*(2), e17019.

Pasotti, L., Zucca, S., Lupotto, M., Cusella De Angelis, M. G., & Magni, P. (2011). Characterization of a synthetic bacterial self-destruction device for programmed cell death and for recombinant proteins release. *Journal of Biological Engineering, 5*, 8.

Patzlaff, J. S., van der Heide, T., & Poolman, B. (2003). The ATP/substrate stoichiometry of the ATP-binding cassette (ABC) transporter OpuA. *Journal of Biological Chemistry, 278*(32), 29546–29551.

Peeters, E., Nguyen Le Minh, P., Foulquie-Moreno, M., & Charlier, D. (2009). Competitive activation of the Escherichia coli argO gene coding for an arginine exporter by the transcriptional regulators Lrp and ArgP. *Molecular Microbiology, 74*(6), 1513–1526.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods, 8*(10), 785–786.

Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J. A., . . . Guyer, M. (2009). The NIH Human Microbiome Project. *Genome Research, 19*(12), 2317–2323.

Phadtare, S., & Inouye, M. (1999). Sequence-selective interactions with RNA by CspB, CspC and CspE, members of the CspA family of Escherichia coli. *Molecular Microbiology, 33*(5), 1004–1014.

Plumbridge, J. (2002). Regulation of gene expression in the PTS in Escherichia coli: the role and interactions of Mlc. *Current Opinion in Microbiology, 5*(2), 187–193.

Pocock, M. J., Evans, D. M., & Memmott, J. (2012). The robustness and restoration of a network of ecological networks. *Science, 335*(6071), 973–977.

Price, N. D., Reed, J. L., Papin, J. A., Wiback, S. J., & Palsson, B. O. (2003). Network-based analysis of metabolic regulation in the human red blood cell. *Journal of Theoretical Biology, 225*(2), 185–194.

Prindle, A., Samayoa, P., Razinkov, I., Danino, T., Tsimring, L. S., & Hasty, J. (2012). A sensing array of radically coupled genetic 'biopixels'. *Nature, 481*(7379), 39–44.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., . . . Finn, R. D. (2012). The Pfam protein families database. *Nucleic Acids Research, 40*(Database issue), D290–301.

Raes, J., & Bork, P. (2008). Molecular eco-systems biology: towards an understanding of community function. *Nature Reviews. Microbiology, 6*(9), 693–699.

Rath, C. M., & Dorrestein, P. C. (2012). The bacterial chemical repertoire mediates metabolic exchange within gut microbiomes. *Current Opinion in Microbiology, 15*(2), 147-54. doi: 10.1016/j.mib.2011.12.009.

Ray, T. S. (1992). An Approach to the Synthesis of Life. *In Artificial Life II*, 371–408.

Reed, J. L., Vo, T. D., Schilling, C. H., & Palsson, B. O. (2003). An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biology, 4*(9), R54.

Regot, S., Macia, J., Conde, N., Furukawa, K., Kjellen, J., Peeters, T., . . . Sole, R. (2011). Distributed biological computation with multicellular engineered networks. *Nature, 469*(7329), 207–211.

Riley, M., Abe, T., Arnaud, M. B., Berlyn, M. K., Blattner, F. R., Chaudhuri, R. R., . . . Wanner, B. L. (2006). Escherichia coli K-12: a cooperatively developed annotation snapshot--2005. *Nucleic Acids Research, 34*(1), 1–9.

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research, 22*(5), 939–946.

Ronchel, M. C., & Ramos, J. L. (2001). Dual system to reinforce biological containment of recombinant bacteria designed for rhizoremediation. *Applied and Environmental Microbiology, 67*(6), 2649–2656.

Ruby, E. G. (2008). Symbiotic conversations are revealed under genetic interrogation. *Nature Reviews. Microbiology, 6*(10), 752–762.

Russell, C. W., Bouvaine, S., Newell, P. D., & Douglas, A. E. (2013). Shared metabolic pathways in a coevolved insect-bacterial symbiosis. *Applied and Environmental Microbiology, 79*(19), 6117–6123.

Sachs, J. L., Mueller, U. G., Wilcox, T. P., & Bull, J. J. (2004). The evolution of cooperation. *Quarterly Review of Biology, 79*(2), 135–160.

Saeidi, N., Wong, C. K., Lo, T. M., Nguyen, H. X., Ling, H., Leong, S. S., . . . Chang, M. W. (2011). Engineering microbes to sense and eradicate Pseudomonas aeruginosa, a human pathogen. *Molecular Systems Biology, 7*, 521.

Sahasrabudhe, S., & Motter, A. E. (2011). Rescuing ecosystems from extinction cascades through compensatory perturbations. *Nature Communications, 2*, 170.

Saier, M. H., Jr. (2000). A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiology and Molecular Biology Reviews, 64*(2), 354–411.

Schilling, C. H., Covert, M. W., Famili, I., Church, G. M., Edwards, J. S., & Palsson, B. O. (2002). Genome-scale metabolic model of Helicobacter pylori 26695. *Journal of Bacteriology, 184*(16), 4582–4593.

Schilling, C. H., & Palsson, B. O. (1998). The underlying pathway structure of biochemical reaction networks. *Proceedings of the National Academy of Sciences of the United States of America, 95*(8), 4193–4198.

Schloss, P. D., & Handelsman, J. (2004). Status of the microbial census. *Microbiology and Molecular Biology Reviews, 68*(4), 686–691.

Shank, E. A., & Kolter, R. (2011). Extracellular signaling and multicellularity in Bacillus subtilis. *Current Opinion in Microbiology, 14*(6), 741–747.

Shin, H. D., McClendon, S., Vo, T., & Chen, R. R. (2010). Escherichia coli binary culture engineered for direct fermentation of hemicellulose to a biofuel. *Applied and Environmental Microbiology, 76*(24), 8150–8159.

Shong, J., Jimenez Diaz, M. R., & Collins, C. H. (2012). Towards synthetic microbial consortia for bioprocessing. *Current Opininon in Biotechnology* 23(5):798-802. doi: 10.1016/j.copbio.2012.02.001.

Shou, W., Ram, S., & Vilar, J. M. (2007). Synthetic cooperation in engineered yeast populations. *Proceedings of the National Academy of Sciences of the United States of America, 104*(6), 1877–1882.

Simmons, S. L., Dibartolo, G., Denef, V. J., Goltsman, D. S., Thelen, M. P., & Banfield, J. F. (2008). Population genomic analysis of strain variation in Leptospirillum group II bacteria involved in acid mine drainage formation. *PLoS Biology, 6*(7), e177.

Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., & Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature, 480*(7376), 241–244.

Smith, D. R., & Chapman, M. R. (2010). Economical evolution: microbes reduce the synthetic cost of extracellular proteins. *mBio, 1*(3).

Sokol, H., Seksik, P., Furet, J. P., Firmesse, O., Nion-Larmurier, I., Beaugerie, L., Dore, J. (2009). Low counts of Faecalibacterium prausnitzii in colitis microbiota. *Inflammatory Bowel Diseases, 15*(8), 1183–1189.

Steidler, L., Hans, W., Schotte, L., Neirynck, S., Obermeier, F., Falk, W., . . . Remaut, E. (2000). Treatment of murine colitis by Lactococcus lactis secreting interleukin-10. *Science, 289*(5483), 1352–1355.

Steidler, L., Neirynck, S., Huyghebaert, N., Snoeck, V., Vermeire, A., Goddeeris, B., Remaut, E. (2003). Biological containment of genetically modified Lactococcus lactis for intestinal delivery of human interleukin 10. *Nature Biotechnology, 21*(7), 785–789.

Steidler, L., Rottiers, P., & Coulie, B. (2009). Actobiotics as a novel method for cytokine delivery. *Annals of the New York Academy of Sciences, 1182*, 135–145.

Stolyar, S., Van Dien, S., Hillesland, K. L., Pinel, N., Lie, T. J., Leigh, J. A., & Stahl, D. A. (2007). Metabolic modeling of a mutualistic microbial community. *Molecular Systems Biology, 3*, 92.

Straight, P. D., & Kolter, R. (2009). Interspecies chemical communication in bacterial development. *Annual Review of Microbiology, 63*, 99–118.

Sumantran, V. N., Schweizer, H. P., & Datta, P. (1990). A novel membrane-associated threonine permease encoded by the tdcC gene of Escherichia coli. *Journal of Bacteriology, 172*(8), 4288–4294.

Swingle, B., Bao, Z., Markel, E., Chambers, A., & Cartinhour, S. (2010). Recombineering using RecTE from Pseudomonas syringae. *Applied and Environmental Microbiology, 76*(15), 4960–4968.

Swingle, B., Markel, E., Costantino, N., Bubunenko, M. G., Cartinhour, S., & Court, D. L. (2010). Oligonucleotide recombination in Gram-negative bacteria. *Molecular Microbiology, 75*(1), 138–148.

Tabor, J. J., Salis, H. M., Simpson, Z. B., Chevalier, A. A., Levskaya, A., Marcotte, E. M., Ellington, A. D. (2009). A synthetic genetic edge detection program. *Cell, 137*(7), 1272–1281.

Taffs, R., Aston, J. E., Brileya, K., Jay, Z., Klatt, C. G., McGlynn, S., Carlson, R. P. (2009). In silico approaches to study mass and energy flows in microbial consortia: a syntrophic case study. *BMC Systems Biology, 3*, 114.

Tamsir, A., Tabor, J. J., & Voigt, C. A. (2011). Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature, 469*(7329), 212–215.

Tang, X., He Ly Fau - Tao, X. Q., Tao Xq Fau - Dang, Z., Dang Z Fau - Guo, C. L., Guo Cl Fau - Lu, G. N., Lu Gn Fau - Yi, X. Y., & Yi, X. Y. (2010). Construction of an artificial microalgal-bacterial consortium that efficiently. *Journal of Hazardous Materials, 181*(1–3), 1158–1162

Tatusov, R. L., Koonin, E. V., & Lipman, D. J. (1997). A genomic perspective on protein families. *Science, 278*(5338), 631–637.

Tchieu, J. H., Norris, V., Edwards, J. S., & Saier, M. H., Jr. (2001). The complete phosphotransferase system in Escherichia coli. *Journal of Molecular Microbiology and Biotechnology, 3*(3), 329–346.

Tenaillon, O., Rodriguez-Verdugo, A., Gaut, R. L., McDonald, P., Bennett, A. F., Long, A. D., & Gaut, B. S. (2012). The molecular diversity of adaptive convergence. *Science, 335*(6067), 457–461.

Tolonen, A. C., Chilaka, A. C., & Church, G. M. (2009). Targeted gene inactivation in Clostridium phytofermentans shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Molecular Microbiology, 74*(6), 1300–1313.

Trotschel, C., Deutenberg, D., Bathe, B., Burkovski, A., & Kramer, R. (2005). Characterization of methionine export in Corynebacterium glutamicum. *Journal of Bacteriology, 187*(11), 3786–3794.

Tsai, S. L., Goyal, G., & Chen, W. (2010). Surface display of a functional minicellulosome by intracellular complementation using a synthetic yeast consortium and its application to cellulose hydrolysis and ethanol production. *Applied and Environmental Microbiology, 76*(22), 7514–7520.

Tuite, N. L., Fraser, K. R., & O'Byrne C, P. (2005). Homocysteine toxicity in Escherichia coli is caused by a perturbation of branched-chain amino acid biosynthesis. *Journal of Bacteriology, 187*(13), 4362–4371.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature, 457*(7228), 480–484.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature, 449*(7164), 804–810.

Turnbaugh, P. J., Quince, C., Faith, J. J., McHardy, A. C., Yatsunenko, T., Niazi, F.,Gordon, J. I. (2010). Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proceedings of the National Academy of Sciences of the United States of America, 107*(16), 7503–7508.

Usuda, Y., & Kurahashi, O. (2005). Effects of deregulation of methionine biosynthesis on methionine excretion in Escherichia coli. *Applied and Environmental Microbiology, 71*(6), 3228–3234.

van den Ende, P. (1973). Predator-prey interactions in continuous culture. *Science, 181*(4099), 562–564.

van Kessel, J. C., & Hatfull, G. F. (2007). Recombineering in Mycobacterium tuberculosis. *Nature Methods, 4*(2), 147–152.

van Opijnen, T., Bodi, K. L., & Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods, 6*(10), 767–772.

van Passel, M. W., Marri, P. R., & Ochman, H. (2008). The emergence and fate of horizontally acquired genes in Escherichia coli. *PLoS Computational Biology, 4*(4), e1000059.

van Pijkeren, J. P., & Britton, R. A. (2012). High efficiency recombineering in lactic acid bacteria. *Nucleic Acids Research, 40*(10), e76. doi: 10.1093/nar/gks147.

Varma, A., & Palsson, B. O. (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Applied and Environmental Microbiology, 60*(10), 3724–3731.

Vrljic, M., Sahm, H., & Eggeling, L. (1996). A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum. *Molecular Microbiology, 22*(5), 815–826.

Waks, Z., & Silver, P. A. (2009). Engineering a synthetic dual-organism system for hydrogen production. *Applied and Environmental Microbiology, 75*(7), 1867–1875.

Walk, S. T., & Young, V. B. (2008). Emerging Insights into Antibiotic-Associated Diarrhea and Clostridium difficile Infection through the Lens of Microbial Ecology. *Interdisciplinary Perspectives on Infectious Diseases, 2008*, 125081.

Wang, H., Kim, HB, Cong, L, Bang, D, Church GM. (2012). Genome-scale Promoter Engineering by Co-Selection MAGE. *Nature Methods, in press*.

Wang, H. H., & Church, G. M. (2011). Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Methods in Enzymology, 498*, 409–426.

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., & Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature, 460*(7257), 894–898.

Wang, M. D., Liu, L., Wang, B. M., & Berg, C. M. (1987). Cloning and characterization of the Escherichia coli K-12 alanine-valine transaminase (avtA) gene. *Journal of Bacteriology, 169*(9), 4228–4234.

Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T. H., Stege, J. T., . . . Leadbetter, J. R. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature, 450*(7169), 560–565.

Wells, J. M., & Mercenier, A. (2008). Mucosal delivery of therapeutic and prophylactic molecules using lactic acid bacteria. *Nature Reviews. Microbiology, 6*(5), 349–362.

West, S. A., Griffin, A. S., Gardner, A., & Diggle, S. P. (2006). Social evolution theory for microorganisms. *Nature Reviews. Microbiology, 4*(8), 597–607.

Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E., & Adami, C. (2001). Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature, 412*(6844), 331–333.

Wintermute, E. H., & Silver, P. A. (2010a). Dynamics in the mixed microbial concourse. *Genes & Development, 24*(23), 2603–2614.

Wintermute, E. H., & Silver, P. A. (2010b). Emergent cooperation in microbial metabolism. *Molecular Systems Biology, 6*, 407.

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America, 74*(11), 5088–5090.

Wollman, E. L., Jacob, F., & Hayes, W. (1956). Conjugation and genetic recombination in Escherichia coli K-12. *Cold Spring Harbor Symposia on Quantitative Biology, 21*, 141–162.

Wright, O., Delmans, M., Stan, G. B., & Ellis, T. (2015). GeneGuard: A Modular Plasmid System Designed for Biosafety. *ACS Synthetic Biology, 4*(3), 307–316.

Wu, D., Daugherty, S. C., Van Aken, S. E., Pai, G. H., Watkins, K. L., Khouri, H., Eisen, J. A. (2006). Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biology, 4*(6), e188.

Yamada, S., Awano, N., Inubushi, K., Maeda, E., Nakamori, S., Nishino, K., . . . Takagi, H. (2006). Effect of drug transporter genes on cysteine export and overproduction in Escherichia coli. *Applied and Environmental Microbiology, 72*(7), 4735–4742.

Yedid, G., Ofria, C. A., & Lenski, R. E. (2008). Historical and contingent factors affect re-evolution of a complex feature lost during mass extinction in communities of digital organisms. *Journal of Evolutionary Biology, 21*(5), 1335–1357.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., . . . Glockner, F. O. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research, 42*(1), D643–648.

You, L., Cox, R. S., 3rd, Weiss, R., & Arnold, F. H. (2004). Programmed population control by cell-cell communication and regulated killing. *Nature, 428*(6985), 868–871.

Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., & Court, D. L. (2000). An efficient recombination system for chromosome engineering in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America, 97*(11), 5978–5983.

Zakataeva, N. P., Aleshin, V. V., Tokmakova, I. L., Troshin, P. V., & Livshits, V. A. (1999). The novel transmembrane Escherichia coli proteins involved in the amino acid efflux. *FEBS Letters, 452*(3), 228–232.

Zhang, Q., He, J., Tian, M., Mao, Z., Tang, L., Zhang, J., & Zhang, H. (2011). Enhancement of methane production from cassava residues by biological pretreatment using a constructed microbial consortium. *Bioresource Technology, 102*(19), 8899–8906.

Zhang, Q., Lambert, G., Liao, D., Kim, H., Robin, K., Tung, C. K., . . . Austin, R. H. (2011). Acceleration of emergence of bacterial antibiotic resistance in connected microenvironments. *Science, 333*(6050), 1764–1767.

Zhuang, K., Izallalen, M., Mouser, P., Richter, H., Risso, C., Mahadevan, R., & Lovley, D. R. (2011). Genome-scale dynamic modeling of the competition between Rhodoferax and Geobacter in anoxic subsurface environments. *ISME Journal, 5*(2), 305–316.

Zimmer, D. P., Soupene, E., Lee, H. L., Wendisch, V. F., Khodursky, A. B., Peter, B. J., . . . Kustu, S. (2000). Nitrogen regulatory protein C-controlled genes of Escherichia coli: scavenging as a defense against nitrogen limitation. *Proceedings of the National Academy of Sciences of the United States of America, 97*(26), 14674–14679.

Zomorrodi, A. R., & Maranas, C. D. (2012). OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Computational Biology, 8*(2), e1002363.

CURRICULUM VITAE