Title
Interactions of Microbes in Communities

Author
Sczesnak, Andrew

Publication Date
2018

Peer reviewed|Thesis/dissertation

Interactions of Microbes in Communities

By

Andrew Sczesnak

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy

with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Adam P. Arkin, Chair

Professor Matthew Traxler

Professor Michael Fischbach

Fall 2018

Abstract

Interactions of Microbes in Communities

By

Andrew Sczesnak

Joint Doctor of Philosophy

with University of California, San Francisco

in Bioengineering

University of California, Berkeley

Professor Adam P. Arkin, Chair

Groups of microorganisms sharing an environment (microbial communities) are ubiquitous in nature. Microbial communities provide essential ecosystem services to other life on Earth by e.g., participating in global biogeochemical processes or interacting with a host's immune system. Such microbes compete for scarce resources, modify an environment for their own purposes, actively war, and occasionally cooperate. Though numerous studies have surveyed the diversity of microbial life in different environments, few have determined the ways in which members of microbial communities interact with one another. Understanding the ways and means by which microbes interact is essential if we are to understand how microbial communities form, persist, and change over time. Knowledge of these processes will allow us to rationally design microbial communities to perform useful functions and predict how our actions might shift the balance of microbes in a community, and thus affect its function. In this work, we develop and apply novel methods for understanding microbial interactions.

Table of Contents

Chapter 1. Introduction

1.1. Diverse bacteria co-occur in communities throughout nature

It has been known for some time that diverse bacteria are found in a wide variety of environments. Before the discovery of DNA, biologists devised many ways of categorizing microbial life, such as cellular morphology, staining, and growth assays in different media. Using these classic methods, it was determined that soil, feces, and wounds, among other environments, harbor microbes that look and act differently from one another.

With the discovery of DNA and the advent of low-cost DNA sequencing, it became possible to quantify the relative abundance of all bacterial taxa in a sample by using the 16S ribosomal gene, and others, as taxonomic markers. Predicated on the theory that all Earthly life shares a common ancestor, the 16S rRNA gene is well-conserved across all bacteria (Cole et al. 2014). Despite high levels of conservation, several "hyper-variable" regions of this gene have accumulated mutations over time, such that the phylogenetic distance of two organisms can be inferred from the similarity of their 16S hyper-variable regions using one of many distance calculations (Schloss 2010).

Applying the technique of 16S sequencing, at first using Sanger's method, and later using short-read sequencers from Illumina and others, biologists have cataloged microbial life in samples as diverse as an acid mine drainage (Bond et al. 2000), a coral reef (Fernando et al. 2015), hyena guts (Heitlinger et al. 2017), smart phones (Lax et al. 2015), and a nuclear waste site at Oak Ridge National Lab (Christensen et al. 2018). These efforts have generated billions of sequences, the analysis of which suggests which bacteria co-occur and in which environments. Though these catalogs have been highly informative as to *what* bacteria are present in an environment, they provide little information as to *why* they are present together. What are the genetic and environmental factors that cause some bacteria to thrive in one community but not another? Among all environmental factors, to what extent do microbe-microbe interactions play a role in community assembly, stability, and diversity?

1.2. Microbial interactions can be computationally predicted

Before asking *why* bacteria co-occur, it is useful to know with some degree of statistical certainty *which* bacteria co-occur. Taking advantage of the enormous amount of available 16S sequencing data, methods have been developed which infer microbial interactions based on co-occurrence across many environments. Though computational methods can't tell us with absolute certainty that two bacteria interact, their results are useful starting points for follow-up experiments.

In recent work, Friedman et al. developed a method to infer correlations between taxa in microbial communities from marker gene sequencing data: SparCC (Friedman & Alm 2012). An important and often overlooked aspect of marker gene (e.g., 16S rRNA) sequencing is that the resulting read count data is compositional, rather than absolute: the relative abundance of all taxa in a sample must sum to 1.0. Intuitively, if the relative abundance of one taxon increases, there must be a concomitant decrease in the relative abundance of other taxa such that the total remains unchanged. If correlations between taxa are calculated without accounting for this fact, they will tend to have negative correlations, "regardless of the true correlation between the underlying absolute abundances" (Friedman & Alm 2012). SparCC estimates the pairwise correlation

between taxa, accounting for compositional effects, by assuming the number of taxa in a dataset is large and the number of true correlations is small.

Improving on this work, Kurtz et al. developed SPIEC-EASI, which additionally accounts for a variety of correlation network topologies (Kurtz et al. 2015). The authors used synthetic data to validate their approach, given the current lack of gold standard microbial interaction networks. Both methods have been successfully applied to predict interactions between microbes in diverse communities. However, they do not address the mechanisms underlying these inferred interactions or whether the interactions affect community function.

1.3. Mechanisms of microbial interactions

Interactions between organisms are thought to take many forms. Perhaps the most intuitive interaction paradigms are competition, where both organisms require a single resource to grow, and cooperation, where both organisms experience a growth benefit in each other's presence (Figure 1.1). Other modes of interaction include commensalism, where one organism provides a growth benefit without itself being affected, and predation, where one organism benefits by harming another (Grosskopf & Soyer 2014).

Computational methods, such as those in 1.2, are only able to label putative interactions as "positive" or "negative." Though we know there are multiple interaction types that fall under either category, such methods are unable to achieve that level of granularity. In addition, computationally inferred interactions lack a time component: they represent only a single snapshot of what is an inherently dynamic process. Though dissecting interactions over time is difficult with current methods, it was attempted in recent work by Venturelli et al. The authors selected a panel of organisms highly abundant in the human gut and subjected them to pairwise co-culture, reading out their relative abundances over time by 16S sequencing. Using this pairwise interaction data, a model was built that predicts the structure and dynamics of communities assembled from subsets of the panel of organisms under study. One interesting finding of this work is that pairwise interactions explain most observed community structure. For example, if organism X interacts with organism Y, and organism Y interacts with organism Z, the influence of X on Z via X → Y → Z is far less important than the direct influence of X on Y, or Y on Z (Venturelli et al. 2018). This has important implications for understanding microbial communities, where the complexity assumed to exist via these "indirect" interactions may be overstated.

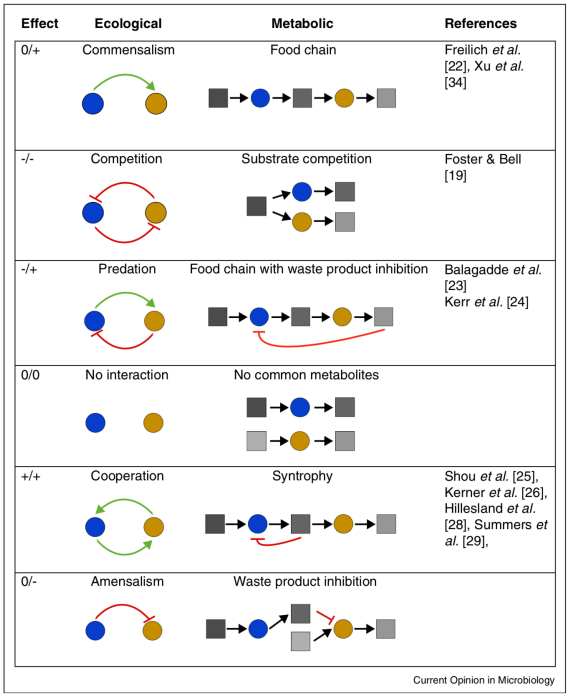

Figure 1.1. The six basic motifs of microbial interactions. Blue and yellow circles denote different microbial strains respectively, while boxes represent metabolites. Stimulating and inhibitory interactions mediated by taxon-specific traits or metabolites are indicated with red and green arrows respectively. In case of a syntrophy, the first microbe in the food chain (blue circle) is inhibited by the accumulation of its own waste product in the environment (mainly via thermodynamic limitations). This inhibition is relieved by the second microbe (yellow organism), which uses the waste product of the first microbe as a food source. Hence, both organisms benefit from the presence of the other. (Figure and caption adapted from Grosskopf et al., 2014)

Among microbes, competition for resources (e.g., iron or a carbon source) is likely the most common mode of interaction. It is assumed to occur between any arbitrarily chosen pair of taxa that occupy the same or similar ecological niche. Perhaps more interesting than instances of direct competition, however, are cases where one microbe has figured out how to get a leg up on its competitors.

Bacteriocins are short, peptide toxins, produced by one taxon to inhibit the growth of other, generally closely-related taxa. Given the likelihood that two closely-related taxa directly compete for resources, bacteriocins are a means by which one taxon might gain a fitness advantage over its relatives. Though the first bacteriocin (colicin) was identified in *E. coli*, similar toxins are broadly found in diverse microbes (Riley & Wertz 2002). A recent survey of the pan-genome of plant-associated *Pseudomonas* identified the presence of at least 13 bacteriocin and 7 antibiotic biosynthesis gene clusters (Loper et al. 2012). The secretion of such molecules can be harnessed to modulate the abundance of microbes in a community and modify its function. For example, some strains of *Pseudomonas* are known to protect crops against fungi by secreting toxic molecules (Haas & Défago 2005). By inoculating crops with biocontrol strains of *Pseudomonas*, the abundance of fungi decreases, and the microbial community becomes less phytotoxic. Methods to identify this type of interaction in a high-throughput way could facilitate the discovery of useful microbes or pathways, and are addressed in Chapter 4 of this work.

Perhaps it is our human bias to think that elements of our culture, like cooperation, exist in other more distantly-related lifeforms. Nevertheless, microbiologists have looked for examples of cooperation among microbes, and have found them. Whereas cooperation between humans is the domain of philosophers, in the study of cooperation between microbes, microbiologists have the benefit of statistical and genetic analysis. A priori, one might conclude that cooperation between unrelated taxa is unlikely to occur in nature: the most foundational premise of evolution is that a gene's highest purpose is to make copies of itself, and itself only. Why would one organism, a collection of genes, help an unrelated organism reproduce? How can organisms with a mutualistic relationship protect themselves from cheaters who consume a shared resource without contributing?

Recent work has demonstrated that not only does cooperation occur in nature, it can spontaneously occur in long-term evolution experiments (Poltak & Cooper 2010). Rather than being detrimental to participants, cooperation may lower the overall energy expenditure of a community (Morris et al. 2013). One early model based on the iterated prisoners' dilemma demonstrates how mutualism between two non-competing taxa can evolve and persist (Doebeli & Knowlton 1998). Another model suggests that cooperating communities limit their susceptibility to exploitation by physically excluding non-participants (Momeni et al. 2013). Both models rely on spatial structure to explain the evolution of mutualism. Supporting this result is the observation that microbial communities often form biofilms, which facilitate the exchange of molecules and serve to exclude cheaters.

Instead of observations of human interactions informing our study of microbial interactions, perhaps lessons learned from microbes can explain human behavior. Like microbial communities, human communities (or biofilms) are defined by the close proximity in which their members live, a lack of trust of non-participants who might exploit their mutualism, and the resulting ways in which they create barriers to exclude non-participants.

1.4. Summary

Methods to study the interactions of microbes in communities have mostly focused on surveys of microbial diversity and the development of ecological models. Some interactions have been computationally identified from sequencing data, but few follow-up studies have been conducted to understand their nature. In general, there has been a lack of methods to dissect the mechanisms underlying microbial interactions in a high-throughput way. In this work, we develop and apply novel methods for understanding microbial interactions.

Chapter 2. Developing a high-throughput 16S sequencing and analysis platform

2.1. Introduction

As discussed in Chapter 1, some methods of inferring microbial interactions rely on accurate assessment of the presence or absence, and if possible, the abundance of microbial taxa in a community. The most common way of obtaining such an assessment is by high-throughput sequencing of the 16S rRNA gene. In this assay, the 16S rRNA gene is used as a taxonomic marker. Under the assumption that all bacteria share a common ancestor, the 16S rRNA gene—which is an essential part of the ribosome and hence an essential part of the cell—has highly conserved regions that are required for function, and highly variable regions under less selective pressure that differ greatly between taxa. The highly variable regions can be used to infer from which taxon a sequence has originated. Using PCR and high-throughput DNA sequencing, it is possible to count the number of occurrences of a highly variable region. Dividing this count by the sum of all such counts yields the relative abundance of a particular highly variable region—and therefore a particular taxon—in a given sample.

In the first half of this chapter, we develop a method of 16S sequencing to suit the needs of our laboratory. Published methods are very good at this task, and were the basis for this work. It was necessary, however, for us to optimize previously-published protocols to reduce cost and increase throughput. In the second half of this chapter, we discuss the results of two experiments performed with this platform.

2.2. Primers and PCR conditions for 16S amplification and sequencing

To facilitate the rapid quantification of bacterial taxa in many samples simultaneously, we designed and optimized a PCR reaction which yields an Illumina-tailed amplicon containing the V3/V4 region of the 16S ribosomal gene. This amplicon can be sequenced directly on Illumina MiSeq. In our method, each well of a 96-well plate contains a different DNA barcode, and the available barcode space allows eight plates to be sequenced in a single run.

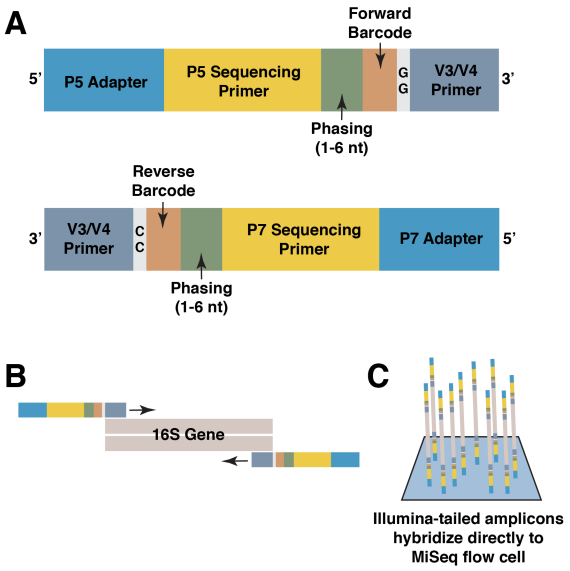

Figure 2.1. Schematic of 16S sequencing primer design. (a) A forward and reverse primer targeting the V3/V4 region of the 16S gene each include a barcode, random phasing sequence, and P5 or P7 Illumina adapters. Adapters on both ends of the amplicon allow for paired-end sequencing. (b) Illumina-tailed primers hybridize to 16S rDNA in a sample, generating complete barcoded 16S amplicons. (c) These amplicons are sequenced directly on Illumina MiSeq, where they hybridize directly to the flow cell without further manipulation.

We optimized this PCR reaction along three axes: number of PCR cycles, primer concentration, and amount of input genomic DNA. It is advantageous to minimize the number of PCR cycles required, as substantial bias can be introduced with increased cycle count (Polz and Cavanaugh 1998). PAGE purified primers such as those used in this assay are expensive, and so minimizing the amount of primer used per reaction can drastically lower the cost per sample. Additionally, it can be difficult to obtain large amounts of genomic DNA from certain samples, and so a protocol which requires a small amount of DNA allows greater flexibility in the types of experiments which can be performed. For example, microbes contained in groundwater are routinely studied in our lab, which are present in concentrations as low as 100,000 cells per liter. The amount of DNA that can be extracted from these samples is often low. Similarly, high-throughput plate-based culture assays in volumes of 100 uL or less can yield little DNA when extracted. As such, we sought to minimize the amount of DNA required by our 16S sequencing protocol while maximizing accuracy. Ultimately, we settled on a protocol that required 10 ng of input DNA, 2.5 pmol primer per reaction, and 20 cycles of PCR amplification.

This 16S sequencing approach is substantially the same as other published approaches (Kozich et al. 2013; Fadrosh et al. 2014), though we added phasing in both the forward and reverse primer. Phasing improves the number of reads which pass Illumina's quality filter by solving the problem of unbalanced base composition at each cycle (Wu et al. 2015). This platform gave members of our lab the ability to perform hundreds of 16S experiments in a single day, without the need to send samples to a core facility and wait weeks for results.

2.3. Bioinformatic analysis of 16S sequencing data

Rather than completely re-implement a 16S analysis pipeline, we chose to perform most pre-processing in-house, then pass the resulting reads to well-established tools such as UPARSE (Edgar 2013) and QIIME (Caporaso et al. 2010). Briefly, overlapping paired-end reads from a 300 or 600 cycle Illumina MiSeq run were first merged using PEAR (Zhang et al. 2014). Next, reads were de-multiplexed according to a parameter file mapping barcodes to plate well IDs. After quality filtering and trimming, the resulting reads were passed to either QIIME or UPARSE for OTU calling and abundance quantification. Our in-house tool was written entirely in Python and is available at http://github.com/polyatail/arkin. Using this pipeline and analysis tools, several 16S experiments were performed, the results of which are briefly discussed in the following subchapters.

2.4. Screening for depletion of sulfate-reducing microbes

This section has been adapted, with permission, from Carlson et al., Environmental Science and Technology, 2015.

In numerous industrial applications, production of hydrogen sulfide (H2S) by sulfate-reducing microorganisms (SRM) is economically and environmentally costly. SRM, and their sulfate substrates, are present in oil and gas pipelines, where corrosive H2S is a primary cause of leaks. Specific inhibition of sulfidogenesis could reduce costs, and prevent loss of fragile ecosystems by contamination with crude oil and derivatives. Despite the industrial need for potent and selective inhibition of SRM, few molecules

are available to perform this task. No studies have systematically evaluated the potency and selectivity of known SRM inhibitors, such as inorganic oxyanions.

We designed a 16S-based assay to test a panel of compounds for their ability to inhibit the growth of environmental SRM, and hence the production of H2S. Briefly, media containing concentration gradients of inhibitory molecules was inoculated with a microbial community obtained from the environment. Following incubation, genomic DNA was extracted, and the relative abundance of each microbe was determined via 16S as described in section 2.1 and 2.2. If a molecule inhibited SRM, the relative abundance of SRM taxa decreased. Using a concentration gradient allowed us to calculate the IC50 of a compound by fitting the relative abundance data to a sigmoidal dose-response curve.

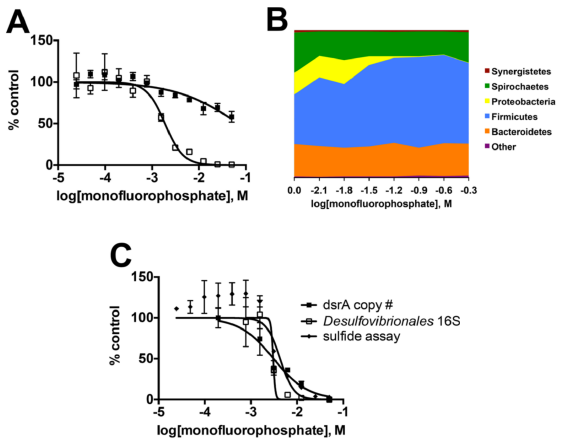

Figure 2.2. Dose−response curves of MFP against growth, sulfidogenesis, 16S amplicon phylum relative abundances, and dsrA copy # in a marine enrichment culture grown in the presence of varying concentrations of MFP for 48 h. (a) Growth (filled symbols) and sulfide (open symbols). (b) Phylum level relative abundances from 16S amplicon sequencing. Desulfovibrionales was the sole Proteobacterium observed. (c) Sulfide, DsrA copy number, Desulfovibrionales relative abundances.

To test the ability of compounds to inhibit SRM, we derived a marine microbial community from sediment obtained from San Francisco By. Sediment was used to inoculate an anoxic continuous flow reactor containing Instant Ocean (Thermo Fisher) marine mix at 35 g/L containing 2 g/L yeast extract. The resulting enriched community was harvested and stored in glycerol stocks at -80 C until use. In one experiment, monofluorophosphate (MFP) was serially diluted in microplates and inoculated with our enriched community at OD=0.02. Following incubation for 48h, the OD and community composition of each well was determined.

Overall growth of the community was inhibited by MFP at high concentrations (Figure 2.2A). However, the reduction in growth was specific to Proteobacteria (Figure 2.2B), of which Desulfovibrionales was the only taxon observed. A follow-up QPCR assay confirmed that copy number of the Desulfovibrionales 16S and dsrA genes correlated well with decreased sulfide production, as MFP concentration increased. This demonstrates that MFP specifically inhibits sulfide production by SRM. Compared to other oxyanions tested with our assay, MFP was among those with the lowest IC50 for sulfide production while maintaining high specificity (Table 2.1). This work demonstrates the potent and selective inhibition of SRM by MFP (Carlson et al. 2015).

Oxyanion attributes			IC_{50} (95% CI) against marine enrichment cultures (mM)		
Molecular formula	Compound name	Group	Growth OD 600	Sulfide OD 660	Selectivity index (Growth IC_{50}: Sulfide IC_{50})
$(CH_3)SO_2$	Dimethyl sulfone	16 (chalcogen)	>100	>100	Not inhibitory
$CH_3SO_3^-$	Methanesulfonate	16 (chalcogen)	>100	>100	Not inhibitory
$NH_2SO_3^-$	Ammonium sulfamate	16 (chalcogen)	74 (52-104)	81 (66-99)	0.9
SeO_4^{2-}	Selenate	16 (chalcogen)	>10	0.38 (0.21-0.72)	>26.3
SeO_3^{2-}	Selenite	16 (chalcogen)	>10	0.57 (0.56-0.59)	>17.5
TeO_4^{2-}	Tellurate	16 (chalcogen)	>1	0.012 (0.010-0.013)	>83
TeO_3^{2-}	Tellurite	16 (chalcogen)	>1	0.053 (0.040-0.071)	>18.9
ClO_4^-	Perchlorate	17 (halogen)	21 (14-31)	2.3 (2.0-2.6)	9.1
ClO_3^-	Chlorate	17 (halogen)	44 (29-60)	1.6 (1.4-1.8)	27.5
ClO_2^-	Chlorite	17 (halogen)	2.8 (1-7.5)	1.17 (0.8-1.7)	2.4
BrO_3^-	Bromate	17 (halogen)	0.59 (0.40-0.86)	0.39 (0.12-1.2)	1.5
IO_4^-	Periodate	17 (halogen)	0.40 (0.27-0.60)	0.40 (0.27-0.59)	1
IO_3^-	Iodate	17 (halogen)	0.37 (0.19-0.72)	0.36 (0.13-0.31)	1
NO_3^-	Nitrate	15 (pnictogen)	46 (34-62)	8.0 (7.0-9.0)	5.75
NO_2^-	Nitrite	15 (pnictogen)	5.5 (3.3-9.3)	0.12 (0.1-0.4)	45.8
HPO_3^{2-}	Phosphite	15 (pnictogen)	>100	>100	Not inhibitory
SPO_3^{2-}	Thiophosphate	15 (pnictogen)	>100	>100	Not inhibitory
FPO_3^{2-}	Monofluorophosphate	15 (pnictogen)	>100	1.9 (0.29-4.4)	53
AsO_4^{2-}	Arsenate	15 (pnictogen)	0.41 (0.21-0.76	0.36 (0.055-2.4)	1.1
AsO_3^{2-}	Arsenite	15 (pnictogen)	0.011 (0.008-0.014)	~0.008	1.4
VO_4^{3-}	Vanadate	5 (Transition metal)	2.7 (0.89-8.2)	0.17 (0.12-0.26)	16
CrO_4^{3-}	Chromate	6 (Transition metal)	0.027 (0.020-0.036)	0.031 (0.026-0.037)	0.9
MoO_4^{2-}	Molybdate	6 (Transition metal)	0.69 (0.24-1.9)	0.0069 (0.0032-0.015)	100
WO_4^{2-}	Tungstate	6 (Transition metal)	1.3 (0.57-3.0)	0.0024 (00075-0.0079)	542

Table 2.1. Inhibition by oxyanions of growth and sulfidogenesis in marine enrichment cultures.

2.5. Screening for microbial interactions by dilute enrichment culture

This section has been adapted, with permission, from Justice et al., Applied and Environmental Microbiology, 2017.

As discussed in chapter 1, it is possible to infer interactions between microbial taxa by analyzing their relative abundances across different environments. Methods to do so are generally applied to marker gene sequencing data post-hoc, rather than to sequencing data that has been explicitly obtained for such purposes as part of an experiment designed to detect interactions. In this work, we developed a method to infer interactions by serially diluting an environmental sample and using it to inoculate large numbers of replicate enrichment cultures. Assuming cells in the dilute inoculum are thoroughly mixed, randomly distributed into each replicate culture, do not interact, and all grow at similar rates, we expect taxa to co-occur at a calculable frequency across all replicates (the null model). Differences between observed co-occurrence frequencies and the null model suggest that one or more of our assumptions are not true, e.g., the taxa interact.

We obtained an initial environmental community from groundwater, which we estimated to have 37,000 cells/mL using the acridine orange direct count (AODC) method. This was used to inoculate 960 replicate enrichment cultures, of which 480 were grown anaerobically in nitrate-reducing conditions, and 480 aerobically. Of these

two conditions, both were further subdivided into 96 replicates each of the inoculum diluted to 1e-1, 1e-2, 1e-3, 1e-4, and 1e-5. Following incubation, genomic DNA was extracted and 16S sequencing performed as described in sections 2.1 and 2.2.

As expected, based on 16S rRNA gene amplicon sequencing data, enrichment cultures started with the highest inoculum concentrations had the highest operational taxonomic unit (OTU) richness. The communities receiving the most concentrated inoculum had statistically similar numbers of OTUs under nitrate-reducing and aerobic conditions (t test, P=0.10), with the nitrate-reducing communities averaging 26.5 OTUs (n=94; standard deviation [SD], 11.27 OTUs) and the aerobic communities averaging 29.2 (n =96; SD, 10.53 OTUs). OTU richness declined in experiments that received less concentrated inocula (Figure 2.3).

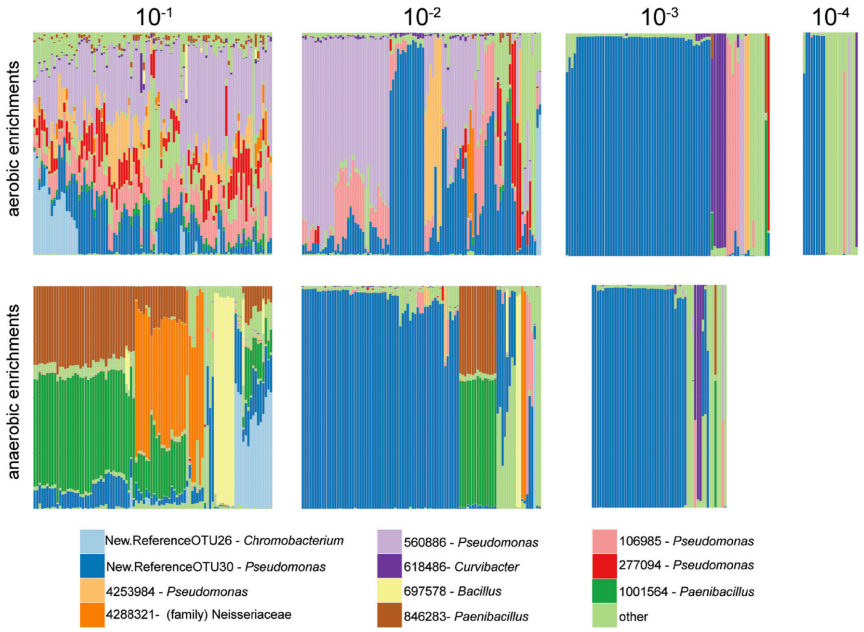

Figure 2.3. Relative abundance of OTUs from (y axes) across all communities (x axes) in the first four dilutions of aerobic enrichments and first three dilutions of anaerobic nitrate-reducing enrichments. Only the most abundant 11 OTU are shown for clarity.

In the 1e-2 dilutions, the aerobic communities tended to have higher species richness than the nitrate-reducing communities (t test, P=2.09e-6), with nitrate-reducing cultures having on average 9.3 OTUs (n=96; SD, 5.7 OTUs) and the aerobically cultivated communities with 13.5 OTUs (n =96; SD, 6.4 OTUs). Aerobic communities that received the most diluted inoculum had on average only 2.3 OTUs (n=3; SD, 2.31 OTUs), and only a single OTU in a single sample was detected in the nitrate-reducing communities begun with the most dilute inoculum. In addition to species richness, we quantified how evenly communities were structured with Pielou's index. At all dilutions, the anaerobic communities showed significantly reduced evenness, despite being seeded from the same populations that seeded the aerobic communities. These results indicate that the anaerobic cultivation conditions favor the outgrowth of a smaller number of taxa, results consistent with stronger selective forces under the anaerobic conditions.

Dilution	No. of samples	No. of species	Total no. of species pair combinations	No. of analyzed combinations[a]	No. of positive interactions[b]	No. of negative interactions[b]	Median power[c]	FDR (%)[d]
NO_3-10^{-1}	94	230	26,335	802	58	47	6.2	1.53
NO_3-10^{-2}	96	124	7,626	317	12	9	7.5	3.02
NO_3-10^{-3}	54	91	4,095	37	1	0	8.2	7.40
NO_3-10^{-4}	0	0	NA[e]	NA	NA	NA	NA	NA
NO_3-10^{-5}	1	1	NA	NA	NA	NA	NA	NA
O_2-10^{-1}	96	164	13,366	1,303	8	8	9.75	16.29
O_2-10^{-2}	96	109	5,886	564	15	3	9.85	6.27
O_2-10^{-3}	79	65	2,080	74	2	0	7.7	7.40
O_2-10^{-4}	22	37	666	12	0	0	NA	NA
O_2-10^{-5}	3	6	NA	NA	NA	NA	NA	NA

[a]Analyzed combinations represent only those species pairs expected to have 1 or more cooccurrences.
[b]Significance at the threshold of a *P* value of <0.001.
[c]Calculated as |observed interactions − expected interactions|.
[d]FDR, false-discovery rate.
[e]NA, not applicable.

Table 2.2. Summary of pairwise co-occurrence analyses for each condition and dilution.

Given the probabilistic nature of how we seeded each replicate, we sought to identify pairs of taxa that may be interacting by observing if they were found more or less frequently together than one would expect by chance. For each condition and dilution, the total number of pairwise comparisons, the number of significant positive and negative associations, and the median strength of the associations for each condition and dilution are shown in Table 2.2. Overall, we identified 115 putative interactions (56 negative and 59 positive) among 34 OTU in the nitrate-reducing samples and 34 putative interactions (23 positive and 11 negative) among 15 OTU in the aerobic samples. There was very little overlap between interaction predictions across conditions, with only 14 OTU and 5 predicted interactions shared in the aerobic and anaerobic communities. Of those five shared interactions, all were positive associations among pairs of closely related OTUs.

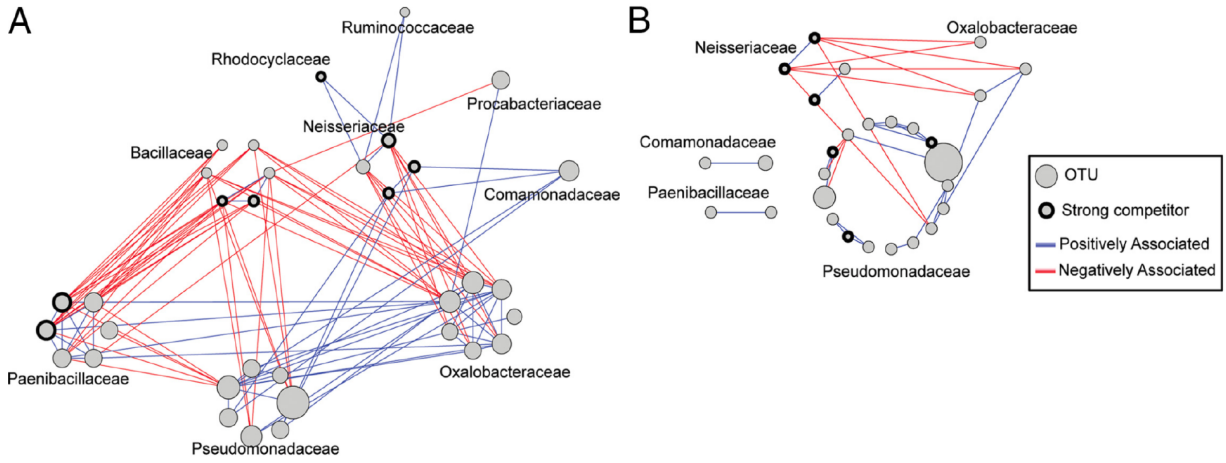

Figure 2.4. Networks depicting positive and negative associations between pairs of taxa in anaerobic nitrate-reducing communities (a) and aerobic communities (b). Graphs were made by the union of interaction graphs at each dilution for aerobic and anaerobic samples, respectively. Positive associations are shown in blue and negative associations in red. OTUs predicted to be strong competitors are indicated with a bold outline. The size of the node for each OTU scales with the estimated number of cultivable units of that OTU in the initial inoculum.

In the anaerobic samples, OTUs of the Pseudomonadaceae were positively associated with members of the Oxalobacteraceae and negatively associated with members of the Bacillaceae and Paenibacillaceae. Oxalobacteraceae, on the other hand, were positively associated with the Paenibacillaceae and negatively associated with members of the Neisseriaceae and Bacillaceae. The Bacillaceae had no positive connections to other families and were negatively associated with members of the Pseudomonadaceae, Oxalobacteraceae, and Paenibacillaceae. In aerobic samples, some positive associations between the Pseudomonadaceae and Oxalobacteraceae were identified, and the Neisseriaceae share negative associations with members of both the Oxalobacteraceae and Pseudomonadaceae families (Figure 2.4).

This work provides a proof-of-concept for detection of interactions between taxa in an environmental sample based on their co-occurrence when grown in the laboratory. Compared to interactions inferred computationally, as described in section 1.2, these interactions can immediately be verified and used as a basis for further experiments: the organisms in question are not uncultivable, and the community from which they derived has been archived. Further, the data from this experiment required only one Illumina MiSeq run and can be obtained in less than a week. Future work surveying microbial taxa in environmental samples could benefit from the addition of this method to existing 16S pipelines (Justice et al. 2017).

Chapter 3. A device for high-throughput enrichment culture in situ

3.1. A culture problem, or microbial dark matter

Many surveys of bacterial life in diverse environments have been performed using 16S sequencing, such as those discussed in Chapter 2. One astonishing conclusion of these studies is that we currently lack the ability to cultivate most microbial life in the laboratory. We have obtained countless 16S and other sequences from the environment that match no known cultivatable organism, and which suggest the creation of a new taxonomic unit (e.g., a candidate phyla). It has been suggested that half of such candidate phyla have no cultivated representatives (Rinke et al. 2013). Recent work has attempted to fill gaps in our knowledge of so-called "microbial dark matter" by algorithmically separating metagenome sequencing reads into bins (Hug et al. 2016), obtaining near-complete genomes of single cells obtained from the environment (Rinke et al. 2013), and cultivating samples in situ to approximate a cell's preferred growth conditions (Nichols et al. 2010). Such studies have, and continue to, broaden our perspective on microbial evolution and diversity.

The reason most microbial life remains uncultivated may be, in part, for historical reasons. For example, *E. coli* is perhaps the most well-studied bacterial taxon, originally isolated from the human gut in 1884 by Theodor Escherich and serving as a model bacterium for the next century (Blount 2015). The ease with which *E. coli* can be isolated from the mammalian gut, and stool, suggests that it is highly abundant in those environments. Several studies, however, including most recently those performed as part of the Human Microbiome Project (Human Microbiome Project Consortium 2012), reveal that *E. coli* is present in extremely low abundance in the human gut. Far more prevalent and far more abundant are Bacteroides taxa, such as *B. thetaiotamicron.* Why, then, aren't we all studying *B. theta*? It turns out that *E. coli* is a facultative anaerobe able to utilize diverse carbon sources, while *B. theta* is an obligate (though aero-tolerant) anaerobe highly adapted to utilize complex dietary polysaccharides (Aaron G Wexler 2017). *E. coli* readily grew on broth available to microbiologists in the late nineteenth century, when exposed to air. *B. theta* was not isolated until 1912—several decades later—with improved knowledge of bacterial metabolism and cultivation techniques.

When cultivating a new organism in the lab, the complexity of the search space of media conditions must be considered. Varying only 10 simple parameters (e.g., pH, carbon source, amino acids, etc), each with 10 possible choices (e.g., pH range 4-8, skip 0.4) results in 10 billion possible combinations—far more than can be methodically tested in a lifetime. Instead, cultivation of new organisms is first attempted in one of many standard media recipes available from e.g., American Type Culture Collection (ATCC) or German Collection of Microorganisms and Cell Cultures (DSMZ). Parameters of these standard recipes can be varied slightly to optimize growth, but the fraction of the search space which can be sampled remains quite limited. Where did these media recipes come from, anyway? They were developed and modified to cultivate microbes that were isolated without too much difficulty early in the history of microbiology—microbes like *E. coli*. This explains why most cultivated organisms belong to only four phyla: Proteobacteria, Firmicutes, Actinobacteria, and Bacteroidetes (Rinke et al. 2013). Growth conditions for a handful of organisms in these phyla were determined early on—in some cases motivated by the desire to cure a specific disease—and used to isolate and cultivate their nearest relatives. Had early microbiologists stumbled upon a different set of microbes first, perhaps we'd be studying them instead.

3.2. Cultivation conditions informed by genomics

One technique which holds promise for the future is the determination of optimal culture conditions for a given microbe from its genome sequence. A genome might encode the ability to utilize e.g., a carbon source, which might be the optimal carbon source to include in a growth medium. Similarly, a genome might *not* encode the ability to synthesize e.g., one or more amino acids, which must then be supplemented in a growth medium.

In one recent study, the authors used the genome and transcriptome of a *Rikenella*-like bacterium to rationally design a growth medium. Found in the gut of the medicinal leech *Hirudo verbana*, the genome of their target bacterium encodes enzymes to utilize glycans derived from host mucin. By performing in vivo RNA-seq, the authors determined that these glycan-utilization enzymes were upregulated by their target bacterium when growing in the leech gut. This suggested that host glycans might be an ideal carbon source to include in a growth medium. By modifying a standard recipe for Eggerth-Gagnon medium to include porcine gastric mucin, the authors were able to cultivate their *Rikenella*-like bacterium in vitro (Bomar et al. 2011).

While this result is intriguing, it is not the norm: most microbes remain uncultivated, and many lack a sequenced genome which can be used to infer cultivation conditions or as a reference for RNA-seq analysis. Perhaps most fascinating is the prospect that there are factors other than those directly knowable by genomic analysis which contribute to the cultivability of an organism. Growth history—the previous conditions in which a bacterium grew—may determine their ability (or lack thereof) to thrive in a new condition (Wolf et al. 2008). Similarly, interactions between a bacterium and its host or other members of a microbial community may be essential (Schnupf et al. 2015). Both possibilities are difficult to investigate using current methods of genomic analysis, and point to exciting new avenues of research.

3.3 Cultivation of microbes in situ using the iChip

Recently, methods of growing microbes in situ have gained much traction. Rather than attempting to isolate microbes by finding ideal growth conditions in the laboratory, these methods place uncultured organisms back into their native environments where presumably, the conditions are already ideal. Intuitively, this makes sense: why spend substantial effort creating something artificial that is, at best, an approximation for something abundantly available in nature? It is true that synthetic media can allow researchers to carefully control experimental conditions. But the vast majority of microbial cultures in the lab are already conducted using complex media containing e.g., yeast extract or blood. Do we truly lose that much control over growth conditions with in situ culture? Given the huge search space of media conditions that goes untested, we are even making use of all that control?

One method is the isolation chip, or "iChip." The iChip is a device which provides microbes their own micro-growth chambers, separated from each other and the environment by porous membranes through which small molecules pass. Each chamber is inoculated with, on average, zero or one cells via limiting dilution (Figure 3.1). The entire chamber is then e.g., buried in soil, where it is incubated long enough for micro-colonies to grow. Following incubation, the chamber is retrieved, and its wells analyzed for growth.

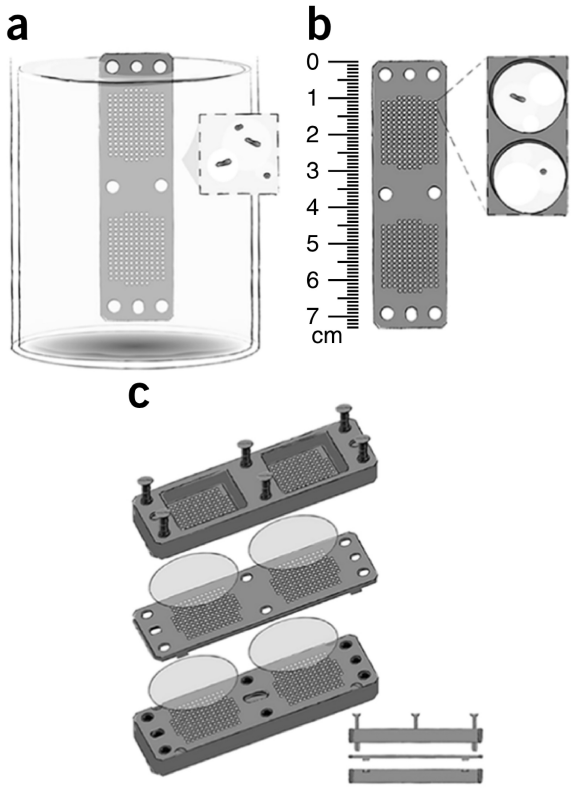

Figure 3.1. Conceptual design of the ichip. (**a**) A plate containing multiple through-holes is dipped into a suspension containing microbial cells. (**b**) The dilution of the suspension is such that each hole captures (on average) a single cell. (**c**) The basic assembly of the ichip: membranes allowing diffusion cover through-holes on each side, and upper and bottom plates containing matching holes press the membrane against the center loaded plate. Once screwed in place, sufficient pressure is applied to seal the contents of each through-hole, which becomes a miniature diffusion chamber. (Figure and caption reproduced verbatim from Berdy, et al, Nat Protoc 2017)

One reason a microbe may resist cultivation is because it lacks an essential (but unknown) growth factor. For example, a culture-resistant *Psychrobacter* strain can be cultured in vitro in the presence of 3.5 nM of a 5-amino acid peptide, LQPEV, which is secreted by a "helper" strain found in the same environment. After growth in laboratory conditions, it was found that some cells evolved to no longer require LQPEV (Nichols et al. 2008). In a later study, the same authors found that many uncultivable microbes could be grown in the lab after forming micro-colonies in an iChip (Berdy et al. 2017; Nichols et al. 2010), perhaps via mechanisms similar to *Psychrobacter*. Future work will determine the prevalence of such signaling molecules in microbial communities, and how they affect community structure and function.

In a recent study, the authors screened extracts derived from several iChips worth of soil bacteria (~10,000) for any which could inhibit the growth of *Staphylococcus aureus*, a common pathobiont. The extract of one microbe, *Eleftheria terrae*, was found to have good activity against *S. aureus* in their assay. Follow-up experiments were performed to find and determine the structure of the inhibitory compound, teixobactin. Genomic analysis of *E. terrae* allowed the authors to find, with high confidence, the non-ribosomal peptide synthesis gene cluster responsible for teixobactin synthesis. Remarkably, they found teixobactin to be active against a broad range of gram positive pathogens and were unable to isolate strains of *S. aureus* or *M. tuberculosis* with appreciable resistance (Ling et al. 2015). Of note is the fact that extracts of 10,000 isolates had to be screened to find a single hit. We saw room for improvement by further miniaturizing the isolation chambers using droplet microfluidics.

3.2. Cultivation of microbes in situ using a novel droplet-based approach

Rather than cultivate microbes in 1mm cylindrical chambers, like the iChip, we proposed to increase throughput by encapsulating single cells in 100um diameter agarose beads generated with a microfluidic device. By shrinking the effective size of the chamber each cell is grown in 1,500-fold, it is possible to incubate approximately 384,000 beads in a device the same size as an iChip (Figures 3.2, 3.3).

The workflow for using our device is largely the same as using the iChip. First, an environmental sample is diluted such that each chamber/droplet receives on average 0 or 1 cells. Agarose beads are generated using a microfluidic device, and then incubated between two diffusible membranes in situ (Figure 3.4). In addition to dramatically increased throughput, our method allows for more rapid screening of micro-colonies in beads via flow cytometry, or with an acoustic printer.

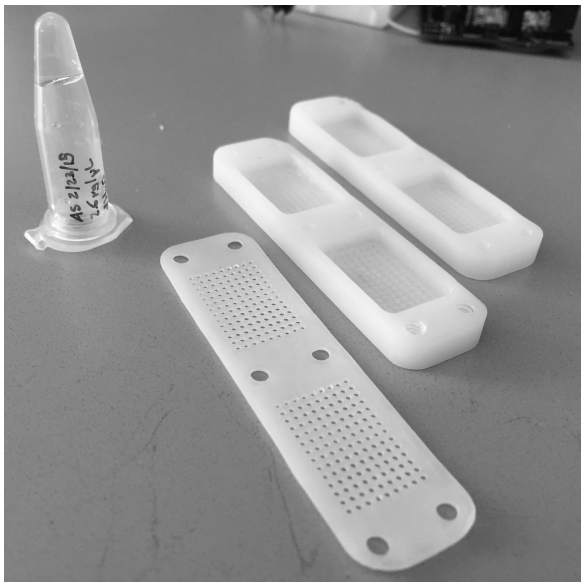

Figure 3.2. A 3D-printed (SLA) model of an iChip (Berdy et al, Nature 2017), produced in the Arkin lab.

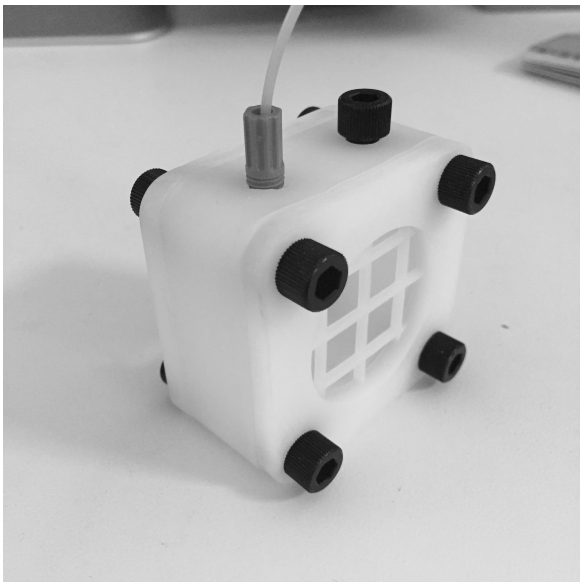

Figure 3.3. A 3D-printed (SLA) model of an in situ growth chamber designed to hold micro-scale beads, produced in the Arkin lab.

Method 1. To screen micro-colonies in beads for their ability to inhibit the growth of another microbe, beads are re-injected into a second microfluidic device. This second device generates water-in-oil-in-water (w/o/w) double emulsion droplets which are sortable by FACS (Sukovich et al. 2017). These droplets each contain a previously-incubated bead (which may or may not contain a micro-colony), suspended in media inoculated with e.g., *S. aureus*. After incubation in conditions permissive for *S. aureus* growth, the droplets are sorted via FACS. Droplets in which no *S. aureus* has grown may have contained an antibiotic, presumably produced by a micro-colony embedded in an agarose bead. Growth of *S. aureus* can be read via GFP, if a suitable GFP-expressing strain is available, or forward- and side-scatter. Droplets negative for *S. aureus* are sorted into plates, with the goal of isolating and identifying the microbe present in the agarose bead.

Method 2. The above method performs the screen in double-emulsion droplets, only to sort them into plates at the end of the process. In this method, beads are instead

acoustically printed onto standard agar plates, at a density such that use of space on the plate is maximized, but the likelihood of cross-contamination between beads is minimized. The agar plates can then be overlaid with e.g., *S. aureus*. Zones of inhibition surrounding beads following incubation suggest the presence of a bead harboring a micro-colony which produced an antibiotic. Micro-colonies at the center of zones of inhibition can be picked by hand or machine for further study.

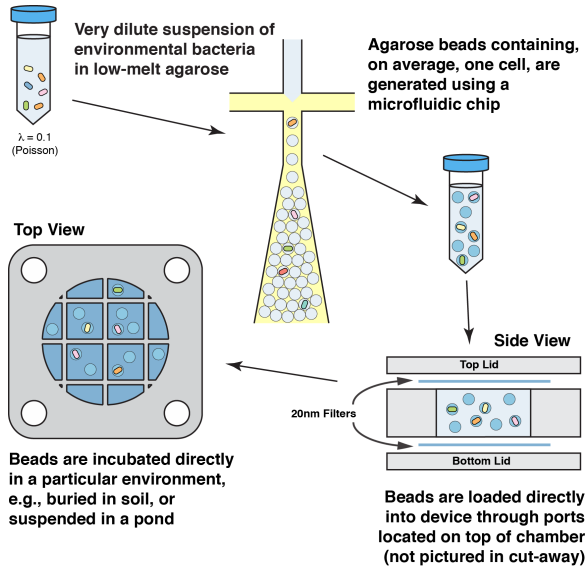

Figure 3.4. Design and operation of a device for in situ cultivation of microbes embedded in millions of agarose beads. An environmental sample is first diluted such that each bead will receive, on average, 0 or 1 cells. Beads are collected in a chamber which is sandwiched between two 20nm filters to allow for passive diffusion. The device is then incubated in e.g., soil or water, prior to analysis of beads via FACS or other assays.

The method proposed here drastically increases the throughput of the iChip assay. Using *Method 1*, millions of micro-colonies can be screened in a single day for inhibition of a target microbe, like *S. aureus*. This represents a 1,000-fold increase over a previous study, which resulted in the successful isolation of the antibiotic teixobactin (Ling et al. 2015). We believe our method could be developed to form a core part antibiotic discovery pipelines in the future.

Chapter 4. Finding genes mediating microbe-microbe interactions

4.1. Introduction

Though many methods have been devised to identify and quantify the abundance of microbes in environmental samples, there have been few methods to computationally predict their interactions, and even fewer methods to confirm interactions in vitro. This has largely been due to the complexity assumed to exist in communities of potentially hundreds of co-existing microbes, and an inability to cultivate many environmental microbes in the laboratory.

In one recent study, the authors tested the idea of co-cultivating a cross-feeding pair of mutant strains of *E. coli* within water-in-oil microfluidic droplets. Mutants unable to synthesize tryptophan (W-) expressed RFP, while those unable to synthesize tyrosine (Y-) expressed GFP. Droplets were generated by encapsulating a very dilute suspension of both mutants, such that the assortment of cells into droplets was approximately a Poisson process. By setting the Poisson lambda parameter to a low value (e.g., 0.1), the authors ensured that most droplets were either empty or contained only a single cell. Following incubation, droplets were examined under a microscope. Those which were GFP+/RFP-, or GFP-/RFP+, contained only one mutant and appeared to divide slowly or not at all over the four day incubation period. Those which were GFP+/RFP+, however, contained both mutants and were observed to have robust growth of each mutant, which was presumably due to cross-feeding of the essential amino acids each auxotrophic mutant was unable to produce on its own (Park et al. 2011). This study proves that it is possible to detect cross-feeding interactions using droplet microfluidics.

In this chapter, we describe a method for detecting cross-feeding and inhibitory interactions between mutants in a barcoded library using droplet microfluidics and Illumina sequencing. Compared to the aforementioned previous study, this method has the advantage of interrogating mutants spanning an entire genome in one experiment, rather than cherry-picking single mutants to test for cross-feeding interactions. To test our system, we used iron acquisition in *E. coli*, a well-studied pathway in which cross-feeding is known to take place (Raymond et al. 2003).

4.2. Pooled fitness assays may obscure interactions between mutants

In bacterial genetics, a forward genetic screen is performed by generating many mutants, screening for a desired phenotype, then identifying mutations presumed to be causative. Prior to the advent of low-cost, high-throughput DNA sequencing, such a screen was rather laborious. Not only would the phenotype screen require isolation of individual mutants as colonies, the location of each mutation in the genome could only be inferred via conjugation mapping, in situ hybridization, or restriction digestion and gel electrophoresis.

Today, this process requires far less work. Our lab has devised a method for simultaneously determining the phenotype of mutants spanning entire bacterial genomes in many conditions. Transposons bearing unique barcodes are randomly integrated into a genome via the Tn5 or Mariner transposon, generating a library of tens of thousands of mutants. The entire library is then inoculated into liquid or solid media and left to undergo several rounds of division. After growth, genomic DNA is extracted from the resulting cells. PCR with Illumina-tailed primers is used to amplify barcoded transposon DNA, which is then sequenced. The abundance of each barcode (i.e., each

mutant) before and after growth can be compared to determine how its representation has changed upon growing in the condition under test. If a mutant is less abundant, it has a growth defect in that condition. If a mutant is more abundant, it thrives in that condition (Wetmore et al. 2015).

One important caveat of this workflow is that interactions between mutants are not considered. It may be difficult to observe growth defects of some mutants when they are complemented by interactions, such as metabolite exchange or cell-cell contact, with other mutants. One such well-studied interaction is iron acquisition in *E. coli*.

4.3. Iron acquisition in *E. coli* is a prototypical positive interaction

Siderophores are small, high-affinity, iron-chelating molecules secreted into the environment. There, they bind their substrate and are re-imported into the cell. Among known siderophores, enterobactin binds iron with the highest affinity. In iron-limited environments, *E. coli* synthesizes enterobactin molecules via a pathway encoded by the genes entABCDEF, imports ferric enterobactin via an ABC transporter encoded by fepABCDG, and releases iron from ferric enterobactin via destructive hydrolysis catalyzed by ferric enterobactin esterase (fes) (Raymond et al. 2003).

An enterobactin molecule secreted into the environment by an *E. coli* cell may be taken up by any other cell. This includes other *E. coli* cells that do not secrete enterobactin and cells belonging to other taxa, such as *Pseudomonas*. In this way, siderophores like enterobactin form a "common good" resource for other organisms to use. In a mixed culture of *ent*, *fep*, and *fes* mutants, not all cells are able to synthesize or utilize enterobactin. These mutants form a prototypical community in which some rely on others to survive (Figure 4.1). Detection of phenotypes for these mutants presents a problem in pooled fitness assays, as discussed in section 4.2. Strains harboring mutations in *ent* genes cannot synthesize enterobactin, yet have no apparent fitness defect in a pooled assay, as they are able to utilize enterobactin molecules secreted by other mutants in the pool.

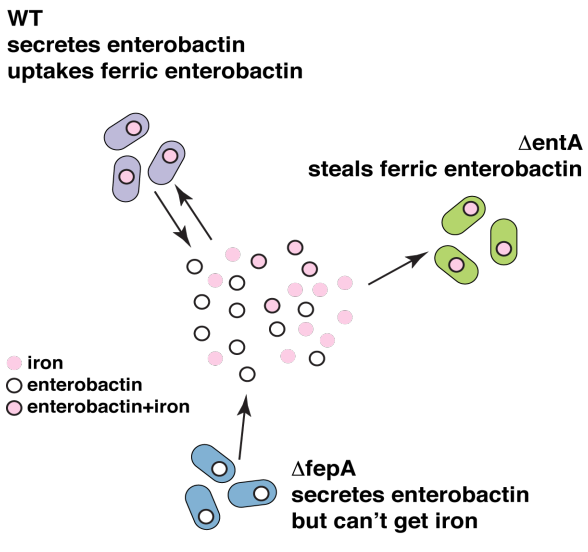

Figure 4.1. *Enterobactin cross-feeding in* E. coli.

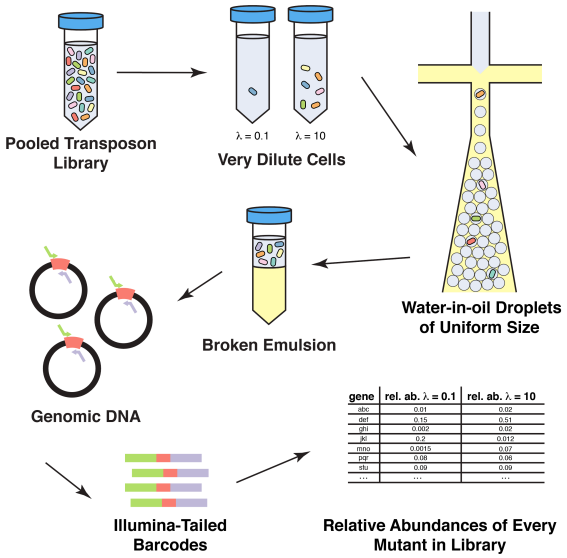

Figure 4.2. Schematic of droplet BarSeq.

4.4. A method to identify cross-feeding obscured by pooled fitness assays

One way to solve the problem of microbe-microbe interactions being obscured in pooled fitness assays is to revert to earlier, more laborious strategies that required each mutant to be isolated and tested alone. Obviously, this is not ideal. Instead, we devised a method to test each mutant on its own without sacrificing throughput. By encapsulating a mutant library in microfluidic droplets such that each receives only a single cell, the effect of growing each mutant in its own test tube (or as a discrete colony on a plate) is achieved (Figure 4.2). As a comparator, droplets can be generated such that each receives an average of 10 cells. In 10 cell droplets, cross-feeding can occur, whereas in single-cell droplets, it cannot. Assuming Poisson statistics, droplets receiving an average of 10 cells will occasionally receive only a single cell. However, the probability of this occurrence is only 1:10,000—sufficiently low that we reason it will be not interfere with the assay.

Using this method with iron-deficient medium, we expect single cell droplets inoculated with *fep*, *ent*, or *fes* mutants to grow poorly, unable to either synthesize or utilize enterobactin. In droplets inoculated with an average of 10 cells (i.e., a random selection of many mutants), *fep* and *fes* mutants will grow poorly, unable to import or release iron from enterobactin. In contrast, *ent* mutants will thrive, acquiring iron via enterobactin secreted by other cells in the droplet. By comparing the results of fitness experiments in single- and multi-cell droplets, we can detect cross-feeding as an increase in mutant fitness when grown in a pool versus a droplet alone.

4.5. Investigating the parameters of iron acquisition in enterobactin mutants

To confirm our expectations of enterobactin-related mutants based on the literature, we performed a series of experiments to quantify various aspects of iron acquisition in *E. coli*. All experiments were performed in M9 minimal media with 22.2 mM D-glucose, 1x Wolfe's Vitamins, and 1x Wolfe's Minerals (without iron), unless otherwise indicated. Prior to inoculation, cells were washed four times in the final media. Media and cultures were stored only in new plastic containers to avoid the unintentional addition of iron from material adhered to glass. Cultures were inoculated at OD=0.0005, the approximate OD at which 80 um spheres (microfluidic droplets) will contain an average of 0.1 cells each (i.e., lambda=0.1, assuming Poisson statistics). Cultures were incubated at 37C for 24h with orbital shaking in a Tecan Spark plate reader.

First, we sought to determine the approximate amount of residual iron left in our cultures after washing. This was accomplished by obtaining growth curves for a panel of iron-related mutant and wild-type strains in a gradient of ascending iron (III) sulfate concentration. Iron (III) was chosen because this is the oxidation state found in aerobic aqueous solution. In these conditions, carrying capacity is presumably limited by iron. By obtaining a titration curve for iron, we can pinpoint the amount of iron at which the carry capacity, as measured by OD, increases. This is related to the amount of iron already present in the culture.

An estimate of the amount of iron present in our washed cultures is between 0.04883 uM and 0.19531 uM (Figures 4.3 and 4.4). This could be 1) carryover from the original rich media, 2) contaminating iron in any component of the media (i.e., vitamins, minerals, glucose, M9 salts), or 3) intracellular stored iron within ferritin complexes (Andrews et al. 2003; Abdul-Tehrani et al. 1999). While this is a non-negligible amount

of iron, it is unlikely to interfere with our experiments, as *ent*, *fep*, and *fes* mutants are unaffected by the addition of iron in amounts as high as 10 uM (Figure 4.5). Interestingly, it appears that *fes* and *fepB* recover, to some extent, in very high (50 uM) concentrations of iron. This may be due to leaks in the ferric enterobactin import and release pathways, which we discuss later.

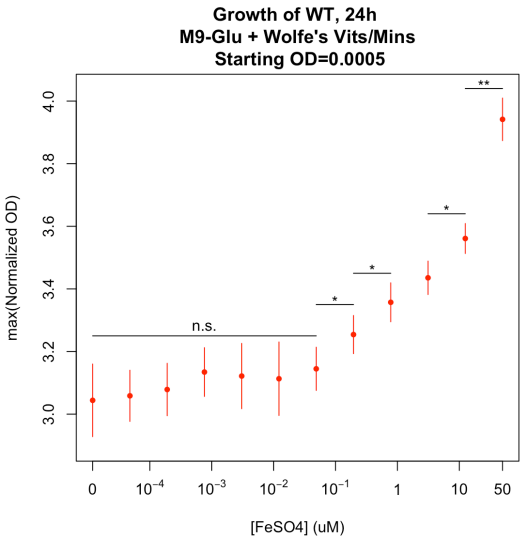

Figure 4.3. Wild-type *E. coli MG1655* was grown in minimal media containing various concentrations of iron. After 24h, max(OD) begins to increase between 0.04883 uM and 0.19531 uM added iron, the first statistically significant difference (* $p < 0.05$, ** $p < 0.01$, Welch's t-test). This provides a range of iron present in culture after washing and inoculation.

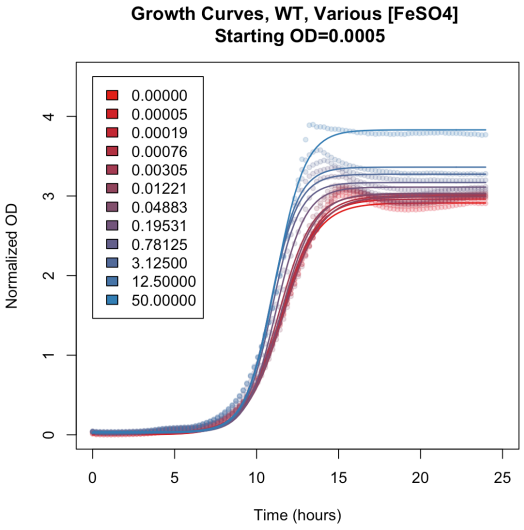

Figure 4.4. Growth curves of wild-type *E. coli MG1655* in minimal media with various concentrations of added iron (uM) indicated by color in legend box. Fitted growth curves show similar lag phases and growth rates, but different carrying capacities.

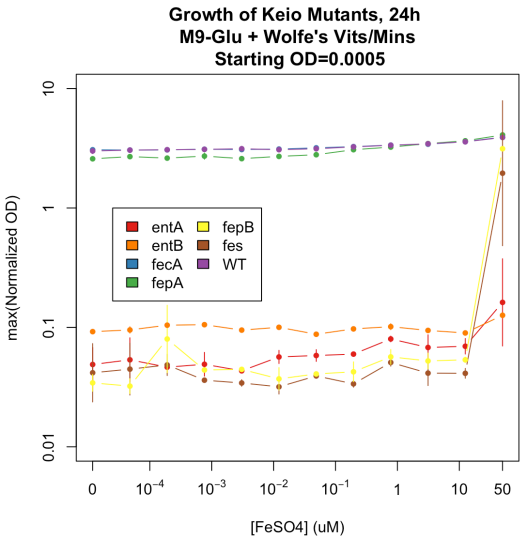

Figure 4.5. Carrying capacity of media containing various concentrations of added iron, for different enterobactin-related mutant strains of *E. coli MG1655*. Mutants deficient in enterobactin synthesis, uptake, and release are unaffected by the addition of iron as high as 10 uM.

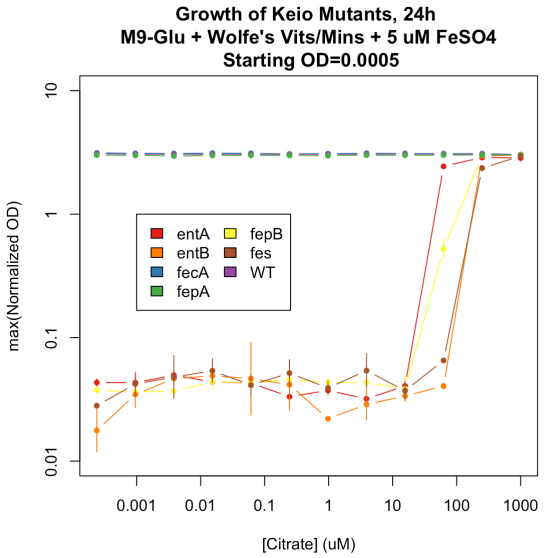

Figure 4.6. Carrying capacity of media containing 5 uM ferric sulfate and various concentrations of citric acid, for different enterobactin-related mutant strains of *E. coli MG1655*. Mutants deficient in the enterobactin pathway recover to wild-type growth levels at approximately 50 uM citrate. At this concentration, iron is effectively chelated to form ferric citrate, which is imported via an ABC transporter.

It is known that many *E. coli* strains have multiple redundant mechanisms for acquiring iron. Of the known mechanisms, three do not rely on exogeneously produced siderophores: enterobactin, ferric citrate, and direct import of ferrous iron. Our experiments are conducted aerobically, where ferrous iron (II) is rapidly oxidized to ferric iron (III), and as such, ferrous iron import is not likely to contribute much to iron availability. However, strains harboring mutations in the enterobactin pathway can still acquire iron via ferric citrate. As such, we sought to determine the extent to which uptake via the ferric citrate pathway was likely to occur in our experiments. Strains were inoculated in media containing 5 uM iron (III) sulfate in a gradient of ascending citric acid concentration. Large amounts of citrate were required to chelate iron such that it could be imported by enterobactin pathway mutants (Figure 4.6). As such, we concluded that uptake via the ferric citrate pathway was unlikely to occur in our experiments, where no citrate or iron is explicitly added to the medium. Mutants lacking *fecA* (Figure 4.6, blue line), the outer membrane component of the ferric citrate

transporter, have no phenotype in this assay. This is because our *fecA* single-gene mutant has an intact enterobactin pathway.

In *E. coli*, transcription of genes related to iron acquisition is regulated by Fur (ferric uptake regulation). Iron-bound Fur represses transcription of gltA (citrate synthase, part of the TCA cycle) and induces transcription of acnA (aconitase A, which converts citrate to isocitrate) during times of iron abundance. In times of scarcity, Fur is unbound by iron, resulting in increased transcription of gltA and the accumulation of citrate. Though this suggests *E. coli* may produce citrate to chelate iron, this phenomenon has not yet been experimentally observed (McHugh et al. 2003). Still, given the large amount of citrate required to chelate iron, we believe it is unlikely for iron-starved cells inoculated at low OD to produce enough within the 24h duration of our experiments.

Stored iron in ferritin complexes accounts for much of the iron in an *E. coli* cell. As such, we sought to determine how much growth could be expected from this source of iron alone. We obtained growth curves of our panel of strains in media with no iron added, at various starting ODs spanning several orders of magnitude. If stored iron is contributing to growth, there should be a fixed number of generations that each enterobactin-deficient strain can complete before intracellular iron stores are exhausted. This will manifest as a direct correlation of starting OD and final OD, where the only parameters determining the final OD of the culture are 1) the starting OD and 2) the number of generations completed on stored iron reserves.

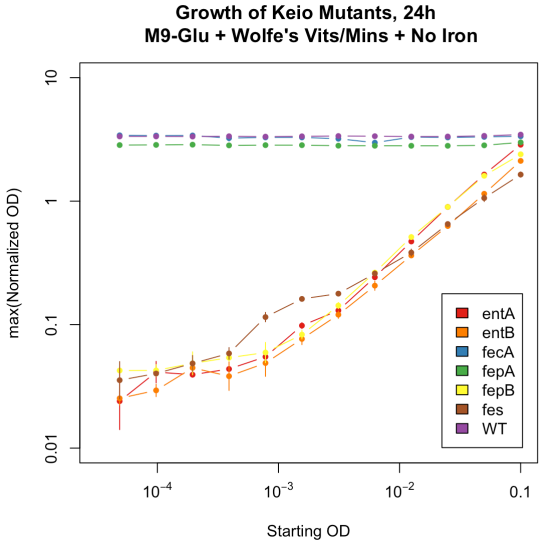

Figure 4.7. Carrying capacity of media with no added iron, for different enterobactin-related mutants. WT requires fewer doublings at higher starting ODs and more doublings at lower starting ODs, to reach the same maximum density. In comparison, *ent* and *fep* mutants are unable to acquire environmental iron and undergo a fixed number of doublings on intracellular iron reserves, reaching densities that depend on starting OD.

Stored iron accounts for 5-6 doublings of enterobactin-related mutants in media with no added iron or citrate (Figure 4.7), as calculated by the log2 ratio of final OD to starting OD. After pre-growth for 24h, at which point stored iron should be depleted, *entA* and *entB* can no longer divide (Figure 4.8), lacking the ability to acquire iron from the environment. This number of doublings is highly plausible, given previously published data quantifying the amount of iron stored by a stationary phase culture grown in iron-rich medium (~0.025% of dry weight) versus the amount of iron stored by a culture grown in iron deficient medium (~0.005% of dry weight) (Abdul-Tehrani et al. 1999). Dividing these two quantities (0.025 and 0.005) suggests that stored iron allows ~5 doublings before reaching an iron-starved state.

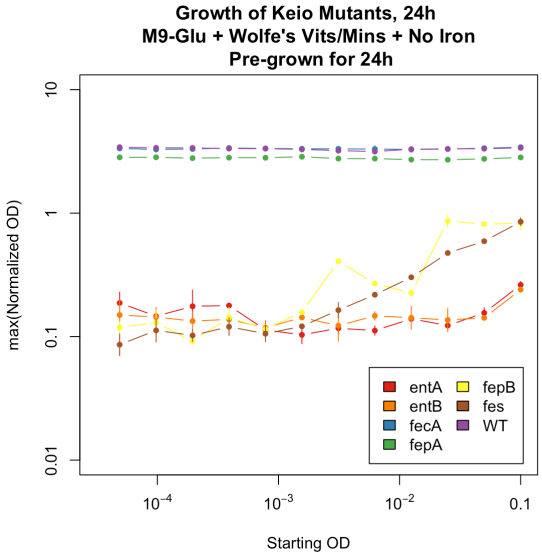

Figure 4.8. After pre-growth for 24h, stored iron is depleted. This results in a de-correlation of starting OD and final OD as compared to Figure 4.7. Strains lacking *ent* and *fep* genes do not divide, unable to acquire environmental iron and lacking stored iron.

Interestingly, *fes* and *fepB* are less affected by the depletion of stored iron than *entA* and *entB*. This may be due to the off-target esterase activity of another enzyme (other than Fes) that can release iron from enterobactin after import. For *fepB*, it seems that import of ferric enterobactin does not strictly depend on all the components of the Fep ABC transporter. In Figures 4.5 to 4.8, *fepA* has little to no fitness defect compared to other mutants. FepA is the outer membrane substrate-binding component of the transporter and is perhaps dispensable when there is no competition for ferric enterobactin. Similarly, FepB is the periplasmic substrate-binding component and its function may be somewhat redundant with FepA—both bind ferric enterobactin with high affinity. Our *fepB* mutant has a substantially greater fitness defect in our experiments, however, suggesting FepB's more essential role in ferric enterobactin uptake. For either the *fes* or *fepB* mutant, if the ability to acquire iron is even marginally intact during pre-growth to deplete stored iron, newly imported iron will offset some of the depletion. This explains the partial correlation between starting and ending OD we observe in Figure 4.8.

Finally, we wanted to be sure that our *ent* mutants would respond to the addition of enterobactin, as expected. Strains were grown in media containing an ascending concentration gradient of purified, non-ferric enterobactin. As expected, *ent* mutants exhibited a dose-dependent increase in carry capacity (as measured by OD), while *fepB* did not (Figure 4.9).

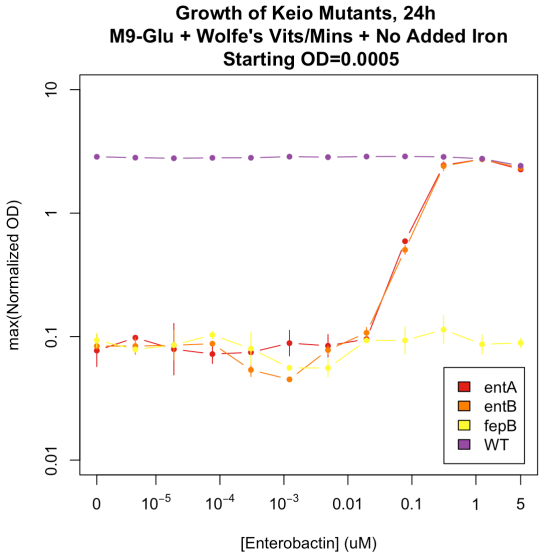

Figure 4.9. Carrying capacity of media containing 5 uM ferric sulfate and various concentrations of enterobactin, for different enterobactin-related mutant strains of *E. coli MG1655*. Mutants deficient in the enterobactin pathway recover to wild-type growth levels at approximately 0.5 uM enterobactin.

In conclusion, we have shown 1) that enterobactin is necessary and sufficient for aerobic growth in iron-limited media in the absence of citrate, 2) that citrate is unlikely to contribute to iron acquisition unless explicitly added to the medium, and 3) residual iron in our cultures is likely intracellular stored iron, which can be depleted by pre-growth for 24h in media without added iron.

4.6. Screening an *E. coli* mutant library for interactions between mutants

To test the method in 4.4, we used a barcoded transposon mutant library generated from *E. coli BW25113*, a strain nearly identical to *E. coli MG1655*. A frozen aliquot of the library was recovered in LB for 4h, then washed four times in M9-Glucose with no added iron. The washed culture was diluted such that an 80 um diameter sphere of the resulting dilution would, on average, contain 0.1 cells. The washed and diluted culture was then used to generate 80 um diameter microfluidic droplets in a T-junction PDMS device, using 2% RAN-008 surfactant (RAN Biotechnologies) in HFE-7500 (3M). HFE-7500 is preferred to FC-40 (3M) as it is far less susceptible to electrostatic coalescence. The device was run for approximately 60 minutes, until 10 million droplets were collected. Assuming 10% of droplets receive one or more cells and that randomly inserted transposon mutants evenly cover all 3,789 non-essential genes represented in the library, 10 million droplets will represent each *E. coli* gene an average of 264 times in our assay. The resulting droplets were incubated for 24h in round-bottom 7 mL polypropylene tubes, at 37C, with shaking at 250 rpm. Tubes must be round-bottomed, and securely attached to the shaker, to reduce mechanical coalescence. Following incubation, 1 uL of droplets were suspended in 50 uL HFE-7500 on a slide and examined for growth under phase contrast.

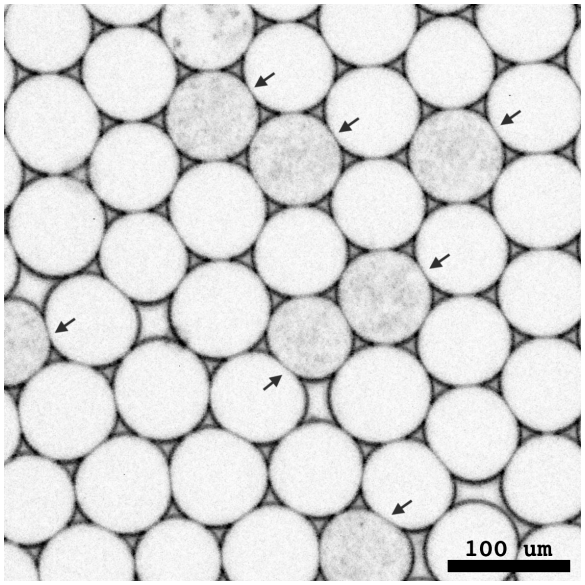

Figure 4.10. Microfluidic droplets containing an *E. coli BW25113* library, after 24h of growth. 100x magnification, Ph3. Full-color original photo was desaturated, inverted, and auto-contrast was applied in Adobe Photoshop. Arrows indicate droplets with detectable growth. White droplets are presumed empty.

Four random frames like Figure 4.10, though larger and containing around 300 droplets each, were used to calculate the percentage of droplets with growth. In total, 84 of 1265 counted droplets grew, or 6.64% (target was 10%). We assumed that cells were distributed into droplets as a Poisson process, and that most droplets encapsulating at least one cell had detectable growth. Given those assumptions, in this experiment 93.6% of droplets received zero cells, 6.2% received one cell, and 0.2% received more than one cell. This represents a 31-fold enrichment of single-cell droplets to multi-cell droplets. The experiment was then repeated, pre-growing the mutant library for 24h to deplete stored iron reserves. In this experiment, 124 of 1319 counted droplets grew, or 9.4% (Figure 4.11). Therefore, 91% of droplets received zero cells, 8.6% received one cell, and 0.4% received more than one cell, for a 21.5-fold enrichment of single-cell to multi-cell droplets.

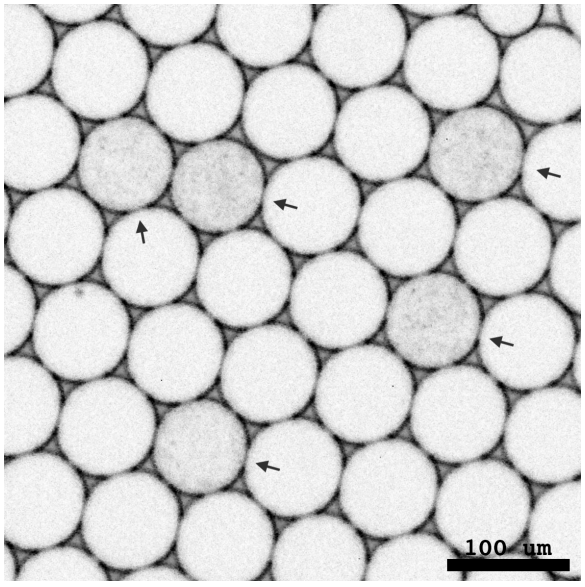

Figure 4.11. Microfluidic droplets containing an *E. coli BW25113* library, after 24h of growth in droplets following 24h pre-growth in media with no added iron. 100x magnification, Ph3. Full-color original photo was desaturated, inverted, and auto-contrast was applied in Adobe Photoshop. Arrows indicate droplets with detectable growth. White droplets are presumed empty.

After growth, the droplet emulsion was chemically broken. First, 2 mL perfluoro-1-octanol (Sigma, 370533-25G) was added to each sample and left to settle for several minutes until an aqueous (upper) and organic (lower) phase were visibly separate. Next, the aqueous phase was carefully removed and transferred to a 15 mL tube. To further separate the phases, each sample was spun briefly and slowly at 100 x g for 1 m such that the cells remained in suspension. The aqueous phase was again carefully removed, placed in a 15 mL tube, and spun at 3.2k x g for 10m to pellet the cells. The cells were then washed in 1mL PBS to remove any residual organic phase and transferred to a 1.5 mL tube. Finally, the cells were pelleted in a micro-centrifuge at 10k x g for 1m, the supernatant was removed, and the samples were kept at -20C until ready to process. Genomic DNA was obtained from each sample using the DNeasy Blood & Tissue Kit (Qiagen, 69506). BarSeq PCR, sequencing, and data analysis were performed as previously described (Wetmore et al. 2015; Price et al. 2018).

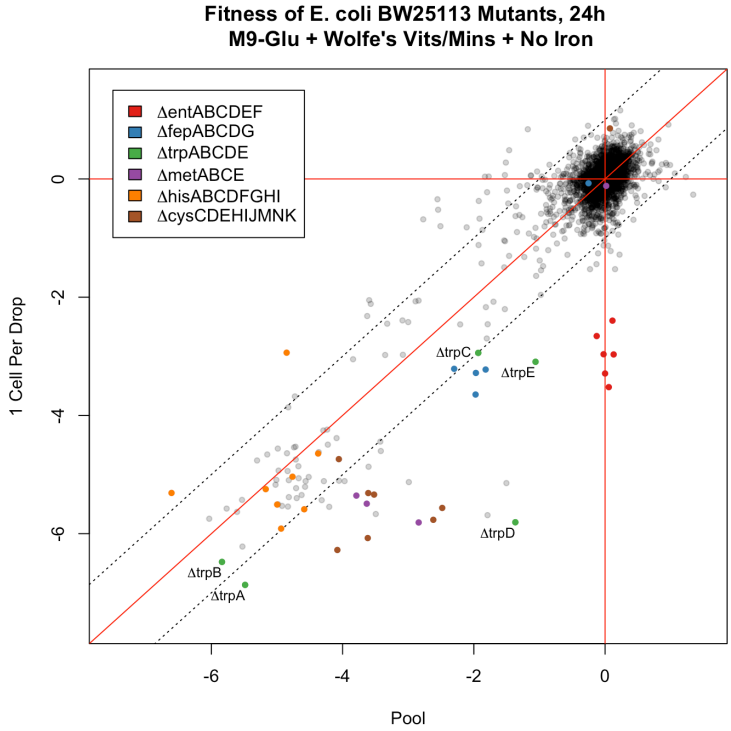

Figure 4.12. Comparison mutant fitness when grown alone and in co-culture confirms cross-feeding of enterobactin by *ent* mutants, but not *fep* mutants. Fitness scores of each gene when grown alone (vertical axis) and in co-culture (horizontal axis) are reported as points. The dashed lines (y = x +/- 1) highlight positive (below the line) and negative (above the line) interactions, respectively. In iron deficient conditions, enterobactin synthesis mutants (entABCDEF, red points) grow poorly alone, but have no fitness defect in co-culture when able to obtain enterobactin secreted by other mutants. In contrast, enterobactin uptake mutants (fepBCDG, blue points) grow poorly both alone and in co-culture, able to secrete enterobactin but not import it to acquire iron. Cross-feeding of some amino acid auxotrophs (green, purple, orange, and brown points) was observed in this assay.

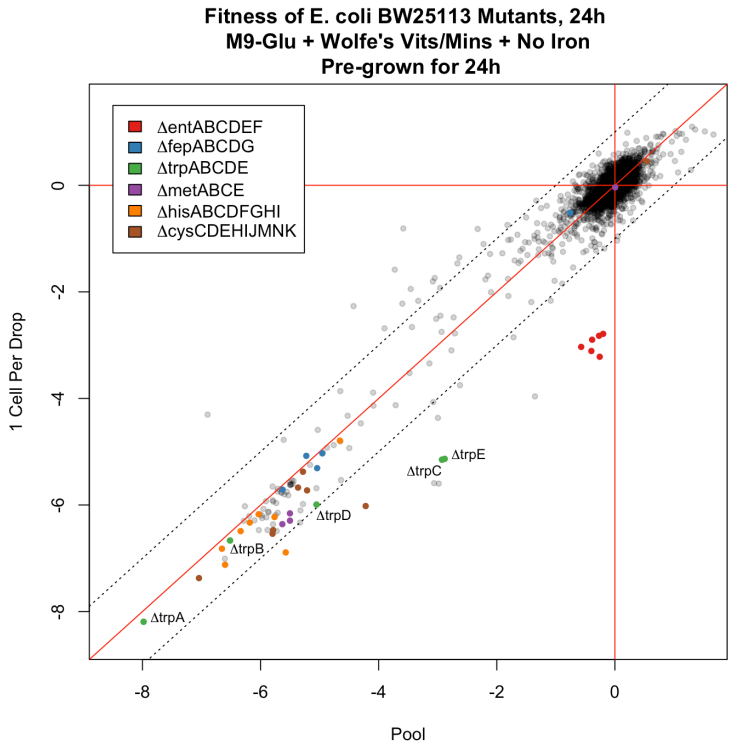

Figure 4.13. Droplet BarSeq with 24h pre-growth to deplete stored iron reserves. Though cross-feeding of enterobactin by *ent* mutants is robustly recovered, potential amino acid auxotroph cross-feeding seen in Figure 4.12, is not.

As expected, *ent* mutants had severe fitness defects (< -2) when grown alone in single-cell microfluidic droplets. When grown in a pooled culture, however, they had fitness values near zero, indicating no growth deficit (Figure 4.12). This confirms that previously reported cross-feeding of enterobactin by *E. coli* can be observed in our high-throughput droplet BarSeq assay. In contrast, *fep* mutants had severe fitness defects in both conditions, as they are unable to import ferric enterobactin. Confirming our own results, *fepA*, which seems to suffer little from iron limitation, had no fitness defect. Other known cross-feeding interactions, such as previously-reported cross-feeding by amino acid auxotrophs, were not robustly observed in this assay. This is likely due to slower growth of auxotrophs in minimal media, even when cross-fed. Prototrophs handily out-compete auxotrophs in the race to divide and consume glucose. Though auxotrophs can grow in the pooled culture—and not in single-cell droplets—they are not able to grow fast enough to have a substantially different relative abundance in our droplet BarSeq assay. Supporting the lack of robust recovery, most amino acid auxotroph cross-feeding interactions seen in Figure 4.12 are not substantial enough to be seen after pre-growth for 24h (Figure 4.13). Further, fitness values less than -4 are mostly indistinguishable from each other due to low read counts. As such, genes which have, for example, a fitness value of -6 in single-cell droplets and a higher fitness value of -4 in pooled culture, are not examples of robust cross-feeding.

4.7. Discussion and future directions

The experiments in this chapter demonstrate that it is possible to screen for positive interactions between microbes in a library using droplet microfluidics. Mutants unable to synthesize enterobactin grew poorly alone in droplets but recovered to wild-type growth levels when enterobactin-producing cells were present. Though amino acid auxotrophs appeared in area of the plot where positive cross-feeding interactions ought to be found (Figure 4.12), their growth rates were not observed to recover to wild-type levels, and did not persist when grown for an additional 24h (Figure 4.13). We believe that amino acid auxotrophs, even if cross-fed, do not double fast enough to maintain a presence in the pooled culture. Were the pooled culture passaged several times, the auxotrophs would likely be lost completely.

One caveat of this assay, which has largely gone unaddressed in microfluidics experiments, is that organic compounds may diffuse out of one droplet, through the continuous phase, and into another droplet. This occurrence breaks the assumption that each droplet is its own isolated micro-reactor. One recent study attempted to quantify this occurrence, finding that fluorescein diffused between droplets on the order of days, while rhodamine diffusion could be detected in as few as 15 minutes (Gruner et al. 2016). Importantly, the authors found that diffusion exponentially decreases with increasing distance between droplets. This suggests that an effective strategy to mitigate droplet cross-talk to is to increase the fraction of continuous phase in an emulsion (i.e., more oil, fewer droplets).

Diffusion across the continuous phase is largely dependent on the surfactant used. Unfortunately, there have been no studies on the widely-used RAN-008 surfactant, the structure of which is proprietary and unpublished. Given the increased stability of droplets generated with this surfactant, we expect it to perform at least as well as common fluorinated PEG surfactants. In our experiments, droplets were thoroughly mixed during incubation in an excess volume of HFE-7500, and around 10% of droplets encapsulated viable cells. The volume of each droplet was around 268 pL, and the total volume of emulsion and excess HFE-7500 was 7 mL. The approximately 1 million droplets encapsulating a viable cell, it follows, made up 3.8% of the emulsion volume. If all such droplets produced a molecule which diffused into the continuous phase, it would effectively be diluted 26-fold. Further, the excess volume of HFE-7500 used places droplets an average of 1.6 droplet-widths apart during incubation. At this distance, we expect diffusion between droplets to proceed 100-fold more slowly than if droplets were packed (Gruner et al. 2016). If rhodamine cross-talk is the upper bound for rate of diffusion, the droplet spacing in our assay would produce detectable effects after 25 hours, at which point our experiments would have been concluded. While we cannot exclude the possibility that cross-talk occurs in our assay, a molecule would have to accumulate quickly in a droplet, diffuse more rapidly through the continuous phase than rhodamine, and be active in target droplets after a 26-fold dilution. Notably, we did not observe cross-talk of enterobactin the continuous phase, which would have resulted in no fitness defect of *ent* mutants in droplets.

Cross-feeding of intermediates of tryptophan biosynthesis is known to occur in *E. coli*. In our assay, we observed a dramatic increase in fitness of *trpD* when grown in a pool, as compared to growth in single cell droplets. Similarly, *trpE* and *trpC* recovered slightly in pooled culture. However, the two final steps in tryptophan biosynthesis, catalyzed by *trpA* and *trpB*, grew poorly in both conditions. Since there are fewer potential cross-feeding intermediates for steps later in the pathway, this is to be

expected. These results suggest that other cross-feeding interactions are detectable using this assay. In this demonstration, conditions were chosen to detect cross-feeding of enterobactin. In future work, conditions might be chosen to detect cross-feeding of amino acids, or resistance to stressors (e.g., antibiotics).

Chapter 5. Concluding remarks

In this work, we discussed current methods for inferring interactions between microbes, developed a method for 16S sequencing and applied it in two separate experiments, outlined a method for high-throughput cultivation of microbes in situ, and developed and tested a method for detecting interactions between *E. coli* mutant strains in droplets. We demonstrated how throughput can be increased by decreasing the size of bacterial cultures using microfluidic droplets, and how this increased throughput facilitates the detection of microbial interactions between members of a large library.

While much effort up to now has been focused on determining what microbes are present in environmental samples and computationally inferring how those microbes interact, there is a lack of methods to determine interaction mechanisms. In this work, we outlined two methods using droplet microfluidics that can (1) increase the throughput of isolating uncultured microbes from the environment, and (2) reveal the genetics underlying microbial interactions. We discussed how isolation of uncultured microbes has led to the isolation and development of novel antibiotics, and will lead to the development of more useful therapeutics in the future. We discussed how, by uncovering genes responsible for microbial interactions, those genes might be harnessed to manipulate microbial communities and change community function.

The application of controlled culture of microbes inside microfluidic droplets will reveal deeper insights into the structure, function, and assembly of microbial communities going forward. In the future, the assay we developed in chapter 4 will be used to identify microbial interactions in bulk, in a variety of conditions. By carefully choosing which culture conditions and microbes to study, genes and pathways with immediate applications in health, agriculture, and environmental remediation will be revealed.

References

Aaron G Wexler, A.L.G., 2017. An insider's perspective: Bacteroides as a window into the microbiome. *Nature Microbiology*, 2(5), p.17026.

Abdul-Tehrani, H. et al., 1999. Ferritin mutants of Escherichia coli are iron deficient and growth impaired, and fur mutants are iron deficient. *Journal of bacteriology*, 181(5), pp.1415–1428.

Andrews, S.C., Robinson, A.K. & Rodríguez-Quiñones, F., 2003. Bacterial iron homeostasis. *FEMS Microbiology Reviews*, 27(2-3), pp.215–237.

Berdy, B. et al., 2017. In situ cultivation of previously uncultivable microorganisms using the ichip. *Nature protocols*, 12(10), pp.2232–2242.

Blount, Z.D., 2015. The Natural History of Model Organisms: The unexhausted potential of E. coli. *eLife*, 4, p.96.

Bomar, L. et al., 2011. Directed culturing of microorganisms using metatranscriptomics. *mBio*, 2(2), pp.e00012–11.

Bond, P.L., Smriga, S.P. & Banfield, J.F., 2000. Phylogeny of microorganisms populating a thick, subaerial, predominantly lithotrophic biofilm at an extreme acid mine drainage site. *Applied and environmental microbiology*, 66(9), pp.3842–3849.

Caporaso, J.G. et al., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5), pp.335–336.

Carlson, H.K. et al., 2015. Monofluorophosphate is a selective inhibitor of respiratory sulfate-reducing microorganisms. *Environmental science & technology*, 49(6), pp.3727–3736.

Christensen, G.A. et al., 2018. Use of in-field bioreactors demonstrate groundwater filtration influences planktonic bacterial community assembly, but not biofilm composition A. Franzetti, ed. *PloS one*, 13(3), p.e0194663.

Cole, J.R. et al., 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research*, 42(D1), pp.D633–D642.

Doebeli, M. & Knowlton, N., 1998. The evolution of interspecific mutualisms. *Proceedings of the National Academy of Sciences*, 95(15), pp.8676–8680.

Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10(10), pp.996–998.

Fadrosh, D.W. et al., 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*, 2(1), p.6.

Fernando, S.C. et al., 2015. Microbiota of the major South Atlantic reef building coral Mussismilia. *Microbial ecology*, 69(2), pp.267–280.

Friedman, J. & Alm, E.J., 2012. Inferring Correlation Networks from Genomic Survey Data C. von Mering, ed. *PLoS computational biology*, 8(9), p.e1002687.

Grosskopf, T. & Soyer, O.S., 2014. Synthetic microbial communities. *Current opinion in microbiology*, 18, pp.72–77.

Gruner, P. et al., 2016. Controlling molecular transport in minimal emulsions. *Nature Communications*, 7, p.10392.

Haas, D. & Défago, G., 2005. Biological control of soil-borne pathogens by fluorescent pseudomonads. *Nature Reviews Microbiology*, 3(4), pp.307–319.

Heitlinger, E. et al., 2017. The Intestinal Eukaryotic and Bacterial Biome of Spotted Hyenas: The Impact of Social Status and Age on Diversity and Composition. *Frontiers in Cellular and Infection Microbiology*, 7, p.689.

Hug, L.A. et al., 2016. A new view of the tree of life. *Nature Microbiology*, 1(5), p.16048.

Human Microbiome Project Consortium, 2012. Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), pp.207–214.

Justice, N.B. et al., 2017. Environmental Selection, Dispersal, and Organism Interactions Shape Community Assembly in High-Throughput Enrichment Culturing. H. L. Drake, ed. *Applied and environmental microbiology*, 83(20), pp.e01253–17.

Kozich, J.J. et al., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and environmental microbiology*, 79(17), pp.5112–5120.

Kurtz, Z.D. et al., 2015. Sparse and Compositionally Robust Inference of Microbial Ecological Networks C. von Mering, ed. *PLoS computational biology*, 11(5), p.e1004226.

Lax, S. et al., 2015. Forensic analysis of the microbiome of phones and shoes. *Microbiome*, 3(1), p.1048.

Ling, L.L. et al., 2015. A new antibiotic kills pathogens without detectable resistance. *Nature*, 517(7535), pp.455–459.

Loper, J.E. et al., 2012. Comparative Genomics of Plant-Associated Pseudomonas spp.: Insights into Diversity and Inheritance of Traits Involved in Multitrophic Interactions D. S. Guttman, ed. *PLOS Genetics*, 8(7), p.e1002784.

McHugh, J.P. et al., 2003. Global Iron-dependent Gene Regulation in Escherichia coli A NEW MECHANISM FOR IRON HOMEOSTASIS. *Journal of Biological Chemistry*, 278(32), pp.29478–29486.

Momeni, B., Waite, A.J. & Shou, W., 2013. Spatial self-organization favors heterotypic cooperation over cheating. *eLife*, 2, p.e00960.

Morris, B.E.L. et al., 2013. Microbial syntrophy: interaction for the common good. *FEMS Microbiology Reviews*, 37(3), pp.384–406.

Nichols, D. et al., 2008. Short Peptide Induces an "Uncultivable" Microorganism To Grow In Vitro. *Applied and environmental microbiology*, 74(15), pp.4889–4897.

Nichols, D. et al., 2010. Use of ichip for high-throughput in situ cultivation of "uncultivable" microbial species. *Applied and environmental microbiology*, 76(8), pp.2445–2450.

Park, J. et al., 2011. Microdroplet-Enabled Highly Parallel Co-Cultivation of Microbial Communities A. Herrera-Estrella, ed. *PloS one*, 6(2), p.e17019.

Poltak, S.R. & Cooper, V.S., 2010. Ecological succession in long-term experimentally evolved biofilms produces synergistic communities. *The ISME Journal*, 5(3), pp.369–378.

Price, M.N. et al., 2018. Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature*, 557(7706), pp.503–509.

Raymond, K.N., Dertz, E.A. & Kim, S.S., 2003. Enterobactin: an archetype for microbial iron transport. *Proceedings of the National Academy of Sciences of the United States of America*, 100(7), pp.3584–3588.

Riley, M.A. & Wertz, J.E., 2002. Bacteriocin diversity: ecological and evolutionary perspectives. *Biochimie*, 84(5-6), pp.357–364.

Rinke, C. et al., 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499(7459), pp.431–437.

Schloss, P.D., 2010. The Effects of Alignment Quality, Distance Calculation Method, Sequence Filtering, and Region on the Analysis of 16S rRNA Gene-Based Studies J. A. Eisen, ed. *PLoS computational biology*, 6(7), p.e1000844.

Schnupf, P. et al., 2015. Growth and host interaction of mouse segmented filamentous bacteria in vitro. *Nature*, 520(7545), pp.99–103.

Sukovich, D.J., Lance, S.T. & Abate, A.R., 2017. Sequence specific sorting of DNA molecules with FACS using 3dPCR. *Scientific Reports*, 7(1), p.39385.

Venturelli, O.S. et al., 2018. Deciphering microbial interactions in synthetic human gut microbiome communities. *Molecular systems biology*, 14(6), p.e8157.

Wetmore, K.M. et al., 2015. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio*, 6(3), pp.e00306–15.

Wolf, D.M. et al., 2008. Memory in Microbes: Quantifying History-Dependent Behavior in a Bacterium P. A. Silver, ed. *PloS one*, 3(2), p.e1700.

Wu, L. et al., 2015. Phasing amplicon sequencing on Illumina Miseq for robust

environmental microbial community analysis. *BMC Microbiology*, 15(1), p.125.

Zhang, J. et al., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics (Oxford, England)*, 30(5), pp.614–620.