Characterizing and Utilizing Droplet-enabled Co-cultivation to Elucidate Interactions in Microbiomes

by

James Yi Tan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Engineering)
in the University of Michigan
2023

Doctoral Committee:

Professor Xiaoxia Lin, Chair
Professor Mark Burns
Professor Gregory Dick
Associate Professor Evan Snitkin
Professor Michael Solomon

James Y. Tan

jamestan@umich.edu

ORCID iD: 0000-0002-0550-932X

Dedication

“Let the favor of the Lord our God be upon us,

and establish the work of our hands upon us;

yes, establish the work of our hands!”

Psalm 90:17 (ESV)

Acknowledgements

My degree itself is not the most valuable thing I have taken away from my time at the University of Michigan. My 10 years at this institution have grown me as critical thinker, intellectual, researcher, teacher, friend, mentor, and man of faith.

Ever since I started research as an undergraduate at the University of Michigan, I had incredibly talented and patient research mentors. The fact I still desire to be in academia is in large part due to these individuals. I thank my first research mentor, Dr. Timothy James. I was very untrained as a laboratory biologist (and was honestly very awkward), but Dr. James was incredibly supportive. I also thank Dr. Tara Webster and Dr. Lutgarde Raskin. Dr. Webster was my graduate student advisor throughout most of my time as an undergraduate researcher, and her enthusiasm contributed to me staying in research. There were personal circumstances that complicated my consistency, but both Dr. Raskin and Dr. Webster saw potential for me as a future doctoral student. That degree of belief in my capability was scary but extremely encouraging. I would like to thank Dr. Mark Burns and Dr. Mike Solomon for serving on my dissertation committee. Most of my meetings with them were very hasty and last-minute, but they provided their feedback to their best of their ability. Lastly, I have great respect and appreciation for my PhD advisor, Dr. Nina Lin. Not many principal investigators would provide the degree of academic freedom that she did, and it was through this that I really developed as an independent researcher. I can say very confidently that she cares about our

personal well-being just as much as our research progress. I am extremely thankful to have Dr. Lin's model: it is refreshing to see a professor at a research-intensive institution that values people so much, despite all the expectations placed on you.

I would like to thank Ford Motor Company for funding the first project I worked on as a PhD student. Dr. Haibo Zhao was our industry partner and was very understanding of the difficulties of juggling courses and research as a first-year student. The National Science Foundation funded work on the method development for droplet barcoding.

Microfluidics development is very difficult to do on your own. Nonformal feedback with Dr. Freeman Lan, Dr. Emilis Gegevicius from Droplet Genomics Inc., and Dr. Roger Nassar from RAN Biotechnology, was critical when I was stuck in workflow development. Their suggestions were immensely helpful and saved months of time.

My first-year transition from an undergraduate in Environmental Engineering to a graduate student in Chemical Engineering was a scary but rewarding one. The graduate courses in Chemical Engineering were very difficult, and my first semester I wrestled frequently with imposter syndrome. There were many nights with Christine Yee, Dr. Grace Tan, Cathleen Chong, Robert Luo, and others within my cohort who I could not have finished this coursework without. Fellow Chemical Engineering student Dr. Gerald Har become a friend during the pandemic years, where we discussed academia, faith, and the local church on breaks. I want to thank the administrative efforts of the ChE staff, especially Barbara Perry, Susan Hamlin, Pamela Bogdanski, Kelly Raickovich, and Benjamin Rodriguez who support the graduate students and the department behind the scenes. I would like to thank my labmates over the years: Dr. Steven Wang, Dr. Scott Scholz, Dr. Chun Wan, Dr. Corine Jackman, Dr. Tatyana Saleski, Dr. Adam

Krieger, Dr. David Carruthers, Dr. Winnie Song, Dr. Peter Mu, Erica Gardner, Sofia Camarero, Adena Collens, Jessica Li, Yanmeng Liu, Dr. Sandy Chiu, and Auden Bahr. As experimentalists, we spend a lot of time in the lab and an engaging and collaborative lab community is critical for a positive PhD experience. In particular, I would like to thank Dr. Saleski for her conceptual contributions to the first chapter of this thesis. Her perspective on the co-culture she developed was critical for the ideation of the project. I would like like to thank Auden Bahr and Jessica Li for helping me with last part of my dissertation, especially when the timeline was very tight.

The Integrated Training in Microbial Sciences (ITiMS) interdisciplinary program was a great community to fill the gap of microbially-minded individuals in my own department and provided funding. I am very thankful for Dr. Betsy Foxman and Dr. Thomas Schmidt for their administrative work to provide a certificate program and community for students and faculty studying microbial systems across the University of Michigan. Through the program, I was able to connect with many generous and kind-hearted faculty, such as my ITiMS Co-advisor Evan Snitkin who served on my committee and was quick to support, despite us never applying genome-scale metabolic modeling like I had originally proposed; Dr. Melissa Duhaime who graciously supported me in applying for a federal postdoctoral fellowship despite a week's notice; Dr. Gregory Dick who taught me in his geomicrobiology course in my undergraduate studies, trained me in metagenomics in his graduate course, served on my dissertation committee, and became my closest collaborator on the University of Michigan campus; and Dr. Foxman herself - for her desire to mentor and be present for all the ITiMS students as a fellow interdisciplinary scientist. Lastly, the students I have met though ITiMS have been

awesome: Dr. Frieda Blostein, Morgan Lindback, Thu Le, Alex Song, Jinny Yang, Mukai Wang, Nikesh Dahal, and others as well. I am very grateful for Dr. Derek Smith, whom I collaborated with in Dr. Dick's laboratory. We worked very well together in the individual analysis of single colonies of Microcystis and their accompanying microbiomes, and I really enjoyed that partnership.

I could never imagine Ann Arbor without Harvest Mission Community Church (HMCC). Truly, you have lived up to your commitment to "transform lost people who will then transform the world" in my own life. I have only realized how amazing it was to have been part of this community for a whole decade. I am very thankful to Rev. Peter Dahlem, Dr. Gina Dahlem, Pastor Joshua Yang, and Erica Yang, who have all served HMCC of Ann Arbor so faithfully during its transitions and challenges. I am thankful to have served alongside the other deacons and elders: Seong Park, David and Sarah Yon, Miesha White, Kristin and Solomon Zheng, Bryan Liu, Tiffany and Dennis Kim, Jennifer Zou, and Nate Jacobson. Your commitment to the local church is something I hope to keep in my life. I am thankful for other and past leaders who I have partnered with, including Michelle and Edwin Tang, Joseph Yang, Amanda Kortz, James Macey, Laura-Anne Wong, and Guy Lin. Jerry Wang, I accomplished things that I thought I would never do with you: break bad habits, run 20 miles, and memorize the Book of Galatians. Jerry Shi, you were a loyal friend since high school and your commitment to doing the most in your capability to overcome evil in the world is something I strive for. Matt Cui, I love that you call me to “shoot the breeze” and tell me about your life even when I rarely call you; your transparency and love for me keeps our friendship going even after you left. David Chang, we became so much closer than I ever would have

thought. Your principle and companionship are things that I take for granted and have learned to appreciate even more these days. Justin Rhee and Michael Baker, you guys have been the best roommates: no greater complimentary exists than between someone who loves playing video games and someone who really loves watching them. Joshua Kammeraad, the way you exemplify the Bible through your life is contagious, and I wouldn't be half the leader and friend I would be now without you. Anita and Peter Li, I respect you both so much and am so honored to be your friend, and I will miss you guys when we are not in the same vicinity or context anymore. Joshua Cheng, your approach to understanding the Bible has expanded my understanding of Christ's love and plan for the world. Amongst Impact, the graduate student ministry at HMCC, there are so many individuals that I wish I could elaborate on their impact and friendship in my life but cannot for sake of space: Dr. Nancy Wu, Torre Puckett, Zach Fritts, Hannah Abraham, Beatrix Yan, Dr. Grace Haeun Lee, Elaine Liu, Alex Wang, Dr. Sam Chen, Daniel Park, Joseph Tu, Dr. Jasmine Jones, Sara Timberlake, Hope Schaeffer, Alex Shen, John Wang, and so many others.

Regarding my family, it has been rough at times, but I am convinced of your love and sacrifice for me. I cannot imagine what it was like for my father to support a family of four with a doctoral student stipend while also getting his PhD at the University of Virginia in three years. My father was my role-model in scientific curiosity from a young age; he challenged me to question, learn, and push myself academically. My mother has always been the "spicy" one in my family, fighting for me to make a name for myself and believing that I had the potential to do whatever I put my mind to. I am very thankful for my sister Shu, who has taken care of me many times and is the only one in the world

who can understand our family struggles. Her husband Dr. Harrison Chen is someone I greatly appreciate for his wit and for taking me and my sister hiking in some of the coolest places in the world.

Finally, I can only give glory to Jesus, the son of God who gave himself for me and loved me. It was not enough that I might aspire for my own ambitions, but He gave me a reason to hope and in midst of confusion. I hope I have been able to demonstrate something about His love for the world through my actions, growth, and my work during my time as a doctoral student.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Microbiomes, the collective communities of microorganisms in a particular habitat, play critical roles in host-associated, natural, and built environments. The cell-cell interactions in microbiomes form expansive and complex networks, driving community structure and behavior. However, they are difficult to characterize, which prevents the rational design of interventions. As an emerging technology for elucidating interactions in microbiomes, microfluidic droplets (microdroplets), which are nanoliter-scale, monodisperse water-in-oil emulsions, show tremendous promise. Microdroplets enable reductionist, high-throughput study of interactions by encapsulating subsets of communities for co-growth and analysis. Yet, major barriers to realizing this potential include: (1) deficiencies in our understanding of how microdroplet parameters affect microbial co-growth, and (2) the lack of generalizable approaches for high-throughput and high-resolution characterization of co-cultivated sub-communities in microdroplets.

One critical parameter in droplet-enabled co-cultivation that has evaded evaluation is the droplet size. Given the same number of initial cells, a larger droplet increases the length scale secreted metabolites must diffuse and dilutes the initial concentration of cells, impacting community dynamics. To evaluate the effect of droplet size on a spectrum of syntrophic interactions, we cultivated a synthetic model system consisting of two E. coli auxotrophs, whose interactions could be modulated through supplementation of related amino acids in the medium. Our results demonstrate that the

droplet size impacts numerous aspects of the growth of a cross-feeding bi-culture, particularly the growth capacity, growth rate, and lag time, depending on the degree of the interaction. This work suggests that the droplet size should be more carefully evaluated based on the system of study or research objectives.

The first approach to address the lack of technical capabilities is the utilization of metagenomic shot-gun sequencing for individual droplets. We demonstrated this approach with the encapsulation and co-cultivation of droplet sub-communities from a human fecal sample. From a selection of 22 droplets, we observed this approach provides accessibility to previously uncharacterized gut commensals. We applied metagenomic sequencing for the de novo reconstruction of genomes from one droplet sub-community and demonstrated the capability to dissect sub-communities with high genomic resolution. Genomic characterization of one novel member of the family Neisseriaceae revealed novel pathways such as the production of atherogenic intermediates. Future adaptation and application of this approach would enable the inference of specific interactions based on genomic complementarity.

The second approach is droplet-resolved, quantitative 16S amplicon sequencing to profile the absolute abundance of thousands of droplet subcommunities. The eventual application of this methodology would be to infer interactions in co-cultivated droplet sub-communities based on membership and the degree of co-growth in each droplet. We developed novel microfluidic and molecular biology workflows for high-throughput, droplet-based barcoding of 16S sequences and the incorporation of molecular standards. We benchmarked the workflow throughput and accuracy with constructed mock communities. The benchmarking data showed good accuracy in

terms of relative community composition, but further work is needed to assess absolute quantification and improve even amplification and sequencing depth across droplets. Additionally, for wide applicability across biological systems, the degree of throughput and generalizability need to be improved.

In summary, we have expanded the understanding and technical capabilities regarding the utilization of microdroplets in deciphering interactions in microbiomes, enabling the technology to further advance fundamental discoveries in the field of microbiome science.

Chapter 1 : Research Background and Objective

1.1 Microbiomes

"Microbes rule the world."

Originally coined from Bernard Dixon in 1998 (1), this statement highlights the importance and prevalence of the microbial world despite remaining largely hidden from the naked eye. Our findings since then demonstrate that this statement has only been proved even more true. The collective of bacteria, fungi, viruses, and protists present in natural systems, coined the "microbiome", house the majority of the function (2–4) and diversity of life (5,6) and are the most likely progenitors of complex eukaryotic life (7). Microbiomes are generally categorized into host-associated, environmental systems, and engineered systems.

1.1.1 Host-associated systems

Just within the human body, bacterial cells outnumber human cells by a factor of 10 (8) and are essential for health. In the context of the gut, the microbiome contributes the majority of carbohydrate-degrading enzymes for the digestion of complex fibers that we consume (9,10) and act as our first primers for our immune response (11). The vaginal microbiome, in a healthy state, inhibits the growth of unwanted anaerobic bacteria and decreases the chances of contracting sexually-transmitted infections (12,13). Studies of the microbiomes of our closest evolutionary relatives – bonobos and

chimpanzees – reveal that our microbiomes have specifically co-evolved over millions of years with us (14), suggesting that this intimate partnership is a major factor in human evolution.

Studies into various animal microbiomes show intimate functional connections as well. Termites rely on highly efficient bacteria for the degradation of consumed wood and the fixation of nitrogen for growth (15). Many marine fish and squids acquire and house bioluminescent bacteria from surrounding water for cloaking strategies to evade predators (16,17). Even the simplest multicellular organisms, sponges, have dense microbiomes comprising 30% of their body mass. The sponge microbiome produces various biologically-active secondary metabolites to protect sponge from predators and acts as a major reservoir for natural products for biotechnological development (18).

Lastly, plant microbiomes are based on complex signaling systems for the selection of specific rhizosphere bacteria from the soil (19) and are known for their importance for disease-resistance (20), nitrogen-fixation (21), and even salt tolerance (22).

1.1.2 Environmental systems

Terrestrial, aquatic, and oceanic systems are major reservoirs for biogeochemical cycles (23,24). A stunning portion of biologically-mediated biogeochemistry is microbially-mediated (25).

For example, ocean microbiomes, which are estimated to be 90% of biomass in the ocean (26), are critical for the flux of global carbon. The cyanobacteria *Prochlorococcus* and *Synechococcus* are the most abundant photosynthesizers, and their activity is hypothesized to account for around 30% of all global primary production

(27). The sedimentation of microbial mass to the ocean floor is a long-term reservoir for the sequestration of carbon dioxide (28,29). However, the lysis of marine bacteria by bacteriophage though the “viral shunt” (30) greatly influences how much of this fixed carbon becomes sequestered and quantifying this is highly desired.

Aquatic freshwater systems, such as lakes and rivers, are similar to oceanic systems in regard to their contributions to biogeochemical cycles, but one major distinction is that aquatic systems are source waters for drinking water. The contamination of these bodies by certain pathogens or biological agents requires proper disinfection and treatment. For example, *Microcystis aeruginosa* is a globally-prevalent cyanobacteria and contributor of cyanobacterial harmful algae blooms (cHABs) which can produce potent toxins and compounds that foul freshwater bodies (31).

In terrestrial forests, the largest carbon reservoirs are bacterial and fungal biomass in soils (32). In permafrost, the rate of accumulation of carbon fixation greatly exceeds respiration, making these systems some of the largest global carbon reservoirs (33). Of major concern regarding global climate change is the observation that increasing temperatures increases rates of respiration, making these largely recalcitrant reservoirs labile (26). Additionally, soil microbiomes are a largely untapped source of natural products (34) and have provided novel families of antibiotics critical for outpacing the rate of antibiotic resistance (35). Lastly, the transformations of iron, especially in acid mine drainage (36), and many other heavy metals are microbially-mediated (37,38) and have important considerations in their mobility and toxicity.

1.1.3 Engineered systems

Many microbiomes include those that have naturally inhabited or are intentionally inoculated into anthropomorphically constructed infrastructure. This can include large centralized systems such as municipal drinking water treatment. While drinking water disinfection deactivates a large portion of organisms, they also select for organisms that can resist disinfection systems such as chlorination (39). Additionally, distribution systems for drinking water hold microbial biofilms that contribute to microbially-mediated corrosion of pipes (40). In addition to physiochemical processes for the removal of contaminants, many wastewater treatment facilities depend on microbiomes. For example, aerobic communities comprised mostly of *Zoogleoa* flocs to absorb the large amounts of organic carbon from municipal wastewater (41), while *Accumulibacter* removes phosphorus (42), and consortia under certain redox conditions remove nitrogen through nitrification and denitrification (43). Instead of only being used for removal, anaerobic microbiomes are also being used for the recovery of energy and nutrients from waste streams in processes such as anaerobic digestion or chain elongation (44).

Many microbiomes inhabit our built infrastructure, including the air and surfaces in our homes, transportation systems, schools, and workplaces. Increases in heat and humidity can increase the degree of metabolite production by the microbiome of our living spaces, which has been associated with certain diseases, such as eczema, onset and exacerbation of asthma, and hypersensitivity pneumonitis (45).

In inspiration from natural systems, the biochemical production of certain commodity chemicals and fuels from waste streams is performed by microbial communities (46,47). These systems are different in that they are not natural

communities but are curated and constructed synthetically for the optimization of specific metabolic outputs.

1.2 Microbial Interactions

To perform their ecosystem function in many of the above-described systems, microbiomes operate as highly interconnected and spatiotemporally complex units. Studies of multiple microbiomes show diverse physical arrangements and network structures, but most microbiomes follow a division of labor strategy at multiple levels. Typically, there is a linear trophic flow of energy from one level to the next. In Figure 1.1, this is illustrated as the degradation of complex polysaccharides to shorter chains, monosaccharides, and eventually fermentation to short-chain fatty acids, such as butyrate and acetate. Each step is performed by a different set of microbes at each trophic level. While this illustration is highly simplistic, some variation on this structure is observed in most natural systems. Another layer of complexity is the exchange of metabolites between different trophic levels. Auxotrophs, organisms that acquire required metabolites from the environment or from other organisms, are prevalent in natural communities and typically exchange amino acids and vitamins. Other complex interactions exist, such as inhibition in which organisms produce bacteriocidal compounds such as antibiotics to inhibit the growth of other organisms. This illustration also did not consider other essential elements, such as nitrogen, iron, and oxygen, which provide further dimensions of complexity.

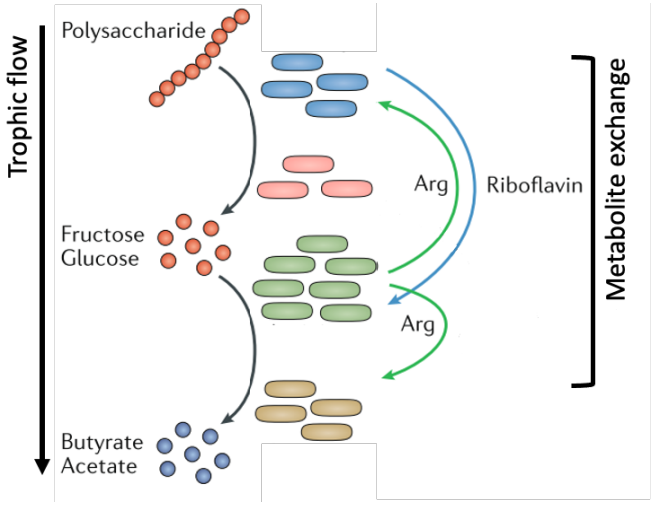


Figure 1.1: Example microbiome interaction network to demonstrate generalized microbiome structure, comprising of a linear trophic cascade for the flow of carbon and metabolite exchange between auxotrophs between the different trophic levels. "Arg" is shorthand for the amino acid arginine. Figure adapted from (48).

The exchange of metabolites between cells is generally referred to as syntrophy. The titular example of syntrophy is in anaerobic communities (Figure 1.2a): when secondary fermenters utilize short chain fatty acids, they produce hydrogen as a waste product which serves as an electron source for methanogens or sulfate-reducers who utilize this residual hydrogen for their own metabolism. This process is mutualistic: excess hydrogen is utilized which renders fermentation by syntrophs thermodynamically favorable and the hydrogenotrophic organisms receive their electron source (Figure 1.1b). This mutualistic metabolic cross-feeding is quite physically-intimate (49,50) (Figure 1.1c). Outside of this specific context, a very common observation in many microbiomes is that the production of amino acids or amino acid precursors is divided amongst many different species, with no one single species self-sufficient on the production of all amino acids (51). These syntrophic interactions are ubiquitous in microbiomes based on computation predictions (52) and observational studies (53–55).

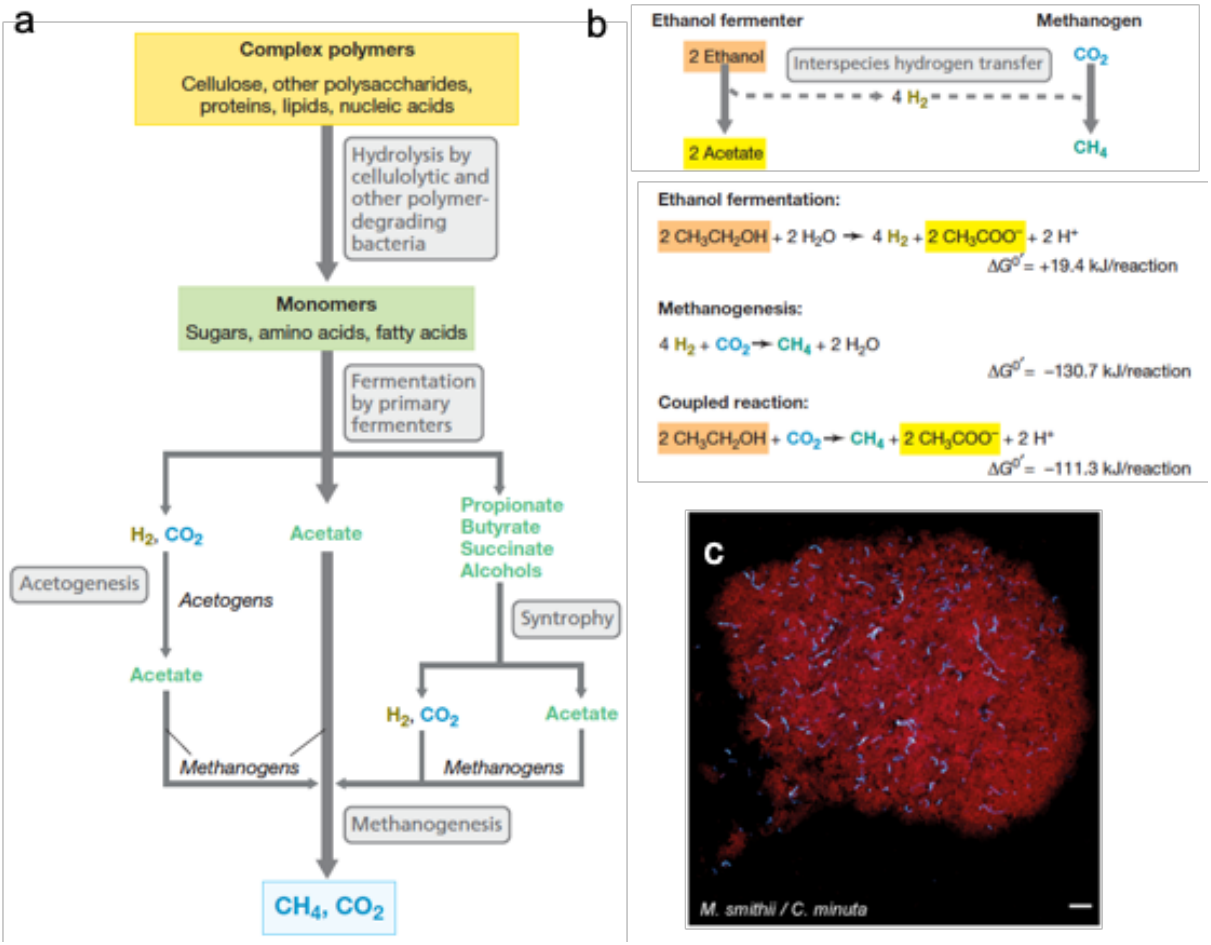


Figure 1.2: Microbial syntrophy in a natural context. (a) General trophic structure in anaerobic communities. (b) Example of syntrophic coupled metabolism in which interspecies hydrogen transfer removes the side-product hydrogen from ethanol fermentation and renders the metabolism as thermodynamically favorable. (c) Fluorescence microscopy demonstrating close physical association between methanogen Methanobrevibacter smithii *(blue) and acetate-producing and butyrate-producing* Christensenella minuta *(red). Scale bar represents 10 μm. Figures adapted from* (56)*. Microscopy image taken from* (50)*.*

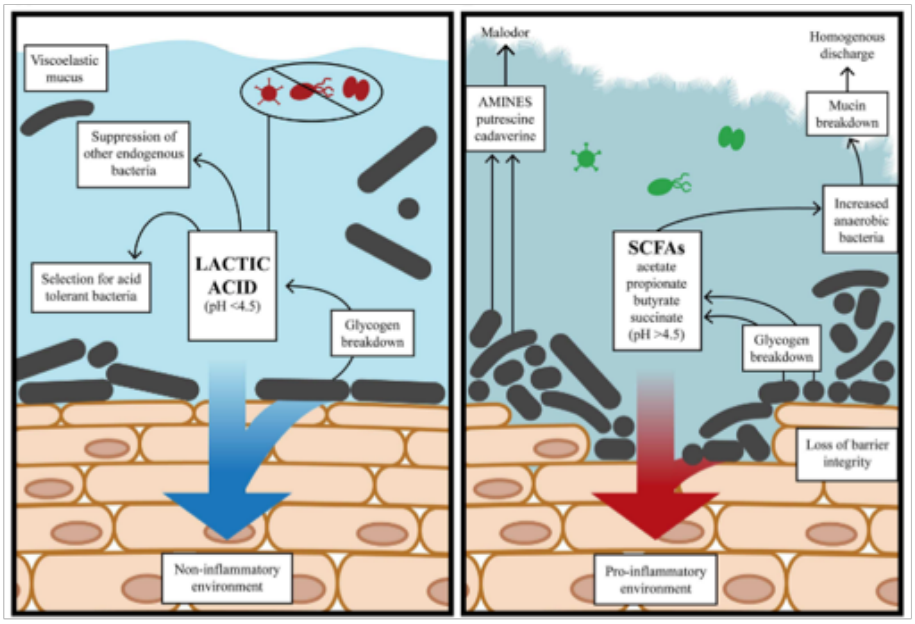
1.3 Predicting and Manipulating Microbiomes

While many microbiomes contribute to healthy functioning, many of them can also contribute to disease, ecosystem damage, and inefficiency within designed systems. Of particular interest are the unhealthy, "dysbiotic" states of the human

microbiome that lead to inflammation (57) and predisposition to disease (12,58) (Figure 1.3a). In environmental systems, due to warming temperatures and higher nutrient runoff, large, toxic cyanobacterial blooms caused by *Microcystis aeruginosa* and its associated microbiome are prevalent in freshwater lakes globally (59). One of the largest examples of this in Lake Erie of the Laurentian Great Lakes (Figure 1.3b) has resulted in the disruption of the ecosystem, drinking water systems, and economies that rely on the lake such as fishing and tourism (60). Lastly, in the context of wastewater treatment, the enrichment of a specific filamentous bacterium *Microthrix parvicella* causes bulking and foaming in secondary treatment processing, decreasing overall water quality (61) (Figure 1.3c).

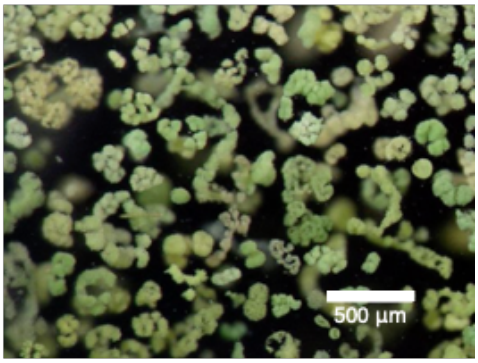
In each of these examples, there is desire to manipulate the system to shift the microbiome composition and function away from this “diseased” state to a desired “heathy” state. However, unlike traditional infectious diseases which rely on one causative agent which can be eradicated or inhibited to return a system to a healthy state, these systems are much more complex. They comprise of thousands to millions of species over spatiotemporal bounds with extensive interactions to form highly complex networks in which higher order, non-intuitive emergent outcomes result. With the perspective that they are ecological networks, they have proven notoriously difficult to understand and provide interventions for (62).

a



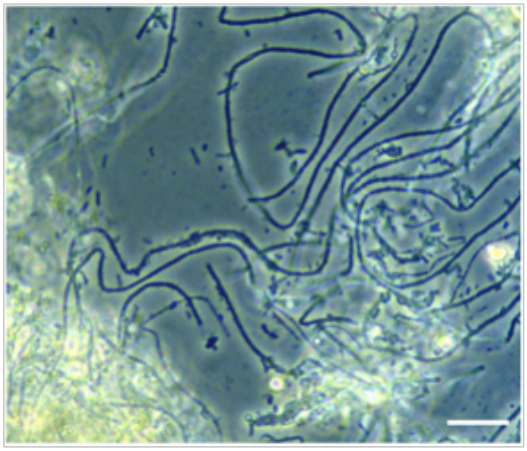
b

European Space Agency (ESA) Envisat (Oct 8, 2011)



NOAA Great Lakes Environmental Research Laboratory

c



Daniel Ayers, Rick Kelly (Brown and Caldwell Tracy Crane)

Figure 1.3: Examples of "unhealthy" dysbiotic microbiome states (a) The healthy (left) and dysbiotic (right) state of the human vaginal microbiome. The healthy state is characterized by higher abundance of Lactobacillus *sp. and a lower abundance of other anaerobic bacteria and other viral and bacterial sexually-transmitted infections. The dysbiotic state is characterized by the opposite: high abundances of anaerobic bacterium contributing to a pro-imflammatory response, mucin degradation, and increased susceptibility to infections. Figure adapted from* (12)*. (b) Large cyanobacterial harmful algal blooms in Lake Erie caused by* Microcystis aeruginosa*. (c) Foaming and bulking in wastewater treatment plants by* Microthrix parvicella*. Scale bar represents 10 μm. Figures adapted from* (56) *and* (61)*.*

Reaching a mechanistic and predictive understanding of microbiomes must be reached. However, even understanding one organism in isolation requires immense time and study. As you increase the number of organisms in the system, characterizing emergent properties from complex systems requires an understanding of not only the individual organism's behavior but also altered behavior as a result of the interactions between them. As a response to this need in 2016, the microbial ecology community called for two things: (1) new tools to understand microbial interactions and (2) the synthesis of experimental and observational studies and computational strategies to generate predictive models (63). Despite the advances since then, the challenges have hardly changed. What has been demonstrated more prominently is the methodological deficiency: we do not have the tools to study interactions between cells at the resolution needed for adequate microbiome science due to two primary challenges: (1) culturability and (2) complexity.

1.4 Barriers and Recent Advances

1.4.1 Culturability

Traditional microbiological techniques relied heavily on the ability to cultivate microbes on agar plates and liquid culture. However, one major challenge to this was the "Great Plate Anomaly", where the number of cells that could be counted on the plate

did not match the total number of metabolically active bacteria (64,65). The study of environmental microorganisms was limited greatly by the availability of medium formulations that could grow cells of interest, which were largely created through a laborious trial and error process.

However, with booming sequencing capabilities, it became possible to sequence environmental DNA extracted from cells. This "metagenomic" approach allowed for the study of both phylogenetic and functional diversity in an environment without the need to culture (66). Utilizing the 16S SSU rRNA gene sequence which is universally and phylogenetically conserved in bacterial and archaeal lineages, it was possible to identify which microbial species were present in a natural sample (67) (Figure 1.4a). While this answered the question of "who is there?" it provided very limited insight into functional potential. To provide more genome-resolved information, it became possible to utilize sequencing coverage (68) and genome-specific signatures, such as GC content (69) or tetranucleotide frequency (70) to resolve and reconstruct individual genomes *de novo* from metagenomic libraries (Figure 1.4b). These approaches expanded the study of microbiology away from the limited culture-limited view of the microbial world. Other culture-independent 'omic approaches, such as metatranscriptomics (71), metabolomics (72), and proteomics (73) sought to provide other critical information into gene expression, secreted and present metabolites, and present protein profiles, respectively of the environments.

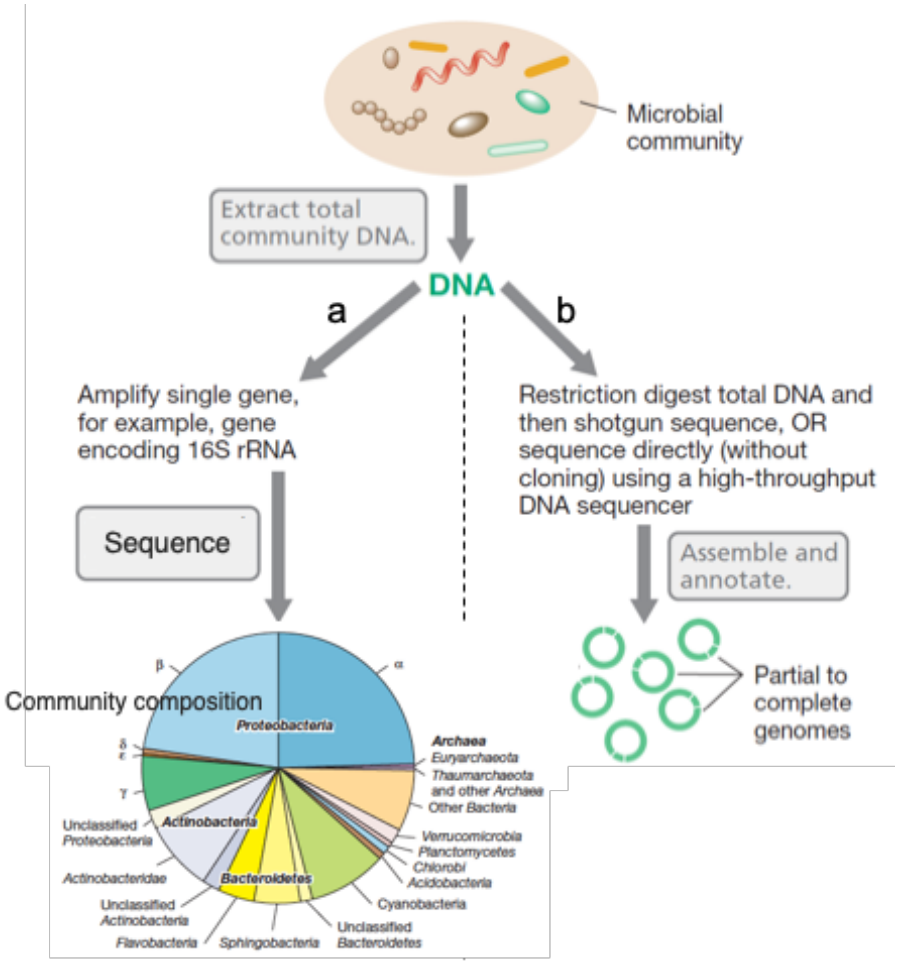


Figure 1.4: Overview of sequencing-based, culture-independent methods to study microbial communities. Both methods require the extraction of the total community DNA from a sample. (a) For phylogenetic characterization of a community, a single marker gene, commonly the gene encoding the 16S rRNA, is amplified with PCR, sequenced, and used with reference databases to provide community compositions. (b) The entire DNA is fragmented and shot-gun sequenced. De novo genome reconstruction can assemble draft genomes, revealing membership and functional potential. Figure adapted from (56).

In addition, other culture-independent assays to study spatial organization and function were developed. Fluorescence *in situ* hybridization (FISH) uses specific marker gene probes to identify cell types in natural samples, allowing the biogeographical study of microbial habitats and aggregates (74–76). Even further, in order to study certain metabolic activity on these spatial scales, research groups have developed tools such as nanoSIMS (77) and BONCAT (78) in combination with FISH to resolve to a

previously unprecedented degree the spatial complexity and functional activity of microbial habitats.

However, what is lost from all these culture-independent studies is phenotypic observations. Information regarding the microbe's growth rate, life strategy, specific metabolic physiology, and substrate preferences (79) is informative for understanding an organism's ecology and how it may have evolved to survive within a community and environmental context. Additionally, metabolic activity may not necessarily correspond to growth, and understanding growth in an environmental and community context is necessary for developing a predictive understanding of microbiome dynamics over time.

1.4.2 Complexity

As described before, due to the sheer number of individual members and spatiotemporal factors, culture-independent data is severely obscured. For example, without extensive sampling and differential coverage information (80), it is difficult to resolve strain-level draft genomes. There are also studies that have been able to resolve an ecological signal within genomes of the same species by differentiating lower coverage regions as genomic islands (81) and regions of high horizontal gene transfer (82), but these approaches require extensive sequencing effort.

Some ‘omics studies have yielded insightful findings by studying simple natural communities. This can be performed by studying smaller scales of spatial resolution to reduce the total complexity (83). Leventhal *et al.* did so by studying individual millimeter-scale granular biofilms to study community composition and diversity at the strain-level (84). In a collaborative effort with microbial ecologists, we also did a similar study on individual cyanobacterial-heterotroph aggregates (85).

Some groups have approached microbial ecology from another end, utilizing isolated, well-characterized microbes in simplified, constructed communities to reduce the complexity. For example, the Venturelli *et al.* has done extensive work to characterize pairwise interactions between a synthetic human gut community to provide a detailed interaction network constructed from growth-informed kinetic models and show that these pairwise interactions explain co-existence at the community level (86) (Figure 1.5). They then expand this framework with more community members to optimize a synthetic community for short-chain fatty acid production (87) and apply machine learning models (88). From this resolved analysis, certain properties of community behavior, such as coexistence, could be explained. Ponomarova *et al.* utilize a three-member synthetic community comprised of two lactic acid bacteria and *S. cerevisiae* to elucidate mutualistic interactions between the bacteria and yeast cell through physiological observations (89). At this resolution, carefully constructed experiments paired with metabolomics and transcriptomics and highly resolved genetic analyses provide a very high-resolution look into this mutualism that could not be determined from ‘omics studies of the natural community alone.

The main criticism of the synthetic ecology approach is its reductionism and the infeasibility of its scalability. Cultured representatives may not necessarily be good representatives of the system of study, and many natural systems simply lack cultured representatives. Additionally, first principles of microbial community behavior and assembly may be elucidated, but it is difficult to posit whether these principles in these simplified studies are even relevant in an *in vivo* context where complex spatiotemporal scales are at play. The scalability is also an issue; while all 45 pairwise interactions

between 10 members may be manageable, 100 members increases the total of possible pairwise interactions to 4950, which becomes much less manageable in terms of manual labor.

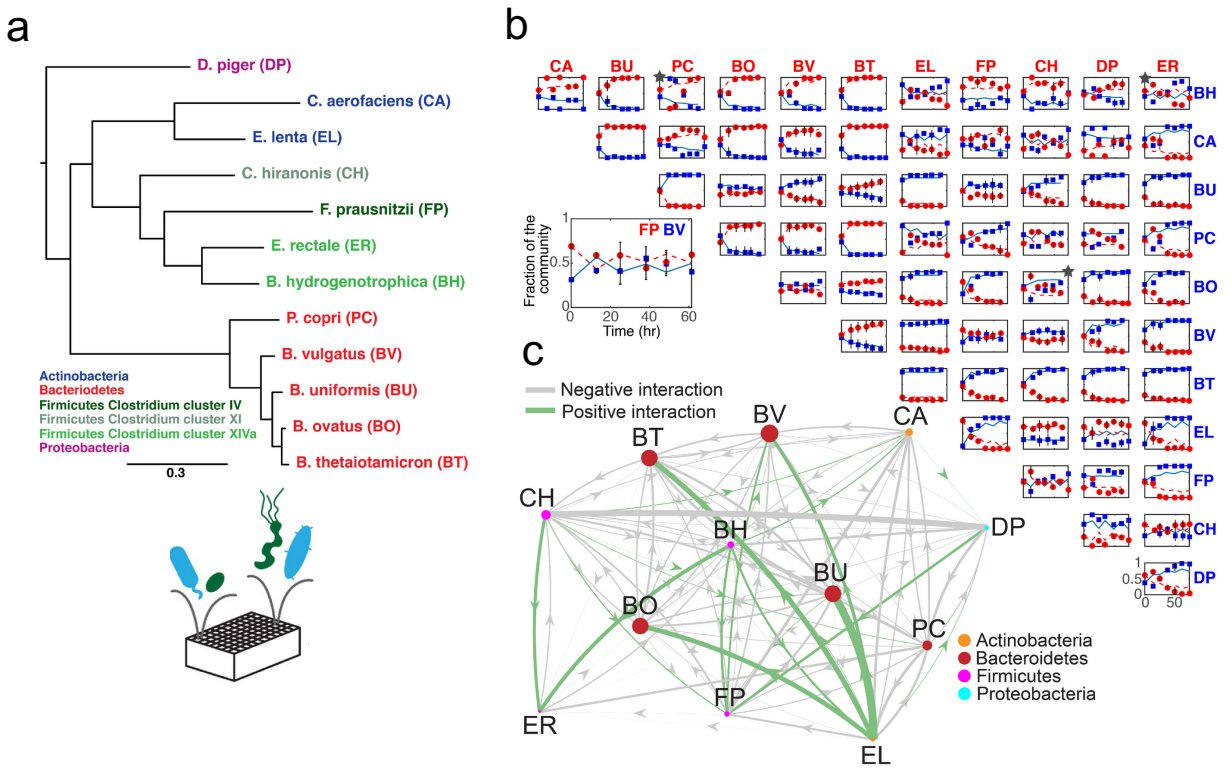


Figure 1.5: Study design from Venturelli et al. *to investigate microbial interactions and emergent community behavior with a synthetic ecology approach. (a) Venturelli* et al. *designed a 12 member synthetic community from human gut microbiome isolates. The synthetic ecology approach involves a high-throughput co-culture platform of different pairwise combinations to generate kinetic growth data. (b) Kinetic growth data from various pairwise combinations. Each growth plot is the relative abundance of the two members (red and blue) over time. (c) From the growth data, different interactions between the different members are inferred and general community structure is determined. Figure adapted from* (86).

1.4.3 The methodological divide

As such, the study of microbiomes is primarily divided into two methodological approaches (Figure 1.6): (1) top-down approaches such as the application of culture-

independent techniques to the study of complex natural communities and (2) bottom-up approaches which involve constructing synthetic communities from cultivated isolates to study emergent properties. Top-down approaches provide a sense of natural diversity, the communities of study are ecologically very stable and robust, and this scale allows modeling of ecosystem phenomenon. However, these studies lack strain-level and phenotypic resolution. Bottom-up approaches are very tractable and allow for genetic manipulation and resolved metabolic modelling, but are limited by cultivability, ecological translatability, and scalability. While both are key to understanding microbial communities, methodologies are needed that better bridge the gap between the reductionistic synthetic, laboratory-constructed nature of bottom-up approaches with the ecologically relevant observational, low-resolution nature of top-down studies.

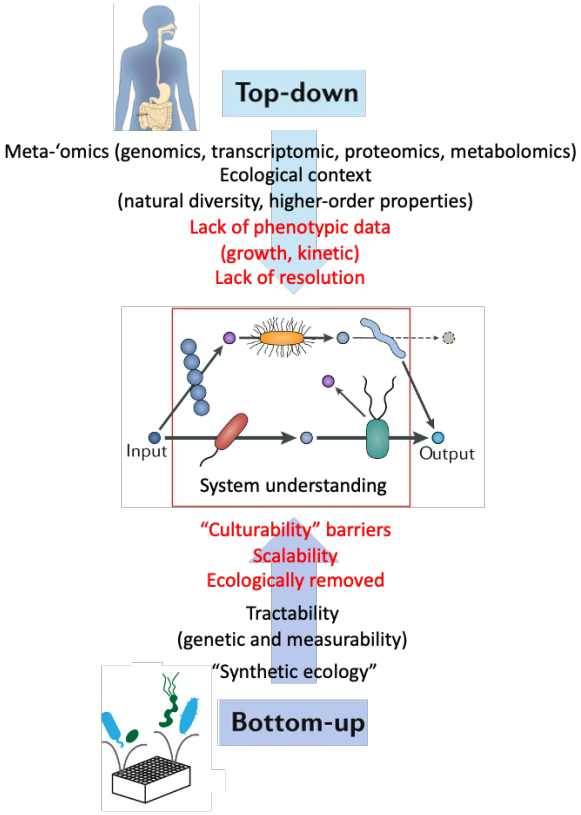


Figure 1.6: Overview of top-down and bottom-up approaches to study microbiomes. Figure adapted from (90) *and images are from Illumina and* (86)*.*

1.5 Microdroplets

Microfluidics provides a means to bridge that gap. Microfluidic droplets ("microdroplets") which are water-in-oil emulsions, typically between 40 and 150 microns in diameter, can be produced by co-flow through a narrow microfluidic orifice in an ultra-high-throughput manner – up to 1000 droplets per sec – (Figure 1.7a) and serve as miniaturized, localized bioreactors for microbial cultivation, rendering high-throughput investigation. Cells from a sample are encapsulated randomly according to Poisson statistics (Figure 1.7b) (91) and can be co-encapsulated and co-cultivated, resulting in high-throughput generation of droplet sub-communities. As a proof-of-concept, a fluorescent *E. coli* auxotrophic bi-culture system, one strain with a genetic knockout for tryptophan and the other with a genetic knockout for tyrosine, was grown in these microdroplets (92) (Figure 1.7c). Droplets with both members present had abundance co-growth while droplets with only one member did not.

Microfluidics provide the ability to study interactions at relevant length scales. While bacteria are typically only microns large, the smallest scale that we typically find comfortable to work with is at the microliter scale (e.g. micropipettes). Interrogating microorganisms at length scales comfortable for us has impeded our ability to study these systems intimately. Studies with interacting communities in biofilms demonstrate that the length scale relevant to interacting communities can be limited to a couple of microns (93). Because of the ability for limited encapsulation of a small number of cells to limit diffusion of molecules only to whatever is contained the droplet, microdroplets

are a great environment for the contained study of microbial interactions at the length-scale relevant for the subject of study.

Microdroplets can bridge top-down approaches and bottom-up approaches. In the study of natural communities, if the cells of interest are present in the sample at a reasonable abundance, co-encapsulation and the high-throughput generation of microdroplets can provide co-encapsulation of that cell type in pairwise or higher order combinations with other members of the community. Complexity is also quickly reduced, decomposing complex communities into smaller sub-communities with more discernable ecological signals that may be enabled by now-possible strain-level metagenomic reconstruction. Bottom-up approaches are benefited by the sheer throughput that increases the scalability of synthetic ecology studies by orders of magnitudes, making the study of more complex synthetic communities less manually intensive.

Hsu *et al.* applied the high-throughput nature of microdroplets to quantitatively infer interactions on fluorescence model cultures co-cultivated in different combinations and environments with imaging and cell quantification (94). While these were very important studies to demonstrate the potential of microdroplets, they were limited to cultured, genetically modified strains with fluorescence markers. The further application of microdroplets to deconvolute and study interactions in natural communities is technically challenging and requires methodological advances.

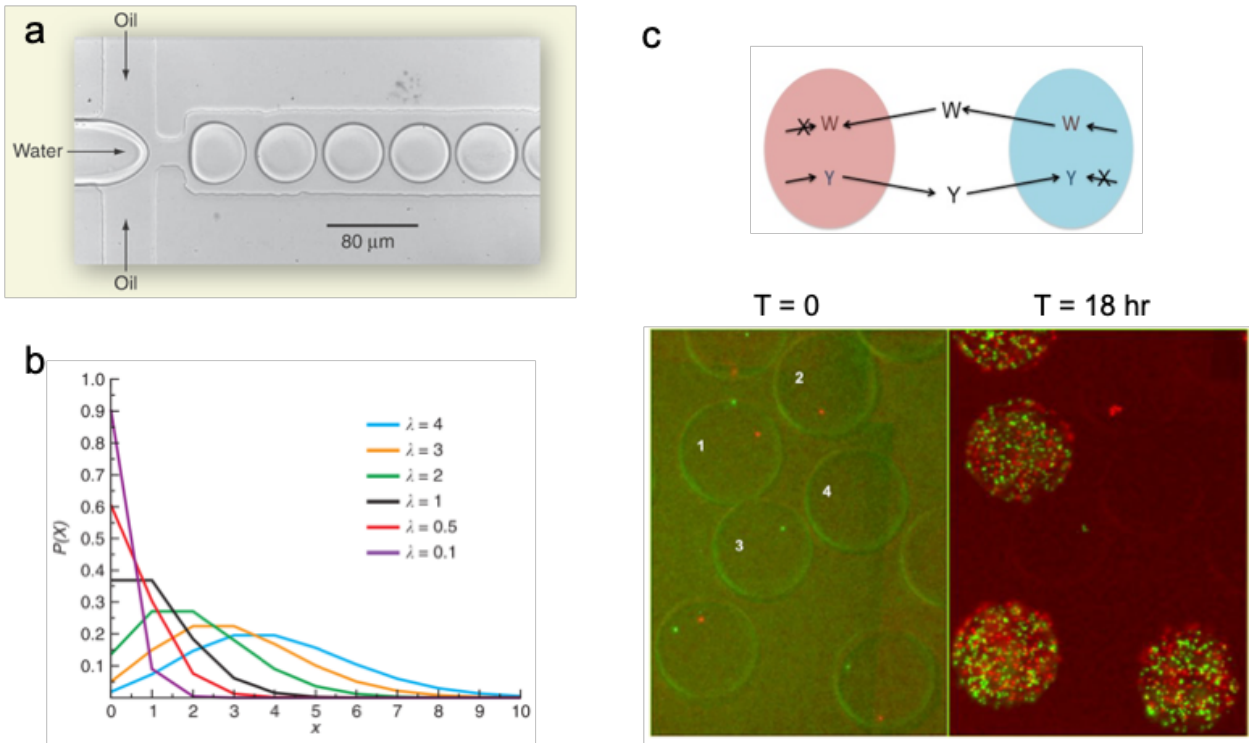


Figure 1.7: Microfluidic droplets (microdroplets) and their demonstration in the high-throughput co-cultivation of encapsulated cells. (a) Droplet generation in a two-phase water in oil flow-focusing device. Image from (95)*. (b) Poisson distributions with different λ values. The probability distribution function P(x) represents the probability of the number of x particles (e.g. cells) per droplet. Figure from* (91)*. (c) Co-cultivation of mutualistic co-growth between an auxotrophic pair of* E. coli*, one auxotrophic for tyrosine and the other for tryptophan. Both are distinctively fluorescently labelled. Figure from* (92)*.*

As already established, utilizing co-growth in microdroplets to study microbial interactions is a critical step in the microbial ecology toolkit. However, there are critical gaps in both our understanding and our methodology. The first gap is in our understanding of the experimental parameters in droplet cultivation. There are two main parameters in droplet cultivation, the number of average cells per droplet and the droplet size. Specifically, in droplet co-cultivation, when encapsulating cells in droplets for the study of pairwise interactions, the target is a low number of initial cells in each droplet to limit the number of cell types. However, droplet size is chosen based on

convention. In many laboratory groups – including my own – size is based on previous studies and the size of droplets that can be produced in-house. The assumption is that droplet size does not have large difference on microbial co-cultivation outcome. This is a spurious assumption: size affects the effective initial cellular concentration, which has drastic effects on the outcome of the co-cultivation studies, yet very few studies have looked into these effects in a droplet context.

The second gap is the lack of technical capability to apply sequencing technologies to study droplet sub-communities. Applying sequencing will extend microdroplet-enabled interaction inference beyond synthetic, fluorescently labelled communities. While many techniques such as 16S amplicon sequencing and metagenomic shot-gun sequencing make it theoretically possible to study natural sub-communities, no one has incorporated these technologies into the microfluidic droplet workflow. Given a fixed amount of sequencing effort, there are two approaches in applying this technology: (1) studying a smaller number of droplets, one-by-one with metagenomic shot-gun sequencing and (2) profiling droplets in a high-throughput manner in parallel with 16S amplicon sequencing. The first approach allows for genomic-level resolution and the inference of metabolic capabilities of whatever community is being studied. However, metagenomic reconstruction requires a comprehensive sequencing effort. With a limited sequencing effort, shot-gun sequencing a large number of droplets may not recover enough sequencing coverage and genomic signal per droplet for effective metagenomic reconstruction per droplet. In the second approach, a more high-throughput analysis of droplets is possible with 16S marker gene sequencing. However, inference of the interaction mechanisms based on

genomic information is not possible if metagenomic genome reconstruction has not been performed on the community beforehand. Both approaches require technical advances beyond what is capable currently.

1.6 Dissertation Overview

This dissertation provides advances addressing these gaps and is summarized in Figure 1.8. The main body of the dissertation is comprised of Chapters 2-4: Chapter 2 centers on the investigation of the droplet parameters on droplet co-cultivation, and Chapter 3 and 4 discuss method development for the study of microbial interactions with co-growth in microdroplets. More specifically, Chapter 2 details the effect of changing the droplet size on the growth of syntrophic microbial co-cultures. Using fluorescently labelled co-cultures, the growth of each member can be tracked, parameterized, and compared under different conditions. In the analysis, droplet-to-droplet variation of co-cultures and monocultures is also measured and studied. Chapter 3 and 4 address methodological advances to study co-growth in a small number of microdroplets with metagenomic shot-gun sequencing and in a larger library of droplets with 16S amplicon sequencing, respectively. Chapter 3, the investigation of a small number of droplets with metagenomic shot-gun sequencing and genome reconstruction, was performed in collaboration with Dr. Sida Wang and applied to human fecal samples as a demonstration on a natural community. It involves the utilization of droplet-separation microfluidic technology that Dr. Wang developed and a bioinformatics pipeline using available, open-source metagenomics software. Chapter 4, the study of growth in a larger number of droplets with 16S amplicon sequencing incorporates the development of droplet-barcoding sequencing technology to preserve a

signal for the resolution of individual droplets as well as a quantitative signal so that droplets with growth could be distinguished from droplets without growth. A preliminary benchmarking was performed with mock communities of different cell types and abundance profiles to evaluate the accuracy and sensitivity of the method. Chapter 5 is a summary of the dissertation work, perspectives on future work, and reflections on performing interdisciplinary work within the scientific community.

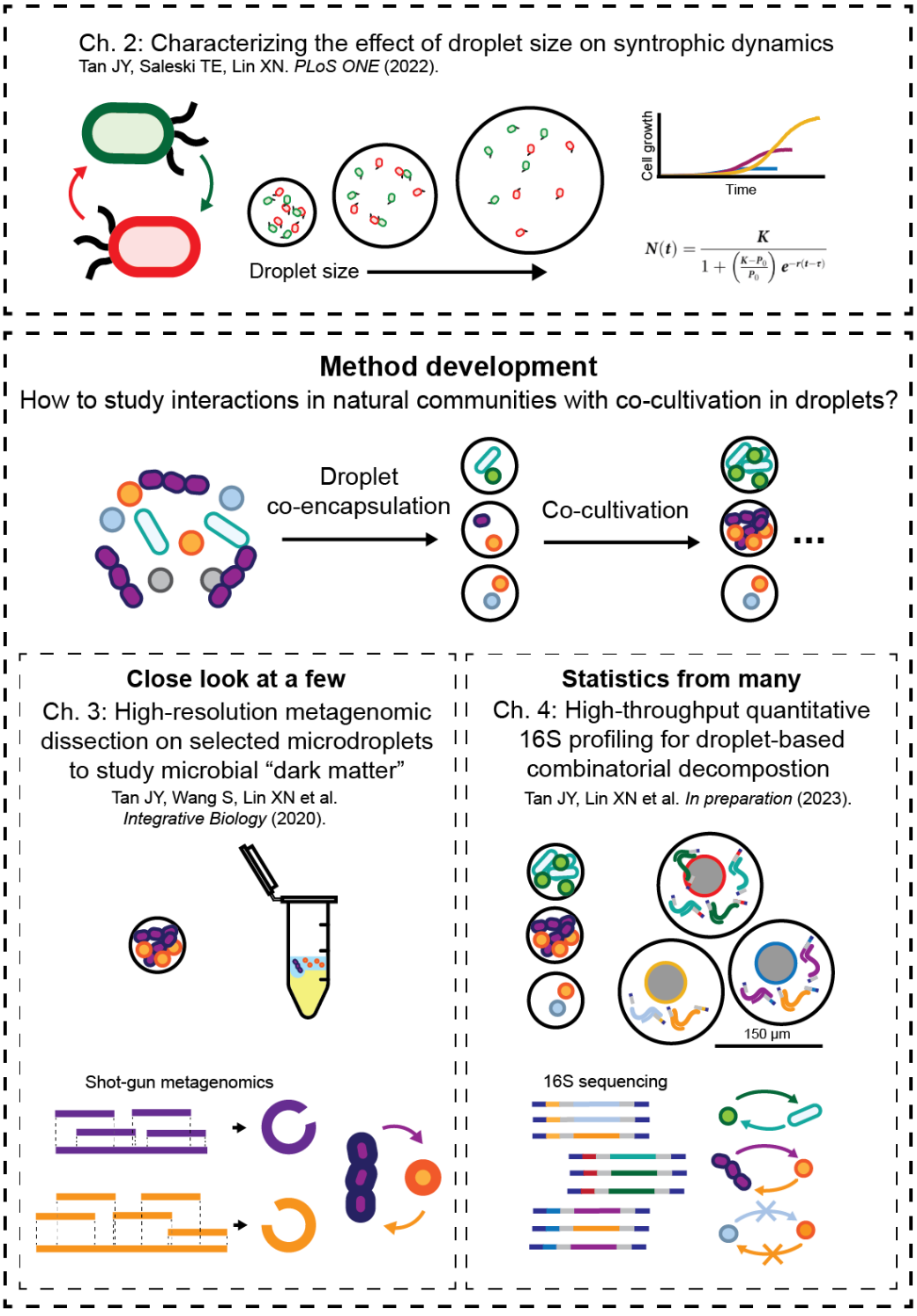


Figure 1.8: Overview of the dissertation. Chapter 2 is an investigation into the microdroplet as a co-cultivation environment for syntrophic growth by using variable droplet size and fixed number of initial cells and fluorescence assays to measure growth response. Chapter 3 and 4 are focused on methodology development to analyze microbial interactions in microdroplets. Chapter 3 is about the study of a small number of droplets with deep shot-gun metagenomic sequencing. Chapter 4 is concerning the high-throughput study of droplets with 16S amplicon sequencing and droplet barcoding.

Chapter 2 : Characterizing the Effect of Droplet Size on Syntrophic Dynamics

This chapter is modified from a previously published work (96).

2.1 Summary

A critical parameter in droplet-enabled co-cultivation that has evaded appropriate evaluation is the droplet size. Given the same number of initial cells, a larger droplet size can increase the length scale secreted metabolites must diffuse as well as dilute the initial concentration of cells and exchanged metabolites, impacting the community dynamics. To evaluate the effect of droplet size on a spectrum of syntrophic interactions, we cultivated a synthetic model system consisting of two *E. coli* auxotrophs, whose interactions could be modulated through supplementation of related amino acids in the medium. Our results demonstrate that the droplet size impacts substantially numerous aspects of the growth of a cross-feeding bi-culture, particularly the growth capacity, maximum specific growth rate, and lag time, depending on the degree of the interaction. This work heavily suggests that one droplet size does not fit all types of interactions; this parameter should be carefully evaluated and chosen in experimental studies that aim to utilize droplet-enabled co-cultivation to characterize or elucidate microbial interactions.

2.2 Introduction

One key assumption in previous studies utilizing microdroplets to co-cultivate and analyze microbial communities is that the droplet environment experienced by the co-encapsulated cells is representative of relevant bulk conditions. This assumption has not been fully validated or assessed in previous studies. In natural systems, transport of extracellular metabolites occurs along a continuum between convective and diffusive mass transfer. For example, the human gut is relatively homogenous and well-mixed (74), whereas in marine environments, chemotaxic symbionts exploit chemical gradients primarily established through diffusion (97). In homogenous, well-mixed environments, convective mass transfer is efficient and is seldom a limiting factor. In contrast, in the microdroplet environment, unless actively applied, convective mass transfer is not present. At micron-scales, nevertheless, diffusion is able to transfer mass in relatively short times. Based on statistical mechanics, the time (t) required for a molecule to travel a distance (L) in three dimensions can be calculated as $t = L^2/6D$. For instance, in specific scenarios investigated in this study where valine serves as a metabolite exchanged between cells and has a diffusivity (D) of approximately 800 μm^2/sec (98), we can estimate that the time taken for a valine molecule to travel the maximum distance in a droplet increases from ~0.63 second in the smallest droplets we examine, of 55 μm diameter, to ~4.7 minutes in the largest ones, of 150 μm diameter. It remains to be investigated whether or not the potential delay incurred by diffusion in droplets of larger sizes is significant enough to alter growth dynamics.

In designing experiments for the growth of microbial communities in microdroplets, two parameters to consider are: a) the lambda (λ) value, i.e. the average

number of cells per microfluidic droplet according to the Poisson distribution (91), and b) the droplet size. These two parameters combined govern the cellular density, which in turn determines the average path length required for diffusion to facilitate transport. The λ value is manipulated to control the number of initial cell density given a droplet size. This value is chosen based on the objective of the experiment. If the objective is single cell encapsulation, λ can be set at 0.1 or 0.2 (i.e. 1 out of 10, or 1 out of 5 droplets contains one cell on average). However, if the objective is co-culturing, the λ could be set at 5 for all cell types so that each droplet contains at least one of each cell type for proper representation of diversity in every droplet. Typically, the λ value is kept as low as possible to maintain single-cell or small population resolution, keeping in accordance to the sensitivity and capacity considerations of microdroplets. For droplet size, there have been a wide range of droplet sizes used in microbial co-cultivation studies, such as 40 μm (99), 55 μm (100), 120 μm (92,101), 135 μm (102), and 150 μm (103) for different types of communities with different mechanisms of interaction. The droplet size utilized for a microfluidic study is highly dependent on the needs of downstream droplet processing, such as whether the droplet will fit into channels or be compatible with optics and sensors, as in the case for fluorescence activated droplet sorting (FADS) (104). As a result, droplet size can be limited by experimental constraints rather than being evaluated as a key factor for the biological system in question. However, the optimal droplet size for the system of study is not a trivial parameter to determine. This can be highly dependent on multiple factors: the secretion and uptake rate of metabolite exchanges, the diffusivity and affinity of the metabolite being exchanged, and the cellular requirement for the metabolite in question.

We hypothesized that metabolite exchanges in one droplet size may elicit very different growth dynamics from the same interacting co-culture in another droplet size due to altered length scales relevant for diffusion. To systematically evaluate the effect of droplet size on co-culture dynamics, this study utilizes a two-member *E. coli* amino acid auxotroph cross-feeding system (100) in a range of droplet sizes. These two *E. coli* strains are fluorescently labelled to allow real-time monitoring of growth dynamics. We first examine the growth of a monoculture in different droplet sizes as a baseline. Then, we study the two-member co-culture in a similar manner to study the dynamics of growth as a result of amino acid exchange and the effect of droplet size on these dynamics. We also alter the environment to modulate the extent of cross-feeding and investigate three scenarios (low, medium, and high degrees of interaction). Overall, we characterize the difference in dynamics across different droplet sizes using this model system, demonstrating that droplet size has a profound impact on co-culture dynamics. In addition, we observe that the low initial cell number arising from droplet encapsulation leads to a significant degree of droplet-to-droplet stochasticity, which should be taken into consideration while analyzing growth patterns in droplets.

2.3 Materials and Methods

2.3.1 Strains and culturing conditions

The *E. coli* strains, S1 Δ*ilvD* and S2 Δ*lysA*, were constructed by Saleski et al. (100). The full genotype of S1 Δ*ilvD* is K12 Δ*ilvD*::FRT Δ*galK*::cfp-bla pSAS31, and S2 Δ*lysA* is JCL260 Δ*lysA::*FRT pSA69/pBT1-proD-mCherry, derived from JCL260 (105,106). pSAS31 was constructed by Schott Scholz (107), and pBT1-proD-mCherry was acquired from Michael Lynch (Addgene plasmid #65823). Both strains have

constitutively-expressed fluorescent protein reporters, mNeonGreen and mCherry for S1 Δ*ilvD* and S2 Δ*lysA*, respectively.

Strains were maintained as glycerol stocks kept at -80°C. New cultures were inoculated from cryostocks into LB broth (Miller) with appropriate antibiotics and incubated overnight at 37°C at 250 rpm. S1 Δ*ilvD* was grown on ampicillin (100 µg/mL) and kanamycin (50 µg/mL), while S2 Δ*lysA* was grown on ampicillin (100 µg/mL), kanamycin (50 µg/mL), and tetracycline (10 µg/mL). 1 mL of each culture was harvested by centrifugation at 4,000 g for 5 minutes, washed twice, and resuspended in 1 mL M9. To determine cellular concentration, both strain suspensions were quantified through cell counting on a disposable C-Chip haemocytometer (SKC Inc, C-Chip) through phase contrast light microscopy on an inverted light microscope (Nikon Eclipse Ti-S). Based on the volumes of a 55 µm, 75 µm, 100 µm, 125 µm, 150 µm diameter sphere, 5 different suspensions were created to achieve a λ value of 5 cells of each strain per droplet.

M9, consisting of M9 salts (47.8 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.55 mM NaCl, 9.35 mM NH_4Cl, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$), micronutrients (2.91 nM $(NH_4)^2MoO_4$, 401.1 nM H_3BO_3, 30.3 nM $CoCl_2$, 9.61 nM $CuSO_4$, 51.4 nM $MnCl_2$, 6.1 nM $ZnSO_4$, 0.01 mM $FeSO_4$), thiamine HCl (3.32 µM) and dextrose (D-glucose) at 5 g/L, was used as the base medium. When both strains were grown together, ampicillin and kanamycin were supplied at the concentrations previously specified. Amino acids, when specified, were supplemented as follows: (1) 3 mM isoleucine, (2) 3 mM valine and 3 mM leucine, and (3) no additional amino acids. All amino acids were the enantiopure L-isomer.

When S2 Δ*lysA* was co-cultivated with S1 ΔilvD, 0.1 mM IPTG was added to induce amino acid production in S2.

2.3.2 Microfluidic device fabrication

To fabricate the microfluidic devices, polydimethylsiloxane (PDMS) base elastomer and curing agent (10:1 ratio of elastomer to curing agent by mass) was poured onto SU-8 molds with the microfluidic device features (Fig S1). The molds with uncured PDMS were vacuumed to remove air bubbles, and heated at 65°C overnight to solidify the polymer. The PDMS layer was removed off of the SU-8 molds, and devices were cut to size. To complete the fabrication, the devices were punched with holes by a biopsy punch (1.0 mm inner diameter) to create openings for channel flow, and bonded on the PDMS base via plasma-activated bonding using a corona discharge wand. The devices were silanized with (tridecafluoro-1,1,2,2,-tetrahydrooctyl)-1-trichlorosilane using a desiccator, and used for droplet generation as described below.

2.3.3 Microfluidic droplet generation

Strains were diluted to achieve a λ of 5 cells per droplet for each strain for 55 μm, 75 μm, 100 μm, 125 μm, and 150 μm droplets. The oil phase was composed of Novec HFE-7500 fluorinated oil (3M) with 2% PEG-PFPE amphiphilic block copolymer surfactant (Ran Biotechnologies, 008-FluoroSurfactant). The aqueous cell suspension and the oil phase were loaded into 1 mL and 3 mL syringes, respectively, with 23-gauge Luer Lock syringe needles (BD 305145) attached. PTFE tubing (0.022” ID, Cole-Parmer) was used to connect the syringe needle to the droplet generation device. Kent Scientific GenieTouch syringe pumps were used to infuse the oil phase and cell

suspension phase into the device to generate the droplets. Two flow-focusing droplet generation devices of different channel dimensions (Figure A.1) were used with different oil/aqueous flow rates to create the full range of droplet sizes. For all droplet sizes, the aqueous phase flow was fixed at 20 µL/min aqueous, but the oil phase flow was adjusted to achieve the desired size. For 55, 75, 100, 125, 150 µm diameter droplets, the oil phase flow was set to 45, 22, 30, 19, 12 µL/min, respectively. For each droplet condition, approximately 450 µL of droplets were collected for approximately 23 minutes after steady-state droplet generation was reached in 1.5 mL microcentrifuge tubes from the outflow of the droplet generation device through additional PTFE tubing. Excess oil was removed to improve pipetting of droplets into microwells in 96-microwell plates. 100 µL additional fluorinated oil with 2% surfactant and 100 µL of droplets were carefully pipetted into individual wells in a black clear-bottom microplate (Greiner 655090) and sealed with a Mylar plate sealer (Thermo-Scientific™ 5701). The plate was loaded into a microplate reader (BioTek Synergy H1) at 37°C reading fluorescence green fluorescence (excitation and emission wavelength: 450 nm, 550 nm, respectively) and red fluorescence (excitation and emission wavelength: 587 nm, 615 nm, respectively) every ten minutes.

2.3.4 Image analysis

For imaging of droplets, 10 µL of droplets/oil were pipetted into a C-Chip chamber, sealed with epoxy, and incubated a 37°C incubator. To ensure that droplets were not squeezed under visualization in the C-Chip, the “Neubauer Improved” grid type was used for droplets with diameters 100 µm or less and the “Fuchs Rosenthal” grid type was used for droplets with diameters larger than 100 µm. Phase contrast and

fluorescence images were taken by viewing the droplets with an inverted light microscope (Nikon Eclipse Ti-S) with a Nikon Intensilight C-HGFI epi-fluorescence illuminator. Red and green filters were used in conjunction for fluorescence imaging of the mCherry and mNeonGreen fluorescent cells, respectively. Images were taken with QCapture Pro 7 software using the QImaging EXi-Blue fluorescence microscopy camera with standardized laser intensity and exposure settings of 250 ms and 500 ms for green and red fluorescence, respectively with an objective lens of 10X. All imaging, besides the initial timepoint due to low cell densities and fluorescence, followed these settings to standardize fluorescence intensity.

Images were processed to increase the brightness of raw output images from QCapture using ImageJ (2.0.0-rc-69/1.52i) to the same degree of augmentation for all green images and for all red images to maintain standardization. The red and green images were merged into one image for each droplet/time point. Using custom scripts in MATLAB, these fluorescence and phase contrast images were analyzed to extract the fluorescence intensity normalized of over 100 droplets for droplet sizes of 55, 100, and 150 μm diameter. Fluorescence intensity of individual droplets was normalized by the area of the respective droplet in the image. The processed images and MATLAB scripts are available at: https://github.com/jamesyitan/coculture-droplet-size.

2.3.5 Growth kinetics analysis

Growth kinetic data during cultivation in droplets was recorded through the BioTek Synergy H1 microplate reader. In most experiments, a fixed λ value was employed across different conditions and hence the initial fluorescence/volume of droplets across droplet sizes are different. For comparison across different droplet

sizes, the background fluorescence was removed from growth data and were normalized by the fluorescence value at the initial time point, which resulted in growth curves representing fold changes and starting at a value of 1.

Growth curves of monocultures were fitted to the logistic growth equation in MATLAB using the curve fitting toolbox:

$$\boldsymbol{N(t) = \frac{K}{1 + \left(\frac{K - P_0}{P_0}\right) e^{-rt}}}$$

where N(t) is the population size at time t, K is the carrying capacity, P_0 is the initial population (which is set to 1 due to the normalization), and r is the maximum specific growth rate. Because of the long lag periods observed in the growth curves, we modified the logistic growth equation to account for lag time:

$$\boldsymbol{N(t) = \frac{K}{1 + \left(\frac{K - P_0}{P_0}\right) e^{-r(t-\tau)}}}$$

where τ is the lag time, the duration from inoculation to start of the exponential growth phase. A curve fitting function in MATLAB would fit growth curve data to this modified logistic growth equation and generate three parameter values: K, r, and τ. The raw and normalized data, as well as the MATLAB scripts and functions used to calculate these values, are available at https://github.com/jamesyitan/coculture-droplet-size.

For cross-feeding bi-cultures, it was not obvious how the growth should be quantified. Nevertheless, it was observed that the growth profile of each auxotroph in the bi-culture exhibited the "S"-shape, characteristic of logistic growth commonly assumed for monocultures. To determine whether or not growth dynamics of the cross-feeding auxotrophs can be empirically approximated by the logistic equation, we carried

out dynamic simulation of the cross-feeding bi-culture using an extended version of the ODE model in Kerner *et al.* (108) and fit the resulting growth curves to the logistic equation (Note A.1 and Figure A.2). It was found that the logistic equation was a satisfactory fit. Therefore, despite differences in the exact molecular and cellular process underlying bi-culture and mono-culture dynamics, throughout this work, we use the logistic equation to empirically characterize cell growth and fit associated model parameters to the experimental data. In addition, the ODE model was used to investigate the effect of increasing droplet volume on growth kinetics by varying the initial cell density. The MATLAB scripts and functions for the dynamic simulation are available at: https://github.com/jamesyitan/coculture-droplet-size.

2.4 Results

2.4.1 Monoculture growth dynamics

Before we investigate co-cultures consisting of cross-feeding partners, it is important to understand the effect of droplet size on simple monocultures as a baseline. We therefore selected a fluorescently labeled *E. coli* strain we developed previously (100), S1 Δ*ilvD*, and cultivated it axenically in microdroplets of three sizes (55 μm, 100 μm, and 150 μm in diameter) with a λ value of 5 cells/droplet in M9 medium supplemented with amino acids required by this strain (i.e. 3 mM each of isoleucine, leucine, and valine). A λ value of 5 cells/droplets was specifically chosen to encapsulate a small number of cells in each droplet while allowing almost all droplets to have at least one cell.

As shown in Figure 2.1a, for all three droplet sizes, extensive growth was observed in a majority of the droplets. In the meantime, it was noted that there was

substantial droplet-to-droplet heterogeneity after cultivation. The distributions of droplet fluorescence normalized by droplet area were empirically determined and showed quantitatively the heterogeneity (Figure 2.1b). These histograms reveal that across different droplet sizes, the ranges of final cell density were similar, but the exact distribution within the range was dependent on the droplet size. The distributions of 55 μm and 100 μm diameter droplets were comparable (student's two-tailed t-test, p-value = 0.509), but the difference between 150 μm droplets and the others (100 μm, p-value = 0.041 and 55 μm, p-value = 0.079) was more pronounced. In particular, the distribution of 150 μm droplets exhibited a much larger variance.

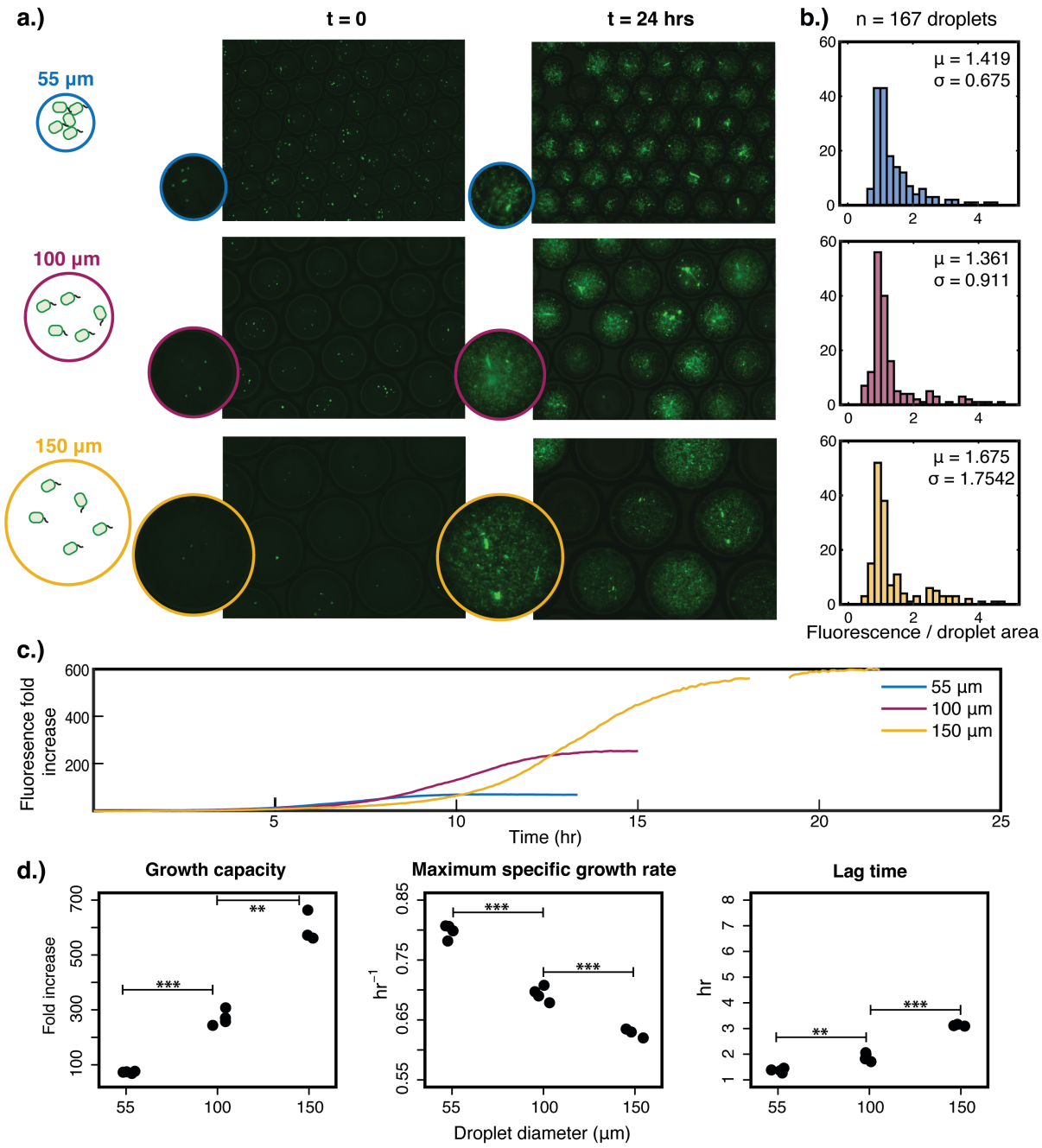


Figure 2.1: Monoculture growth of S1 ΔilvD in a range of microdroplet sizes with λ=5 cells/droplet. (a) Representative images of overlays of fluorescence microscopy and phase contrast at the initial and post-cultivation time points for microdroplets of diameters 55, 100, and 150 µm. (b) Histograms of the post-cultivation fluorescence normalized by droplet area (a.u./pixel) for images of microdroplets for a total of 167 droplets analyzed per droplet size, with associated mean and standard deviation. (c) Growth curves averaging replicate wells containing monoculture microdroplets for each droplet size. Within a single well in a 96-well plate, the number of droplets can range from 70,000 (for 150 µm diameter) to 1,500,000 (for 55 µm diameter). Replicates are shown in Fig S3. (d) Growth model parameters extracted from the growth curve measurements. Growth capacity is the fold increase of cellular density after cultivation in the microdroplets, using fluorescence as a proxy. Statistical significance is indicated by a p-value of less than 0.05 (*), less than 0.01 (**), and less than 0.001 (***), determined through a two-tailed student's t-test.

Kinetic data of growth in droplets were obtained through real-time fluorescence measurements of droplet populations in wells of microplates. Fig 1c illustrates the change of fluorescence averaging across three or four wells, each of which contained approximately 850,000, 140,000, and 42,000 droplets, for 55, 100, and 150 μm diameter droplets, respectively. The full set of growth curves, with well replicates, are in Figure A.3. From these monoculture growth curves, we extracted three parameters, the growth capacity, maximum specific growth rate, and lag time. While we cannot approximate the carrying capacity, i.e. the exact cell density at saturation, from the fluorescence measurements alone, the fold increase of fluorescence is a metric that quantifies the degree of growth possible in a microdroplet as a function of its size. This metric is referred to as the “growth capacity” of the droplet in the rest of the study. In monoculture, the maximum specific growth rate is extracted from the growth curve and is informative in determining whether or not growth is facilitated or hampered. Finally, the lag time is defined as being the time between the inoculation and the start of detectable exponential growth. While the lag time has a physiological basis, the molecular mechanisms for the duration of a lag time is not always obvious. Nevertheless, the lag time is a very important growth parameter dependent on the stress or unfavourability of a condition to cell growth (109). We extracted these three parameters from the growth curves (Figure 2.1d). As expected, the 150 μm diameter droplets were capable of supporting the most growth (598.9 ± 56.0 sd fold increase), followed by the 100 μm (269.7 ± 27.5 sd), and then the 55 μm diameter droplets (73.0 ± 3.4 sd). Interestingly, there were statistically significant differences between the

maximum specific growth rate of the monoculture in different sized droplets. As droplet diameter increased, the maximum specific growth rate decreased from 0.79 ± 0.01 (sd) to 0.69 ± 0.01 (sd) to 0.63 ± 0.01 (sd) hr^{-1} as the diameter increased from 55 μm to 100 μm to 150 μm, respectively. The lag time also lengthened as the droplet size increased, from 1.3 ± 0.01 (sd) to 3.1 ± 0.04 (sd) hr. While unintuitive at first, population-level interactions among bacterial cells mediated by diffusible small molecules, such as competition for limited nutrients or cell-to-cell communication, have been documented extensively (110,111) and are likely underlying these findings.

To gain further insight on how the droplet size impacts the cell growth dynamics, we carried out a separate experiment where S1 Δ*ilvD* was inoculated at the same initial cell density across the different droplet sizes, as opposed to the same λ value (i.e. the average number of cells per droplet). It was observed that when the initial cell density was fixed across droplet sizes, the growth capacity, maximum specific growth rate, and lag time remained largely unchanged (Figure A.4). The only exception was a slightly higher growth capacity in the 55 μm diameter droplets than that in the 100 μm (with a modestly statistically significant p-value of 0.045). These findings suggest that the effect of droplet size on cell growth is likely facilitated through changing the average cell-to-cell distance in the droplet microenvironment, which could influence various factors affecting cell growth such as quorum sensing or diffusion of exchanged molecules.

Another interesting observation was the droplet-to-droplet variability observed in Figure 2.1a. Despite containing cells of the same genotype and the same medium, some droplets had extensive proliferation, some had moderate growth, while some had virtually none. We hypothesized that the variation was due to the cell-to-cell

stochasticity that amplified under the small-number condition of λ=5 cells/droplet, and if the initial number of cells was higher, this variability would decrease. To test this hypothesis, we conducted a monoculture droplet cultivation of S1 Δ*ilvD* at λ=5 and λ=20 cells/droplets to compare the variability between the two droplet populations. We observed that when λ was increased from 5 to 20, the standard deviation of the distribution of the droplets' normalized fluorescence was decreased by 48% in droplets of 100 µm diameter (Figure 2.2) and by 32% in both 55 µm and 150 µm droplets (Figure A.5). Interestingly, not only were the proportion of droplets without growth reduced, but also those with higher-than-average fluorescence. These results suggest that cell-to-cell stochasticity in relevant properties, particularly viability, growth and fluorescence, is indeed a major contributor to droplet-to-droplet variation.

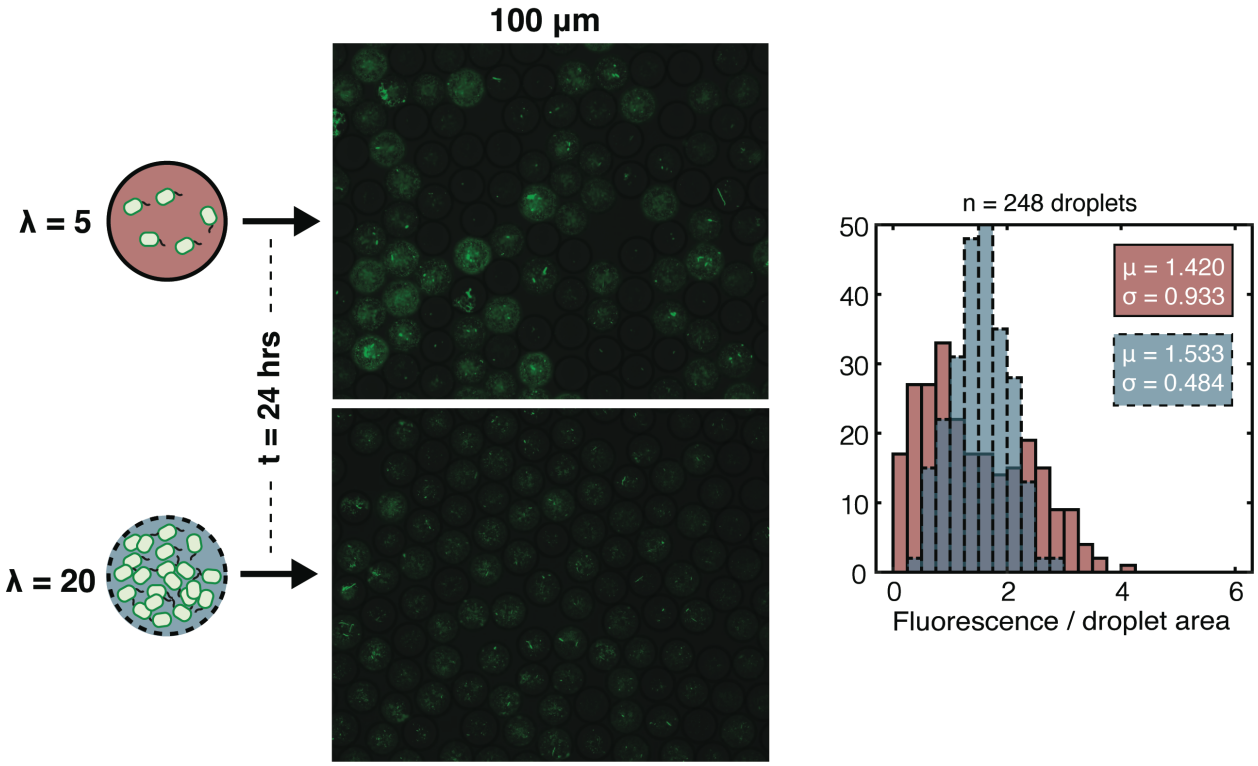


Figure 2.2: Droplet-to-droplet variation of fluorescence for cultivation of S1 ΔilvD with λ=5 and λ=20 cells/droplet in droplets in droplets of 100 μm diameter. Droplet-to-droplet variation is illustrated in representative images of populations of droplets after cultivation for 24 hours under both conditions. Image analysis of a large population of droplets (248 droplets under each condition) was performed to quantify the degree of droplet-to-droplet fluorescence variation through histograms with associated statistics (mean and standard deviation). The distribution and statistics of the population under the λ=5 initial condition is in magenta with a solid boundary; while those under the λ=20 condition in blue with a dashed boundary.

2.4.2 A model system of cross-feeding amino acid auxotrophs

To study co-cultures in droplets, we made use of a model system consisting of two cross-feeding *E. coli* strains previously developed by our group (100). The two strains are: S1 Δ*ilvD* and S2 Δ*lysA*, called S1 and S2 due to their roles as complementary secretors. S1 Δ*ilvD* requires extracellular valine, leucine, and isoleucine due to its knockout of *ilvD*, which encodes a dihydroxy-acid dehydratase responsible for converting 2,3-dihydroxy-isvalerate into 2-ketoisovalerate, a precursor of valine and leucine, as well as 2,3-dihydroxy-3-methylvalerate to 2-keto-3-methylvalerate, a precursor of isoleucine. Similarly, S2 Δ*lysA* is auxotrophic for lysine and requires extracellular lysine. In this study, three amino acid supplementation conditions were utilized to represent conditions of low, medium, and high degrees of interaction between the two auxotrophs: the addition of isoleucine; valine and leucine; and no addition, respectively. Because of differences in related biosynthetic pathways for these amino acids, isoleucine is the more biosynthetically-expensive amino acid for a cell to produce, with lysine, leucine, and valine following in order (55). In bulk cultivation (e.g. in 200 μL in microplate wells), the co-culture grows much faster when supplemented with isoleucine than when supplied with valine and leucine, and the slowest condition is when no amino acid is supplemented (Figure A.6). With different extents of interaction

in the microdroplet, we expected the interaction dynamics to be unequally affected by changes in droplet size. Under all these conditions, a λ value of 5 cells/droplet for each strain was chosen to ensure that almost all droplets would contain at least one cell of each strain.

2.4.3 Lowest degree of interaction

We started with the co-culture condition of supplementing isoleucine in the medium, removing the most biosynthetically-expensive metabolite from syntrophic exchange (Figure 2.3a). We observed under this condition of low degree of interaction that the composition of bi-cultures in droplets was dominated by S1 Δ*ilvD*, which relied less on S2 Δ*lysA* for syntrophic exchange than S2 Δ*lysA* relied on it, as shown by the heavily green fluorescent droplets and significantly higher fold increase of fluorescence from S1 Δ*ilvD* over that from S2 Δ*lysA* (Figure 2.3b,c). Similar to the monoculture cultivation, droplet-to-droplet variation is noticeable, most likely arising from the low number of initial cells of each strain in each droplet, introducing significant variations (Figure 2.3b).

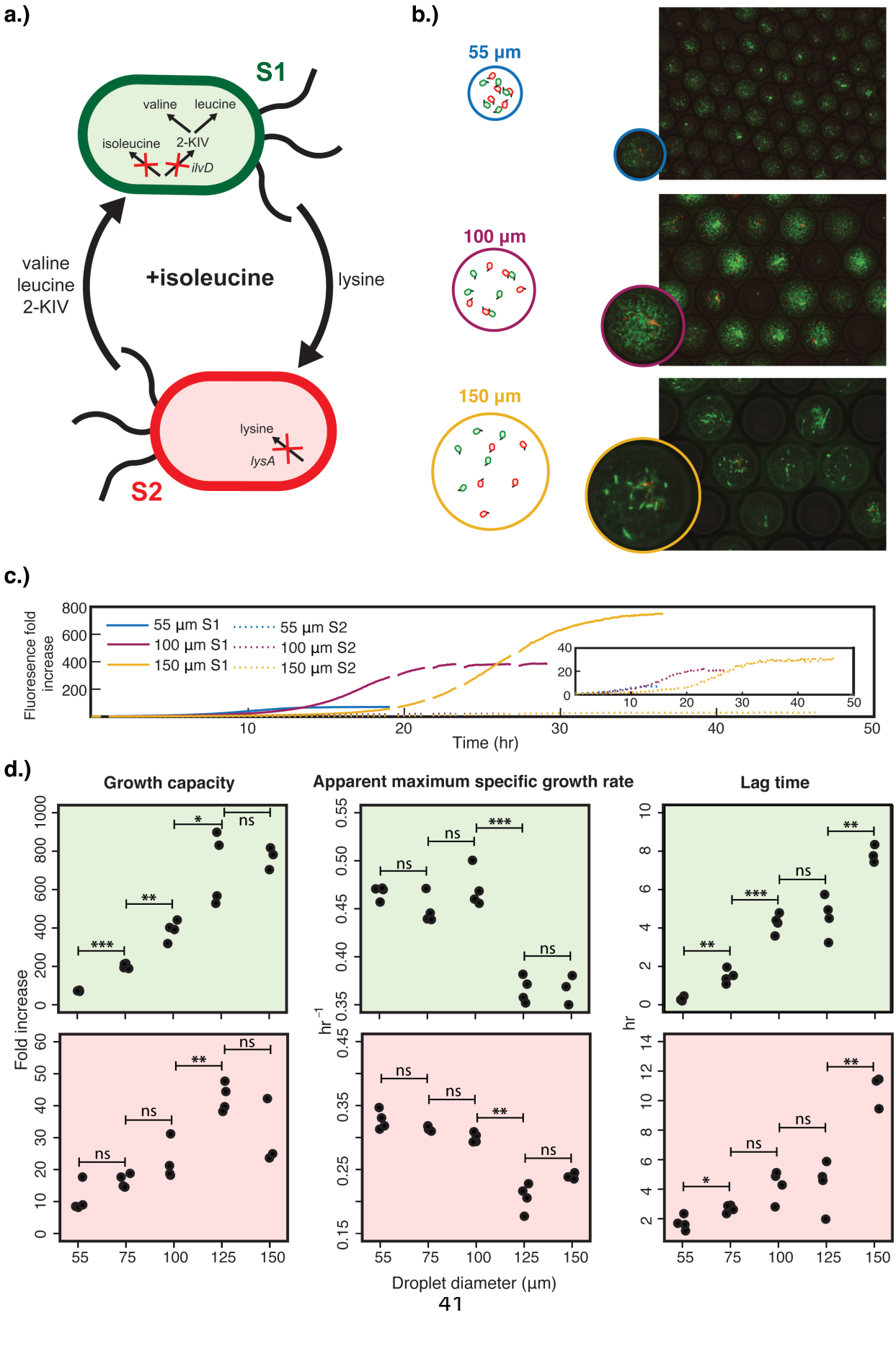

a.)
S1
valine
leucine
isoleucine
2-KIV
ilvD
valine
leucine
2-KIV
+isoleucine
lysine
lysine
lysA
S2
b.)
55 µm
100 µm
150 µm
c.)
Fluorescence fold increase
55 µm S1
100 µm S1
150 µm S1
55 µm S2
100 µm S2
150 µm S2
Time (hr)
d.)
Growth capacity
Apparent maximum specific growth rate
Lag time
Fold increase
hr⁻¹
hr
Droplet diameter (µm)

Figure 2.3: Co-growth of S1 ΔilvD and S2 ΔlysA under the low degree of interaction (with supplementation of 3 mM isoleucine, the most biosynthetically-expensive cross-fed amino acid) in a range of droplet sizes with λ=5 cells/droplet. (a) The cross-feeding between S1 ΔilvD and S2 ΔlysA under isoleucine supplementation. (b) Fluorescence microscopy and phase contrast overlays of droplets of 55, 100, and 150 μm after co-cultivation, with representative images illustrating the bi-culture densities. S1 ΔilvD has a constitutively-expressed mNeonGreen (green) fluorescence reporter and S2 ΔlysA has a constitutively-expressed mCherry (red) fluorescence reporter on a plasmid. (c) Average growth curves of the bi-culture. For each droplet size, there are two growth curves, for S1 ΔilvD (bold line) and S2 ΔlysA (dotted line), respectively. Each curve represents the average growth of the strain in co-culture across multiple wells in a microtiter plate, each of which contains thousands of droplets. Full replicates are shown in Fig S7. (d) Logistic equation parameters extracted from the growth curves for both S1 ΔilvD (light green) and S2 ΔlysA (light red). Growth capacity is the fold increase of cell density after cultivation in the droplets, using fluorescence as a proxy. Apparent maximum growth rate is the fitted value for maximum growth rate from the growth curve, acknowledging that the growth dynamics of a cross-feeding co-culture is not exactly a standard logistical growth curve (Note S1 and Fig S2). Statistical significance is defined by a p-value of higher than 0.05 (ns), less than 0.05 (*), less than 0.01 (**), and less than 0.001 (***), determined through a two-tailed student's t-test.

In terms of the effects of droplet size on growth under this condition, they were quite noticeable and reflected in the growth dynamics of both S1 Δ*ilvD* and S2 Δ*lysA*. Both strains demonstrated higher growth capacity when in larger droplets up until 150 μm, with the trends of S1 Δ*ilvD* being much more statistically significant than those of S2 Δ*lysA* (Figure 2.3d). The most interesting trend was in the apparent maximum specific growth rate (Fig 3d). In the range of droplet diameters from 55 to 100 μm, both strains' apparent maximum specific growth rates remained largely constant. In sharp contrast, when the droplet diameter further increased to the range of 125 - 150 μm, both strains showed reduced maximum specific growth rates, about 20% slower than those in the smaller droplets. It was also interesting that we did not observe a gradual decrease of the maximum specific growth rate while the droplet size increased as previously seen in the monoculture baseline experiment (Figure 2.1d). Lag time increased as droplet size increased (Figure 2.3d), similar to the trend observed in the monoculture. However, the range of lag times observed in the bi-culture with isoleucine supplementation (up to 12 hr) was much wider than observed in the monoculture (up to

3 hr). We also examined whether the final community composition changed when the droplet size increased while all other factors (amino acid supplementation and λ value) were kept constant. Looking into the ratio between fold increase of S1 Δ*ilvD* and S2 Δ*lysA* with isoleucine supplementation, we noted that increasing the droplet size shifted the bi-culture towards one more dominated by S1 Δ*ilvD* (Figure A.8).

As an attempt to dissect mechanisms underlying the above observations, we co-cultivated S1 Δ*ilvD* and S2 Δ*lysA* from the same initial cell density across different droplet sizes, as opposed to the same λ value, which was similar to efforts described earlier for monocultures. Specifically, the average initial cell number (i.e. λ value of the Poisson distribution) for each strain was 5, 30, and 100 cells/droplet in the 55, 100, and 150 μm diameter droplets, respectively. Under this condition, we found that the previously observed effects of droplet size on the bi-culture growth when λ was kept constant vanished. In fact, the trend was even reversed to some extent. Specifically, the growth capacity decreased as the droplet size increased, whereas the apparent maximum specific growth rate increased when the droplet size was increased from 55 to 100 μm in diameter (Figure A.9). These findings are in partial agreement with observations in the monoculture experiments, suggesting again that the cell-to-cell distance in the droplet microenvironment is a key factor determining growth dynamics. Furthermore, the impact of this factor is even more profound when cells need to interact through exchange of primary metabolites essential for growth, like in our model cross-feeding bi-culture.

2.4.4 Moderate degree of interaction

We next investigated the growth dynamics of the cross-feeding bi-culture under the condition of moderate degree of interaction, by supplementing leucine and valine in the medium. This requires the more biosynthetically-expensive amino acid, isoleucine, to be cross-fed in exchange for lysine (Figure 2.4a). In comparison to the previous experiment with isoleucine supplementation, this condition increased the dependence of S1 ΔilvD on S2 ΔlysA, which subsequently reduced the ratio of S1 ΔilvD over S2 ΔlysA and led to bi-culture compositions that were more balanced between the two strains (Figure 2.4b,c). Interestingly, the morphology of *E. coli* cells was notably filamentous under this condition, signifying cellular stress as previously observed under other stressful conditions (112).

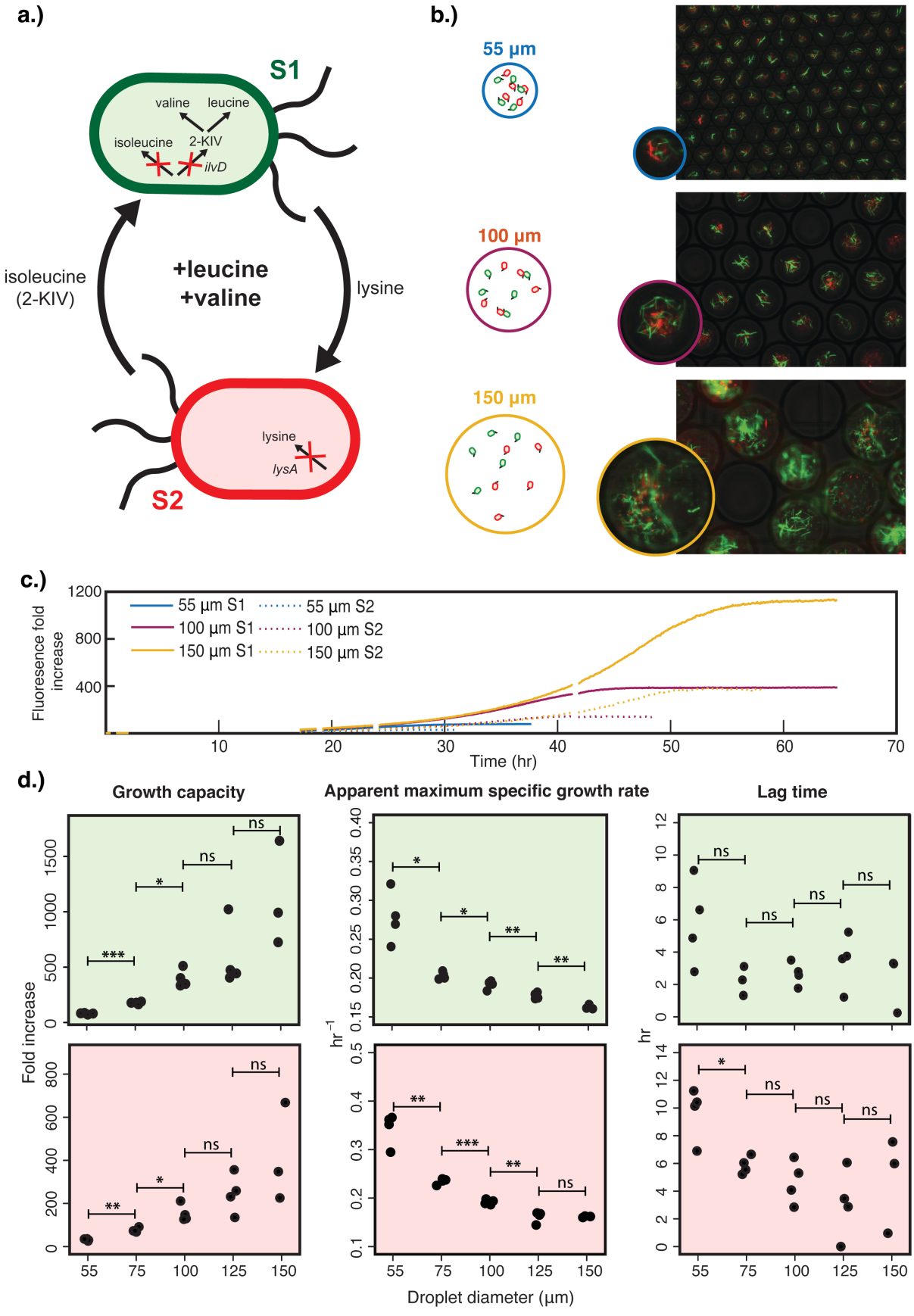

a.)
S1
valine
leucine
isoleucine
2-KIV
ilvD
isoleucine
(2-KIV)
+leucine
+valine
lysine
lysine
lysA
S2
b.)
55 µm
100 µm
150 µm
c.)
Fluorescence fold increase
1200
800
400
55 µm S1
100 µm S1
150 µm S1
55 µm S2
100 µm S2
150 µm S2
10
20
30
40
50
60
70
Time (hr)
d.)
Growth capacity
Apparent maximum specific growth rate
Lag time
Fold increase
hr⁻¹
hr
ns
*
**
***
55
75
100
125
150
Droplet diameter (µm)

Figure 2.4: Co-growth of S1 ΔilvD and S2 ΔlysA under the intermediate degree of interaction (with supplementation of 3 mM leucine and 3 mM valine, which are intermediately biosynthetically-expensive amino acids to produce) in a range of droplet sizes with λ=5 cells/droplet. (a) The cross-feeding between S1 ΔilvD and S2 ΔlysA under leucine and valine supplementation. (b) Fluorescence microscopy and phase contrast overlays of droplets of 55, 100, and 150 μm after co-cultivation, with representative images illustrating the bi-culture densities. (c) Average growth curves of the bi-culture. For each droplet size, there are two growth curves, for S1 ΔilvD (bold line) and S2 ΔlysA (dotted line), respectively. Each curve represents the average growth of the strain in co-culture across multiple wells in a microtiter plate, each of which contains thousands of droplets. Full replicates are shown in Fig S7. (d) Logistic equation parameters extracted from the growth curves for both S1 ΔilvD (light green) and S2 ΔlysA (light red). Statistical significance is defined by a p-value of higher than 0.05 (ns), less than 0.05 (*), less than 0.01 (**), and less than 0.001 (***), determined through a two-tailed student's t-test.

The three sets of parameters obtained from fitting the logistic equation to the growth data were shown in Figure 2.4d. Growth capacity followed the typical trend exhibited in previous experiments, with less of an increase in 125 and 150 μm diameter droplets. The apparent maximum specific growth rates under this condition, however, exhibited different features. First, as expected, these rates were substantially lower than those in the previous experiment with isoleucine supplementation (Figure 2.3d). Intriguingly, in contrast to the abrupt decrease of growth rates from 100 to 125 μm diameter droplets under the isoleucine-supplemented condition, here we observed a gradual decrease of the growth rates in the full range of droplet sizes, with the most significant reduction of approximately 23% occurring in the transition from 55 to 75 μm droplets and a 40% reduction from 55 to 150 μm diameter. Finally, trends of lag time across increasing droplet size were not as evident or statistically strong as in the previous experiment with isoleucine supplementation, particularly with S2 ΔlysA. We also examined the effect of droplet size on the final community composition and noted that increasing the droplet size had no significant effect on the community composition, a trend unlike in the previous experiment (Figure A.8). This may be due to the higher

degree of interdependence between the two strains, which could dominantly govern their ratio and obscure the effect of other factors including the droplet size.

2.4.5 Highest degree of interaction

Lastly, we investigated the bi-culture growth dynamics under a third condition of the most strenuous interaction, in which no amino acids were supplied externally (Figure 2.5a). It was observed that the bi-culture only showed the ability to co-grow reliably in small droplets of diameter 55 μm, whereas in larger droplets (diameters 100 and 150 μm) there was a stark divergence between droplets with and without significant co-growth (Figure 2.5b). This high variation was not only visible in images of the droplets, but also reflected in the widened distribution of size-normalized fluorescence (Figure 2.5c). More specifically, the variance was substantially higher for the larger droplets of diameters 100 and 150 μm; and the distribution shape was significantly distinct from that of the 55 μm droplets (student two tailed t-test, p-values < 0.005). In sharp contrast to the previous two experiments with amino acid supplementation, we were not able to obtain high-quality growth curves from droplet pools in wells of microtiter plates, due to the large subpopulation of droplets that failed to establish bi-culture growth.

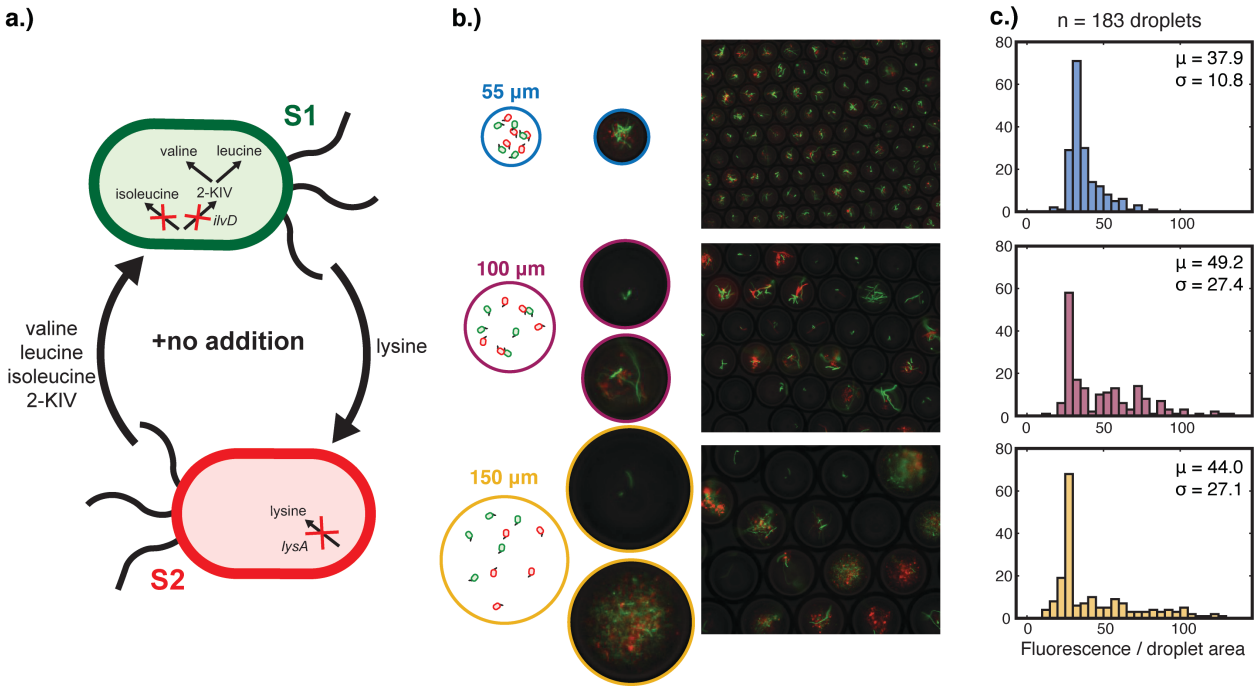


Figure 2.5: Co-growth of S1 ΔilvD and S2 ΔlysA under the high degree of interaction (no supplemented amino acids) in a range of droplet sizes with λ=5 cells/droplet. (a) The cross-feeding between S1 ΔilvD and S2 ΔlysA under no amino acid supplementation. (b) Fluorescence microscopy and phase contrast overlays of microdroplets of 55, 100, and 150 µm after co-cultivation, with representative images illustrating the bi-culture densities. (c) Histograms of the post-cultivation total fluorescence normalized by droplet area (a.u./pixel) for a total of 183 droplets analyzed per droplet size, with associated mean and standard deviation.

2.5 Discussion

2.5.1 Interaction response in growth to changes in droplet size

In this study, we observed that the effect of droplet size on cross-feeding growth dynamics was highly dependent on the degree of the interaction. Under all conditions, the growth capacity generally increased as the droplet size increased, as expected, due to larger droplets giving rise to lower initial cell densities and hence allowing more doublings in cell growth before reaching saturation. In terms of the effect on growth rates and lag times, we noted that there were two very different types of trends. In one

type (Type 1), there is no significant change in the growth rate whereas the lag time is lengthened as the droplet size increases, as observed under the isoleucine supplemented condition when the droplet diameter was increased from 55 to 100 μm (Figure 2.3). In the other type of trends (Type 2), there is a continual gradual decrease in the growth rate with no significant change in the lag time, as observed under the valine and leucine supplemented condition (Figure 2.4).

The mechanisms underlying these distinct growth profiles are not obvious. It is possible that diffusion plays some role as droplet size increases, but the order-of-magnitude calculations suggest that while diffusion may take longer in larger droplets, the time scale required for diffusion in the largest droplet (estimated to be about a few minutes) is still very short compared to the time scale associated with cell growth (i.e. hours). Another possibility may be that the variety of growth dynamics arise from the inherent complexity of this ecosystem consisting of interacting subpopulations. To explore this possibility, we utilized the ODE model detailed in Note A.1 and carried out simulations with a range of parameter values related to the secretion and growth requirement of cross-fed metabolites. We found that these parameters could have a profound impact on the growth dynamics of the simulated bi-culture. Furthermore, the two types of trends noted above regarding how the droplet size affects the bi-culture growth dynamics could be recapitulated *in silico* to a large extent (Figure A.10). Specifically, it can be shown that the lag time increases with insignificant change of the maximum specific growth rate as the initial cell density decreases, characteristic of the experimentally observed Type 1 trend, under the condition that cellular secretion of the cross-fed molecules is of low level (i.e. value of the α parameter in the ODE model is

relatively small). In contrast, the lag time is largely not affected by the initial cell density under the condition that cellular requirement for the cross-fed molecule is high (i.e. value of the β parameter in the ODE model is relatively large), replicating in part the experimentally observed Type 2 trend. In light of these findings from ODE modeling which assumes diffusion is instantaneous, we speculate that inherent properties related to intercellular interactions in complex ecosystems, besides molecular diffusion, may play an important role in shaping the growth dynamics of co-cultures in microdroplets. We do recognize that our ODE model could not explain all the experimental observations, such as the step-like decrease of the maximum specific growth rate when droplet diameter was increased from 100 to 125 μm in Fig 3. As another example, even with the same initial cell density, we observed noticeable changes in the growth capacity and maximum specific growth rate when changing the droplet size, especially for the bi-culture (Figure A.9). The mechanism underlying these different growth dynamics is not clear. Future investigations, both experimental and computational ones, will be needed to elucidate the full spectrum of mechanisms driving co-culture dynamics in droplets. One promising direction would be to develop a more advanced mathematical model to incorporate additional potential mechanisms, e.g. transport of key molecules exchanged between cells, through diffusion and/or other means. In particular, recent studies have revealed that certain bacteria have evolved direct cell-to-cell contact for transporting metabolites more efficiently (113,114). It would be an interesting topic for future research to investigate how droplet size affects co-culture growth dynamics with this alternative mechanism for cross-feeding.

2.5.2 Droplet-to-droplet variation

As demonstrated in this work and previous literature, substantial droplet-to-droplet variation is an inherent feature of microbial cultivation in microfluidic droplets. One of the major advantages of droplet based co-cultivation is the ability to study interactions between a small number of cells. However, the inclusion of only a small number of cells also introduces stochasticity from cell-to-cell variation and encapsulation statistics. We observed that despite a λ of 5 cells/droplet, a large portion of droplets did not show growth, potentially due to a significant portion of non-viable cells. Increasing the λ value reduced the number of these no-growth droplets, but also reduced the number of droplets exhibiting higher-than-average fluorescence, suggesting that at sufficiently large λ, cell-to-cell differences average out and reduce droplet-to-droplet stochasticity. One source of significant cell-to-cell variations in our experimental system may be caused by plasmid variability and stability. In particular, the fluorescent proteins in both strains and three enzymes for enhanced-production of cross-fed molecules in S2 Δ*lysA* are encoded by genes carried on plasmids, which can introduce a large degree of cell-to-cell variability (115,116). As such, at low λ, droplets can be exploited for investigating single-cell level differences, but one should not expect the exact dynamics of a community in one droplet to be exactly reproduced in another. As evidenced by other microfluidic studies (117,118), even in a single population, these dynamics are quite stochastic due to cell-specific quorum sensing capability and expression (117). This stochasticity must be considered appropriately in studies which cultivate populations or communities from low cell numbers. For instance, Hsu et al. (99) addressed this issue by determining the strength of interactions in a three-member system with statistical inference across a large number of droplets.

The precise control of the λ value in this work allowed us to study cell populations at low cell numbers, with one particular observation being the effect of low initial cell numbers on the lag time. Our results are consistent with previous studies that used empirical data from single cells of *E. coli* K12 and showed that at low cell number inocula (1 to 100 cells) in a fixed volume, the lag time would increase as the inoculum size was lowered (119). *E. coli* K12, the base strain of the auxotrophs utilized in this study, has been demonstrated to be affected by quorum sensing. In particular, autoinducer-2 (120,121) could be one of many signaling molecules involved.

2.5.3 Implications for utilizing co-cultivation in droplets to infer interactions

Inappropriate selection of the droplet size for the specific system of study or objective can result in failure to capture the intricacy of ecological interactions. Practically, we acknowledge that the objectives of microfluidic co-cultivation studies are diverse, and how thorough the consideration of droplet size should be will depend on the questions being investigated. For example, microdroplets are being utilized in ultra-high throughput screening (100,122,123). In these scenarios, the objective is usually to identify and retrieve droplets in the top percentile of a vast droplet pool analyzed by single droplet measurements such as fluorescence or optical density. While a larger droplet may allow for a larger dynamic range for screening due to the higher growth capacity, too large a droplet may not be able to reliably render intercellular interactions, as we observed with our model bi-culture system under the no-supplementation condition. Another popular application is the investigation of individual communities encapsulated in single droplets. Due to the difficulty in tracking single droplets, most droplet-based investigations have not focused on the growth of the same individual

droplets over time. For instance, in this study, the growth dynamics in a large population of droplets was studied through the use of averaged characteristics. We expect, however, as droplet technologies continue to advance, future work studying microbial community dynamics would shift to the finer single-droplet resolution and it would be crucial to take into full account the effect of droplet size as well as the inherent stochasticity arising from cell-to-cell variations and random encapsulation, while designing specific experiments.

Chapter 3 : High-resolution Metagenomic Dissection on Selected Microdroplets to Study Microbial "Dark Matter"

This chapter is modified from a previously published work (124).

3.1 Summary

While the "unculturable" majority of the bacterial world is accessible with culture-independent tools, the inability to study these bacteria using culture-dependent approaches has severely limited our understanding of their ecological roles and interactions. To circumvent cultivation barriers, we utilize microfluidic droplets as localized, nanoliter-size bioreactors to co-cultivate subsets of microbial communities. This co-localization can support ecological interactions between a reduced number of encapsulated cells. We demonstrated the utility of this approach in the encapsulation and co-cultivation of droplet sub-communities from a fecal sample collected from a healthy human subject. With the whole genome amplification and metagenomic shotgun sequencing of co-cultivated sub-communities from 22 droplets, we observed that this approach provides accessibility to uncharacterized gut commensals for study. The recovery of metagenome-assembled genomes from one droplet sub-community demonstrated the capability to dissect the sub-communities with high genomic resolution. In particular, genomic characterization of one novel member of the family Neisseriaceae revealed implications regarding its participation in fatty acid degradation and production of atherogenic intermediates in the human gut. The demonstrated

genomic resolution and accessibility to the microbial “dark matter” with this methodology can be applied to study the interactions of rare or previously uncultivated members of microbial communities.

3.2 Introduction

A combination of top-down culture-independent approaches and bottom-up culture dependent approaches holds promise for addressing the challenges in microbial ecology (2,23,125–127). However, there are numerous barriers to traditional cultivation (128), including microbial lifestyles such as obligate syntrophy (129) and slow-growth (130), competition, and extremely low abundances in diverse natural communities. Microfluidic droplet technology presents a unique solution to circumvent many of these barriers by manipulating microbial communities at the nanoliter scale. While used extensively in the past for the high-throughput encapsulation and manipulation of single cells (131,132), microfluidic droplets can also be utilized as micron-scale bioreactors to encapsulate and propagate subsets of natural microbial communities, decomposing a bulk community into a large number of much smaller sub-communities. This compartmentalization limits complexity while allowing the study of microbe-microbe interactions. While microfluidic droplets have been employed for single-cell or “mini-metagenomic” investigation of the microbial dark matter (133–137) or for synthetic systems (94,101), these studies did not incorporate co-cultivation for the investigation of microbial communities. Microfluidic co-cultivation of microorganisms provides a bridge between culture-dependent and -independent tools, allowing for the exploration of ecological interactions between cells through co-growth at microscopic scales. This circumvents technical difficulties of culture-dependent work while enabling use of high-

resolution ‘omics tools in studies of interacting microbial consortia comprised largely of uncultured species.

To demonstrate the potential of utilizing microfluidic droplets to study the “dark matter” of natural communities, this study uses microfluidic droplets to dissect a complex human fecal sample into subcommunities for highly parallel co-cultivation. Afterwards, 22 individual droplets with strong bacterial co-growth were selected with microfluidic techniques. Multiple displacement amplification (MDA), a whole genome amplification method, amplified nucleic acid from individual droplet sub-communities for 16S rRNA amplicon sequencing, and one sub-community was metagenomically shotgun sequenced. Of particular interest relating to the human gut microbiome, a partial-genome of a representative of a novel genus within the Neisseriaceae was found in this droplet, highlighting the capability of microfluidic co-cultivation to access and study uncharacterized microbial diversity.

3.3 Materials and Methods

3.3.1 Microfluidic devices and fabrication

The work here utilizes two microfluidic devices: a droplet generation device and a droplet spacing device. The droplet generation device is a modified cross-flow droplet generation device made from polydimethylsiloxane (PDMS) as described previously by Carruthers *et al.* (103). The droplet spacing device used to separate individual droplets is a modified droplet generation device with three layers to construct a pressure-controlled membrane valve to precisely manipulate droplet flow through a cross-flow junction (Figure B.1) (138,139). The valve is controlled manually by an external pressure pump to switch on and off the flow of individual droplets. When pressure is on,

the membrane valve closes and stops the flow of droplets while allowing the oil phase flow to continue.

To fabricate molds with the microfluidic device features, photo-mask creation and SU-8 mold etching were performed in the Lurie Nanofabrication Facility at the University of Michigan. SU-8 molds were made by negative etching on a silicon wafer, which was spin-coated with SU-8 2035 at a thickness of 50 μm. The wafer was pre-baked at 65°C and at 95°C, exposed, and post-exposure baked at 95°C. After baking, the wafer was silanized with tridecafluoro-1,1,2,2,-tetrahydrooctyl-1-trichlorosilane in a desiccator. For the PDMS layers with device features, PDMS with Sylgard® 184 curing agent (10:1 mass ratio of PDMS to curing agent) was poured on top of respective SU-8 molds, vacuumed to remove air bubbles, and heated at 70°C to solidify the polymer overnight. The devices were then cut to size. The membrane between the valve layer and the channel layers for the spacing device was made from PDMS (15:1 mass ratio of PDMS monomer to curing agent) spun on a glass wafer at 1000 rpm to create a 50 μm membrane. The membrane was heated to 80°C for 15 min and plasma-activated bonded to the other layers of the spacing device using a corona discharge wand. To complete the fabrication, the devices were punched with holes by a biopsy punch (1.25 mm ID) to create openings for channel flow, cleaned with acetone and rinsed with water, and bonded on cleaned glass microscope slides via plasma-activated bonding.

3.3.2 Droplet generation, cultivation, and processing

All droplet work was performed in an anaerobic chamber (Coy© vinyl anaerobic chamber) with an atmosphere of 5% hydrogen, 10% carbon dioxide, and balance nitrogen to protect and provide favorable conditions for obligate anaerobes in our

sample. To remove any trace oxygen, a fan box to recirculate air through a Coy© palladium catalyst Stak-Pak is employed in the chamber. The cellular suspension for flow through the droplet generation microfluidic device was derived from a fecal sample obtained from a healthy human subject at the University of Michigan Hospital in October 2011. The fecal sample was stored at -80°C in a 2 mL cryovial and thawed anaerobically. 1.5 mL of phosphate buffer saline (PBS) was added to the fecal sample, and the mixture was centrifuged at 1000 rpm to separate fecal debris from the bacterial suspension for microfluidic droplet generation. The cell density of the bacterial suspension was determined with a hemocytometer (C-Chip™ disposable hemacytometer, Fisher Scientific 22-600-100) under a Nikon inverted contrast phase microscope (Nikon Eclipse Ti-S) and was determined to be $1.9x10^8$ cells/mL. The suspension was diluted to the appropriate concentration determined by the desired average cell number per droplet (λ) divided by the volume of droplet. To provide different droplet sub-community complexities, two λ values were used for droplet generation: 2 and 10. Control droplets without cells from the fecal sample were also generated to demonstrate no bacterial contamination from reagents was present (Figure B.2). To dilute the volume and provide the nutrients for growth, two different media for the cultivation of anaerobic intestinal bacteria were used: Brain Heart Infusion (BHI) (140) (Becton-Dickinson) and the Schaedler Medium (SM) (141,142) (Oxoid).

Droplet generation was performed in the anaerobic chamber. The droplet generation device was placed on a lab compound microscope (Amscope, M150C) with a USB connected camera. Syringes were connected to tubing (Cole-Parmer, EW-0641721) leading to the droplet generation device and placed on a syringe pump (CMA

102 Syringe Pump). Oil phase – fluorocarbon oil (HFE-7500, 3M) containing 2% perfluoropolyether-polyethyleneglycol surfactant (RAN Biotechnologies) – was placed into one syringe and the bacterial suspension in the other. The syringe pumps adjusted oil and aqueous flows to achieve generation on-chip. Droplets generated flowed outwards from the droplet generation device into Eppendorf tubes attached with tubing. Approximately 500 µL of droplets were collected for each sample and covered with sterile mineral oil to prevent evaporation. The droplets were incubated at 37°C in the anaerobic chamber for one week, which provided sufficient time for co-growth to occur. For imaging of droplets, 10 µL of droplets were inserted into a C-Chip hemocytometer before and after incubation and visualized under the inverted phase contrast microscope.

3.3.3 Spacing of high cell density droplets

Incubated droplets were stored and transferred in 1.5 mL micro-centrifuge tubes using two ports installed in the tube lid. In one port, Teflon tubing was inserted to the bottom of the tube, and the other port had a syringe needle. Droplets were transferred from the bottom of the tube to the spacing device by applying pressure through the syringe tip. Droplets were introduced into the middle junction of the device and held at the junction by flowing 1.5% surfactant in HFE oil while the membrane valve was closed. After the device reached steady state, Teflon tubing was inserted into the outlet channel. Spacing of individual droplets was performed through manual control of a pressure pump connected by tubing to the valve channel. After the spacing of a droplet, the membrane valve was shut for 1 minute to ensure the isolated droplet had traveled through the entirety of the tubing and was collected into a 1.5 mL microcentrifuge tube.

While in the spacing device, droplets were individually visualized with microscopy for identification of droplets with clear bacterial co-growth (Figure B.3). Through usage of the spacing device, chosen droplets were isolated into separate 1.5 mL microcentrifuge tubes and kept for downstream processing. Droplets not selected were discarded.

3.3.4 Cell lysis and whole genome amplification of isolated droplets

For each isolated droplet, 5 µL of surfactant destabilizer (RainDance Technologies, RDT 1000 droplet destabilizer), an additional 4 µL of PBS, and 3 µL of cell lysis solution (3 µL) was added to each tube. After a 10-minute incubation period at 65°C and the addition of a lysis stopping solution (3 µL), the multiple displacement amplification (MDA) reaction mixture (Single Cell Repli-G kit, Qiagen 150343) for whole genome amplification was added to the aqueous phase, and the reaction was allowed to proceed for eight hours at 30°C. The polymerase was deactivated by heating the solution to 65°C, and amplified DNA was collected from the mixture by pipetting out the aqueous phase and then diluted 100-fold.

3.3.5 16S and metagenomic shotgun sequencing

The V4 region of the 16S rRNA gene was amplified and sequenced from MDA-amplified samples for all 22 droplet samples at the University of Michigan Center for Microbial Systems Sequencing Core using a dual-index PCR library preparation and sequencing strategy as specified by Kozich et al. (143). Information on the library preparation procedure is found at https://github.com/SchlossLab/MiSeq_WetLab_SOP/blob/master/MiSeq_WetLab_SOP_v4.md.

Metagenomic sequencing was conducted on a selected MDA-amplified sample at the University of Michigan Advanced Genomics Core. Sequencing core staff prepared barcoded Illumina libraries of a target fragment size of 450 bp using the IntegenX Apollo 324 PrepX ILM DNA Library kit and custom Illumina compatible BioScientific barcoded adapters. Libraries were amplified with the KAPA library amplification kit, quality controlled on the Advanced Analytical Fragment Analyzer, quantified using the KAPA Illumina library quantification kit, and then sequenced for paired end 150 cycle Illumina HiSeq 4000 sequencing.

3.3.6 16S OTU clustering and analysis

16S V4 reads were processed for quality control, chimera removal, and analysis with mothur (144) according to the MiSeq standard operating procedure (https://www.mothur.org/wiki/MiSeq_SOP). Classification was performed with the Ribosomal Database Project (145) v.16 training set from February 2016 (https://www.mothur.org/wiki/RDP_reference_files). Operational taxonomic units (OTUs) were clustered at the 97% similarity with OTUs less than 78 reads, 1 percent of the lowest sample's sequencing depth, removed. To determine if these droplet OTUs were abundant in the bulk fecal sample, the partial V4 region overlap between previously performed 16S V45 454-pyrosequencing reads (processed through mothur according to https://www.mothur.org/wiki/454_SOP, resulting in 8017 reads from the bulk) and the droplet amplicons was compared. Distances less than 0.05 signified matching of a droplet OTU to an OTU from bulk sequencing.

3.3.7 Analysis of metagenomic shotgun sequencing data

Metagenomic shotgun reads were processed in the University of Michigan's high-performance computing (HPC) cluster Flux. To improve quality of the sequence reads, reads for both microdroplet samples underwent dereplication (script available at https://github.com/Geo-omics/scripts/blob/master/DerepTools/dereplicate.pl), Illumina adapter residual removal by Scythe (v. 0.993) (https://github.com/vsbuffalo/scythe), low quality region removal by Sickle (v. 1.33.6) (https://github.com/ucdavis-bioinformatics/sickle), and interleaving (script available at https://github.com/Geo-omics/scripts/blob/master/AssemblyTools/interleave.pl) to associate forward and reverse reads. FASTQC (v. 0.10.1) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to assess quality of the reads before quality control and after.

To normalize the highly nonuniform coverage of reads generated from stochastic amplification of MDA across the sample and subsequently improve assembly, bbnorm.sh from BBTools (http://jgi.doe.gov/data-and-tools/bbtools/bb-tools-userguide/bbnorm-guide) was used to remove high coverage reads to a target k-mer coverage of 50 and remove reads with a k-mer coverage less than two. Before metagenomic assembly, bbsplit.sh from BBTools was used to remove sequence reads from the human genome by mapping to a reference genome provided by the 1000 Genomes Project (146). This resulted in removal of 36 Gb of reads associated with the human genome and a remainder of around 6 Gb presumably prokaryotic-associated. Metagenomic assembly of these cleaned reads was done with metaSPAdes (v. 3.9.0) (147), using the flags "sc" for MDA amplified DNA and "meta" for metagenomic reads. As suggested for a multi-cell data set with longer Illumina paired reads lengths, the

iterative k-mer lengths used for assembly were: 21, 33, 55, 77, 99, and 127. QUAST (v. 4.3) (148) was used to check assembly quality. Visualization and manual binning of scaffolds was done with anvi'o (v. 2.3.0), following the metagenomic workflow (149). Due to uneven MDA amplification, coverage information was not relevant for binning; therefore, read mapping information and the anvi-merge command were not used. As a result, anvi-profile was run with the "--blank-profile" parameter as specified (http://merenlab.org/2016/06/06/working-with-contigs-only/). Binning was performed using the automatically generated tree based on tetranucleotide frequency profiles and assisted with taxonomic information provided by Centrifuge (150) for contig splits in anvi-interactive visualization. Bin assessment was performed initially with anvi'o, utilizing the marker gene set from Campbell *et al.* (151), and finally with CheckM (152) to determine degree of genome completeness and contamination.

Annotation of genes for genomes of high completeness and low contamination was performed using RAST (153). Pathway inference and subsequent visualization and analysis was performed with the PathoLogic (154) component of Pathway Tools (155) utilizing the MetaCyc database (v.22.0) (156). A custom python script was used to convert the Genbank annotation files for each individual metagenomic scaffold for each genome from RAST into files compatible for PathoLogic genome inference. Incorporation of genomes into the microbial tree of life was done with PhyloPhlAn (157), which uses percent identity of proteins from reconstructed microbial genomes with conserved protein sequences from microbial genomes from the Integrated Microbial Genomes (IMG).

3.3.8 Data availability

All draft genomes recovered from metagenomic assembly are available in GenBank under Bioproject ID PRJNA597463.

3.4 Results

3.4.1 General workflow

We present a microfluidic droplet based technological pipeline that extends previous development in droplet co-cultivation and single-droplet processing (138,139) with metagenomic sequencing (Figure 3.1). A complex environmental sample, in this case a human fecal sample, has been processed to form a microbial suspension that flows through a microfluidic droplet generation device (92) (Figure 3.1a). The microfluidic device utilizes biphasic flows of the aqueous microbial suspension and an oil phase through a T-junction to generate monodispersed, nanoliter-sized microdroplets at a rate of 1 to 500 droplets/second (Figure 3.1b). The distribution of cells encapsulated in individual droplets follows Poisson statistics and can be manipulated through droplet size and initial cell concentration of the microbial suspension. Afterwards, the droplets are incubated under appropriate conditions (Figure 3.1c) and visualized with microscopy. Droplets meeting selection criterion are isolated and transferred to a separate receptacle using a microfluidic droplet spacing device (Figure 3.1d). Lysis reagents and droplet destabilizer are added to the droplets to release and lyse cells, and MDA is used to amplify minute amounts of DNA to a quantity required for sequencing (Figure 3.1e). Based on results from 16S amplicon sequencing, droplets containing sub-communities with interesting taxa can be selected for metagenomic shotgun sequencing (Figure 3.1f) for further study, followed by bioinformatic analysis.

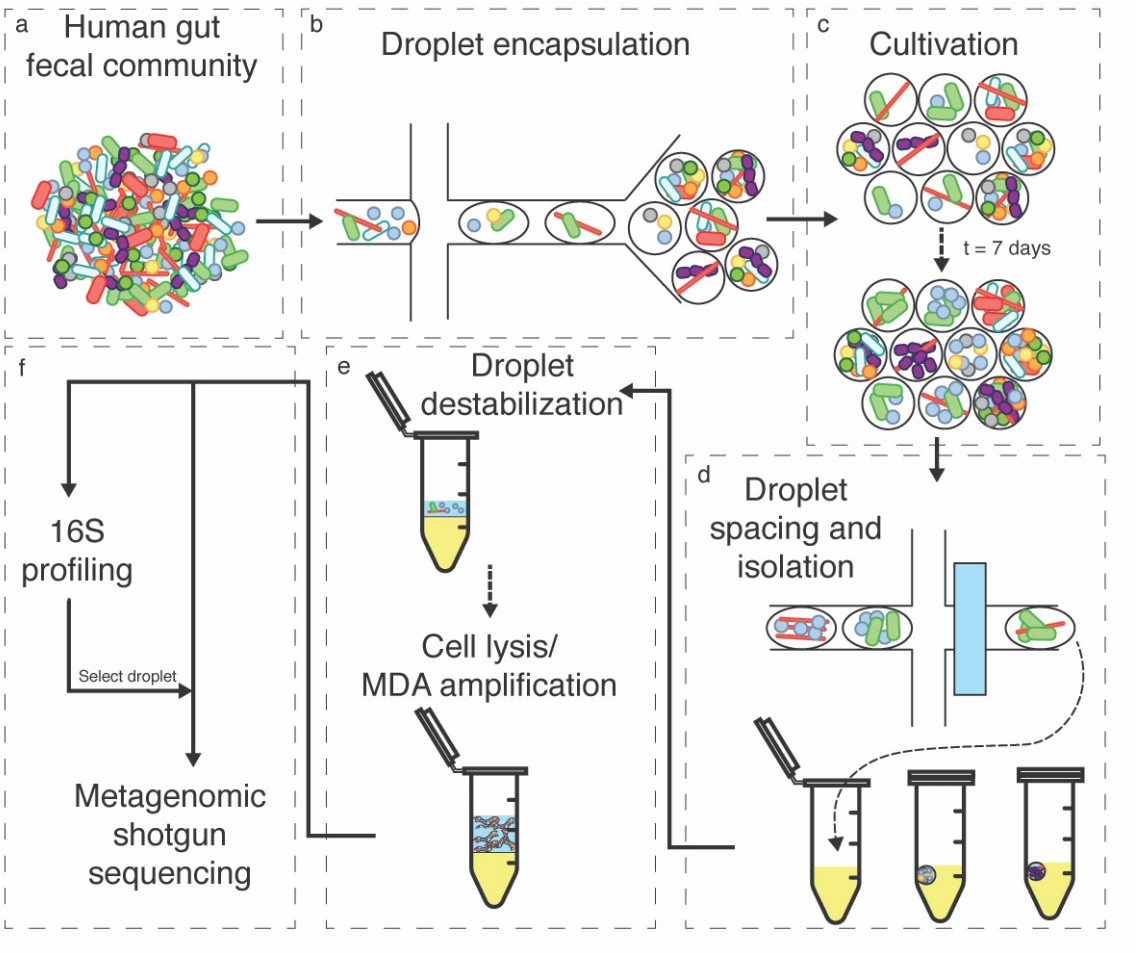


Figure 3.1: Overview of the microfluidic droplet cultivation and processing pipeline. (a) A microbial suspension derived from a human fecal sample is prepared. (b) Using biphasic flows of aqueous and oil phases, random combinations of bacteria are encapsulated in microdroplets at frequencies according to a Poisson distribution. (c) These droplets are incubated anaerobically for a week to allow for co-cultivation of the sub-communities. (d) With a droplet spacing device, microdroplets are isolated and processed individually. (e) Upon droplet destabilization, cells released from individual droplets are lysed and their genomes are amplified with multiple displacement amplification (MDA) to generate sufficient nucleic acid material for downstream sequencing. (f) 16S amplicon and metagenomic libraries are prepared with amplified DNA and sequenced. 16S profiling of individual droplets is used to determine which droplet to submit for metagenomic shotgun sequencing.

3.4.2 Highly parallel co-cultivation of microbial sub-communities

Two rich media for anaerobe cultivation, brain heart infusion (BHI) and the Schaedler medium (SM), were used to cultivate microbial sub-communities in individual droplets (Figure 3.2). Initially, most sampled droplets contained only a few cells (Figures

3.2a and 3.2c,), but a subset of droplets exhibited substantial co-growth after incubation (Figures 3.2b and 3.2d). A range of morphologies were observed across these cultivated sub-communities. For instance, some droplets showed communities with distinctly long-rod morphologies (Figure 3.2b-3, 3.2d-8), while others consist of communities with distinct cocci morphologies (Figure 3.2b-4, 3.2d7). Interestingly, only a subset of droplets supported growth in rich media and with a high λ of 10 to ensure all droplets contained viable cells, demonstrating that not all sub-communities encapsulated were capable of co-growth in these conditions. Droplets that were individually selected from droplet spacing based on dense co-growth are shown in Figure B.4.

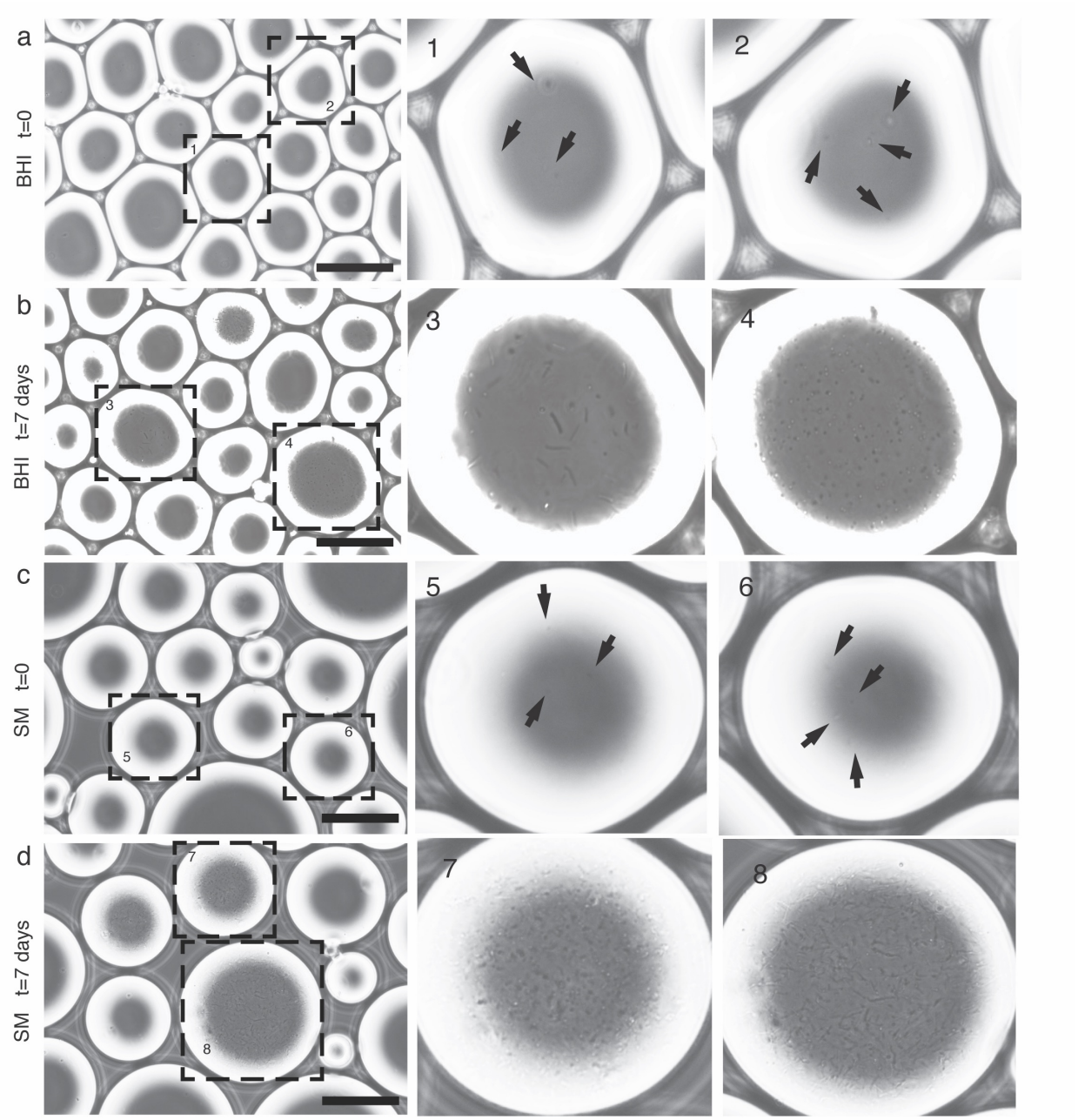


Figure 3.2: Microbial sub-communities in generated microfluidic droplets before and after co-cultivation. A sample pool of droplets with encapsulated microbial sub-communities before (a, c) and after anaerobic cultivation for a week (b, d). Droplets were cultivated in two rich media: brain heart infusion (BHI) (a,b) and Schadler media (SM) (c,d). Droplets are not tracked over time, so each droplet viewed is distinct. Dashed boxes on the left correspond to the magnified droplets in each subpanel on the right, identified by the numerical marker. Arrows distinguish single cells in the pre-incubation microfluidic droplet. Scale bar is 100 microns.

3.4.3 High variation in sub-community composition

A total of 22 droplets with high cell densities at the end of the incubation period were isolated and underwent 16S amplicon sequencing. Because amplification bias introduced with MDA heavily obscures any quantitative signal in sequencing (158), OTUs were analyzed by presence or absence to compare community composition across droplet sub-communities. The community membership of these 22 droplets was highly variable between droplets (Figure 3.3). Due to the stochastic nature of cell encapsulation during droplet generation and the high diversity of the original human gut microbiome sample, a relatively large number of OTUs appear in a relatively small number of co-cultivated sub-communities. However, OTUs of certain genera, such as *Staphylococcus* (OTU1), *Propionibacterium* (OTU27), and *Corynebacterium* (OTU4) (Table B.1), are pervasive across these sub-communities. Hierarchal clustering of droplet sub-communities demonstrated that two major clusters emerged, partitioning generally by growth medium. While this clustering is not significant (unbiased p-value < 95%) most likely due to the relatively small number of droplets analyzed, certain OTUs are highly associated with SM, such as *Parabacteroides* (OTU5) (indicator value = 90.9, p-value = 0.021) and *Butyricimonas* (OTU3) (indicator value = 75.3, p-value = 0.046) (Supplementary Table 2). This suggests that the medium composition is one factor determining which microbes are co-cultivated in the droplets.

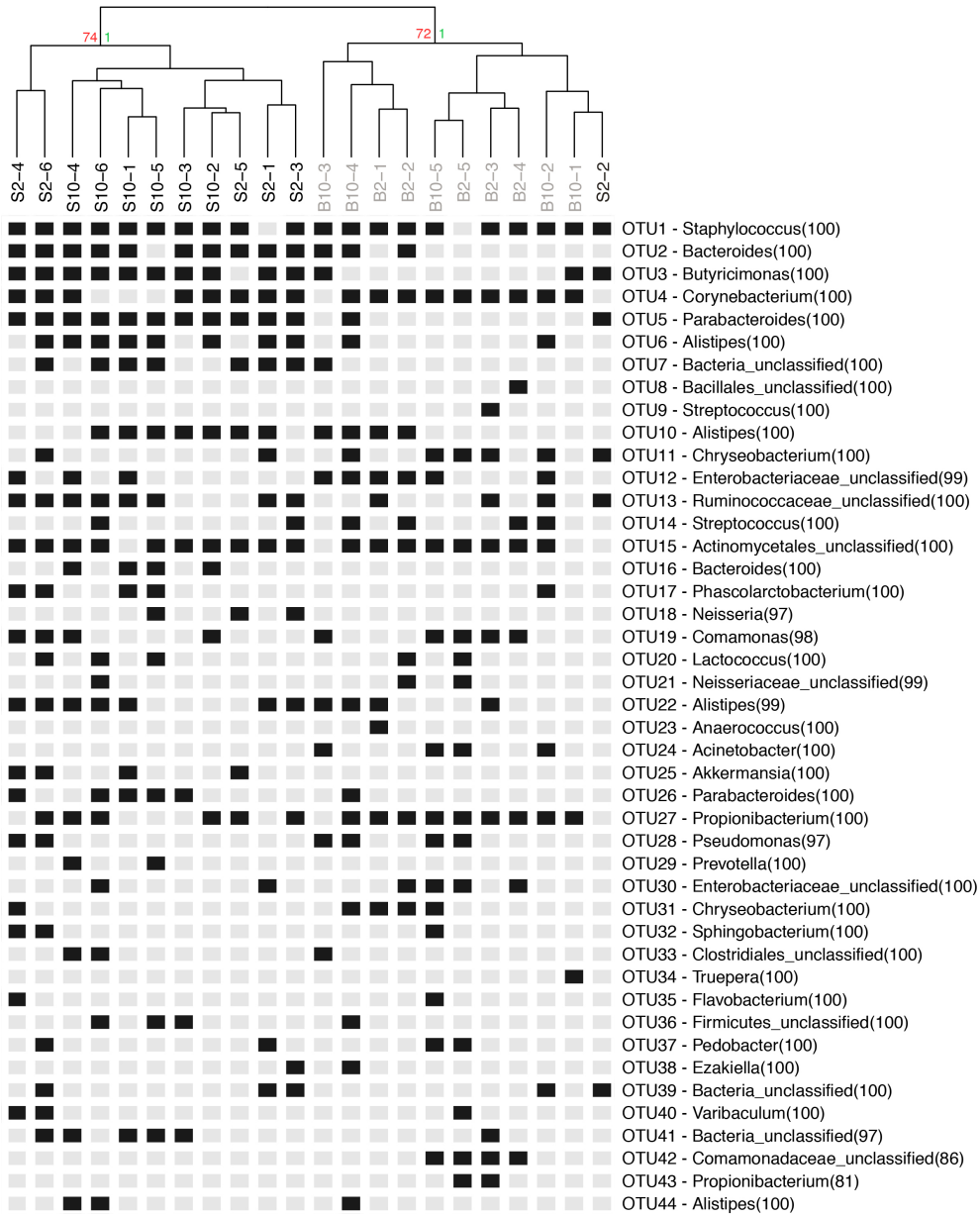


Figure 3.3: 16S OTU profiles of 22 isolated droplets sub-communities. OTU classification and bootstrap values are provided. Droplet identity nomenclature is based on Schaedler media (S) or BHI media (B) and with the initial λ value (2 or 10) and a numerical identifier. Because MDA introduced significant bias, quantitative information is not shown and OTUs are presented as either present (black) or absent (light gray) in a sample. Hierarchical clustering of the droplet taxonomic profiles provides two distinct clusters. P-values provided by pvclust in R do not signify statistical significance (AU < 95%).

Interestingly, while the droplets were generated such that the average initial cell number per droplet (λ) was either two or ten, most of the selected and analyzed sub-communities were found to each contain more than 10 OTUs. In fact, the difference in the species richness between droplets with λ of 2 and those of 10 was not statistically significant (p-value = 0.711, student's two-tailed T-test). In terms of community composition, a subset of the λ=10 droplets appeared to group together in the cluster dominated by droplets with SM, but no obvious patterns emerged in the other cluster dominated by droplets with BHI (Figure 3.3). Additionally, statistical testing showed that there were no strong associations between individual OTUs and either group of droplets with a specific λ (Table B.2). There are several possible reasons for this somewhat unexpected result. First, it is important to note that even when the λ value is 2 in the Poisson distribution, there are still droplets that each contain a relatively large number of cells (e.g. droplets containing 6 or more cells per droplet exist at a frequency of over 1.6% theoretically). Second, the droplets generated in this study were non-uniform, resulting in higher variation in the distribution of cell numbers than we would expect theoretically; in particular, larger droplets had more cells per droplet than expected and smaller droplets had less. Finally, our criterion for selecting droplets heavily favored those with strong bacterial co-growth. All these factors combined likely have led to the observed high diversity of the analyzed droplets, even those from the experiment with a low λ value of 2.

It is also important to note that the incubation time was one week, biasing our analysis towards sub-communities that demonstrated co-growth by the end of the

incubation time. Given a longer period, slower-growing bacteria could have been enriched and included in the analysis as well.

3.4.4 Accessibility to low-abundance microbial "dark matter"

Many genera observed in the droplet sub-communities are representative gut commensals. For example, representatives of the glycan-degrading *Bacteroides* and *Prevotella* (9,10,159), *Akkermansia* (160,161), and lactic acid bacteria (162,163) have established functional roles in host health that are relatively well understood as a result of culturing-dependent experimentation. However, commensals of the Neisseriaceae (164) and Clostridiales (165) are present in droplet sub-communities and have low sequence identities to NCBI 16S representatives (Table B.1). In addition to many of these members being largely uncharacterized, many of them are low-abundance members. To demonstrate this, we compared the amplicons in the droplet sub-communities with those in the bulk fecal sample from which our microfluidic droplet sub-communities were derived. A small fraction (15/44) of the OTUs in droplet sub-communities were detected in the bulk community, and 9 OTUs of those were represented at an abundance lower than 0.1% in the bulk sample (Table B.3). This result demonstrated the ability of microfluidic droplets to encapsulate and co-cultivate representatives at very low abundances in microbial communities.

3.4.5 Metagenomic reconstruction of a single droplet sub-community

We performed metagenomic shotgun sequencing, assembly, and binning on an individual droplet sub-community of interest (B2-2) (Table 3.1). B2-2 was chosen for further analysis due to the presence of lactic acid bacteria and several phyla that have

been characterized to a very limited extent. From 16S V4 amplicon sequencing, 13 OTUs were detected in B2-2. Out of 13 metagenomically assembled genomes (MAGs), 6 were of high quality (completeness above 75% and contamination below 10%) were recovered, 6 were of lower quality, and 1 (*Alistipes*) was missing. Additionally, it is interesting to note that in the metagenomic library, a partial *Lactobacillus* genome was recovered despite *Lactobacillus* not being detected in the 16S library of this droplet sub-community. We also report high degree of contamination by host DNA: 85.5% (347 million out of 406 million pair-end reads) of metagenomic reads mapped onto the human genome, resulting in a small fraction of sequencing effort contributing toward recovery of the bacterial community. Nevertheless, even from this relatively small fraction of sequencing reads, a substantial portion of the droplet sub-community, including several nearly complete genomes, was recovered.

16S V4 OTU identity	Genome identity	Length (Mb)	# of contigs	GC%	Completeness (%)	Contamination (%)
Propionibacterium (100)	*Propionibacterium acnes*	2.8	70	60.01	99.35	2.01
Corynebacterium (100)	*Corynebacterium* sp.	2.4	54	58.65	98.46	0.37
Actinomycetales unclassified (100)	*Lawsonella* sp.	1.6	25	52.4	95	0
Staphylococcus (100)	*Staphylococcus* sp.[1]	2.6	95	32.65	94.76	14.93
Lactococcus (100)	*Lactococcus lactis*[1]	2.1	198	35.81	85.72	13.26
Neisseriaceae unclassified (99)	*Neisseriaceae* nov. gen. nov. sp.	1.7	127	42.29	76.05	8.56
Streptococcus (100)	*Streptococcus* sp.	3	115	39.73	97.72	81.63
Enterobacteriaceae unclassified (100)	*Escherichia* sp.	3.9	249	50.86	69.17	2.96
N/A	*Lactobacillus* sp.	1.2	124	46.47	36.21	0
Enterobacteriaceae unclassified (99)	*Klebsiella* sp.	2.3	171	55.77	30.7	0
Chryseobacterium (100)	*Chryseobacterium* sp. (100)	1.5	143	38.35	25.16	5.17
Bacteroides (100)	*Bacteroides* sp. (100)	2.4	283	43.61	22.41	3.45
Bacteroides (100)	*Bacteroides* sp. (100)	1	119	38.27	8.88	1.75
Alistipes sp.	Not detected in metagenomic dataset					
1. Contamination values are below 10% for alternative marker gene set (40)						

Table 3.1: Assembly statistics of genome bins recovered from the bacterial sub-community in droplet B2-2 and their inferred taxonomies from CheckM and NCBI BLASTN.

3.4.6 Study of novel functional and phylogenetic diversity

From droplet B2-2, we recovered the genome sequence of an uncultivated member of the Neisseriaceae. Several lines of evidence indicate that this genome represents a novel genus within the family Neisseriaceae. BLASTN results against the NCBI 16S rRNA database signify the closest phylogenetic match as *Snodgrassella alvi* with 93% identity (Table B.1, OTU21). By phylogenetic analysis of a comprehensive set of conserved bacterial proteins (166), this member clearly falls within the Neisseriacaea but is quite distinct from other representatives (Figure 3.4a).

As its phylogeny suggests, the functional diversity of this novel representative of the Neisseriaceae is also distinct from other members of this family and the gut microbiome as a whole. Key metabolic characteristics from pathway reconstruction include a complete fatty acid oxidation pathway, acetate fermentation from pyruvate via acetyl-CoA, complete glyoxylate cycle, and carnitine degradation (Figure 3.4b). Fatty acids are primarily absorbed in the small intestine; however, under high fat diets, a significant fraction of fatty acids reach the large intestine and affect the microbiome and host health (167–169). Noting that the genome of this member was not completely recovered, it is still worth pointing out that the TCA cycle, among the best characterized and documented pathways, was not detected. This along with the presence of the acetate fermentation pathway suggests an anaerobic lifestyle, like many other members of the distal gut microbiota. In addition, the pathway for the conversion of carnitine, a nutrient in red meat, to γ-butyrobetaine (γBB) was present. γBB is an important intermediate for the microbe-dependent conversion to trimethylamine (TMA) (170,171),

which is eventually converted to trimethylamine-N-oxide (TMAO). TMAO has been demonstrated to increase risk to cardiovascular diseases (172,173), and the utilization of carnitine by gut microbiota has been demonstrated to play a crucial role in accelerating atherosclerosis in mouse models and potentially in human hosts as well (171).

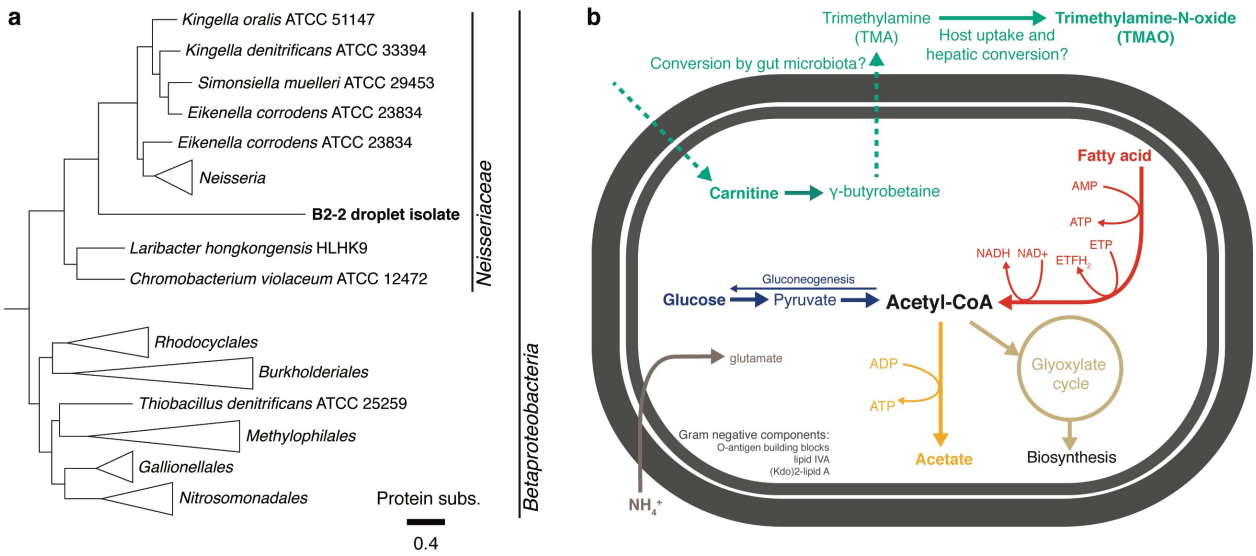


Figure 3.4: Phylogenetic and metabolic description of a novel member Neisseriaceae observed in droplet B2-2. (a) Phylogenetic tree comparing conserved protein sequences between the recovered genome with other Betaproteobacteria genomes from IMG. (b) Metabolic reconstruction of the most distinctive pathways.

3.5 Discussion

3.5.1 Summary

Recognizing the fact that bacterial species rarely live in isolation, we utilized microfluidic techniques and metagenomic approaches to dissect, cultivate, screen, and analyze complex communities comprised of uncharacterized bacteria. Rare and

uncharacterized members of the microbiome were analyzed with this methodology, including a novel Neisseriaceae with metabolic characteristics potentially important for human host health. The detection of this novel representative, which has complete fatty acid oxidation and carnitine degradation pathways, suggests possible niches that members of the Neisseriaceae occupy in the gut. This study is also the first to apply whole genome amplification, sequencing, and state-of-the-art bioinformatic tools to reconstruct and resolve multiple genomes from a co-cultivated sub-community in a microfluidic droplet. This approach can be used to address questions in a variety of natural systems where bacteria cannot be studied with traditional cultivation methods, such as those that have slow-growing or fastidious life styles.

We would like to note the improved genome recovery of certain MAGs (up to 99% completeness) in this study compared to previous studies using MDA to amplify single cell genomes, which recovered an estimated 70-78% of the total genome (174,175). This improvement is enabled by droplet co-cultivation, which allows growth-mediated “amplification” from a small number of single cells to populations, thereby increasing the DNA material and aiding the recovery of nearly complete genomes. While other microfluidic studies to recover genomes from complex microbiomes were able to achieve a completeness of 90% or higher, they required multiple droplets for sequencing and an amalgamation of contigs from the sequencing of multiple droplets (133,134). In contrast, all the draft genomes reconstructed in this study were from sequencing a single droplet.

3.5.2 Limitations

As much potential as this approach has, there are technical limitations. Because whole genome amplification, demonstrated here with MDA, introduces significant stochastic amplification bias, quantitative insights regarding community composition are lost. Alternative 16S amplification methods which better preserve quantitative community composition could be employed in parallel with MDA-enabled metagenomic sequencing. For example, touchdown polymerase chain reaction (TD-PCR) has been used to study low-biomass microbial systems (176) and could be applied to 16S amplicon sequencing of individual microfluidic droplets. In addition, although bacterial contamination was not present in appropriate co-cultivation controls (Supplementary Fig. 2), contamination during metagenomic sequencing and genomic recovery of droplet sub-communities was present, which is common for amplification from low-biomass samples. In addition to MDA indiscriminately amplifying host DNA from the human fecal sample, highly pervasive OTUs, such as *Staphylococcus*, *Propionibacterium*, *Corynebacterium*, and other members of Actinomycetales, are more typically skin commensals (177) and are suspicious of being molecular contamination as well. In the future, more thorough washing steps to reduce nucleic acid contamination from microbial suspensions and extensive sequencing controls of reagents should be performed as well.

One crucial aspect of the workflow demonstrated in this work is how co-cultivated sub-communities in individual droplets are selected for further sequencing and analysis. We carried out manual selection using a previously developed microfluidic device for droplet spacing (138,139). Future enhancement could leverage automated droplet sorting (104,178) to increase the throughput and robustness of the technology pipeline. It is also important to note that we selected droplets with high cell densities

hypothesizing that a high biomass yield may be the result of positive interactions between bacteria, which are prevalent in natural systems (52) and have been shown to occur effectively, for instance in the form of metabolic cross-feeding, in the microfluidic droplet environment (92,100). However, due to the complexity of the growth media employed in this work, inferring what ecological mechanisms led to enhanced co-growth is difficult. In addition, most of the analyzed droplet sub-communities comprised of a relatively large number of species, in part due to our selection criteria for high total biomass after co-cultivation. Communities with a higher diversity corresponding to a higher biomass productivity is a typical ecological phenomenon (179). Possible mechanisms for higher productivity increasing with species richness in these droplets include higher overall usage of droplet resources through niche complementarity, facilitation through cross-feeding, and higher chances of including faster-growing members that grow abundantly regardless of surrounding microbial partners. While the methodology demonstrates potential for microbial co-cultivation in droplets and the resolution of analysis, the specific experimental design is limited in capacity to determine mechanisms for co-growth.

3.5.3 Conclusion

With more specific questions, metagenomic analyses of co-cultivated sub-communities in microfluidic droplets has the capability to shed light on the ecology of many uncharacterized microbial systems. In particular, cross-feeding is believed to be widespread throughout complex bacterial communities (48,52), but most demonstrated examples of this are in highly-simplified synthetic communities (55,89,180). Such ecological interactions can be studied not just for simple synthetic systems, but sub-

communities of naturally occurring complex microbiomes, including those which have eluded cultivation-based efforts. Possible applications of this methodology can especially be applied to less complex, yet undefined, natural communities. Dissecting and co-cultivating these communities in droplets in defined media can elucidate complementarity of genomic pathways across species which leads to co-growth. Further development of single droplet resolution applied to bacterial transcriptomics and metabolomics would provide the means to elucidate mechanisms for even more complex systems.

Chapter 4 : High-throughput Quantitative 16S Profiling for Droplet-based Combinatorial Decomposition

4.1 Summary

The ability to infer interactions in microbiomes has been limited to genomic inference from metagenomic data and in laboratory-constructed synthetic co-culture of representative isolates. Both approaches are limited, by the complexity of natural systems and metagenomic resolution and of tractability and characterization in the laboratory. By quickly reducing the complexity and due to the random encapsulation statistics, microdroplets can quickly decompose a natural community into highly parallel combinatorial subsets and allow for the inference of interactions within these subsets. Doing so requires the ability to evaluate many droplets for statistical inference and for each, analyzing the membership of droplets and evaluate the extent of co-growth to determine if an interaction is present. This chapter elaborates the methodological development of a droplet-resolved, quantitative 16S amplicon sequencing library preparation workflow for application on the combinatorial decomposition and co-cultivation of a natural community. To demonstrate the capability of the method, we generated two mock community droplet libraries of known microbial composition with different abundance profiles and applied the method on them. This is the first development of this methodology and has implications for the establishment of

microdroplets as a staple tool in microbial ecology and the study of microbial interactions.

4.2 Introduction

One of the major limitations of culture-independent approaches is the lack of phenotypic data. While genomic information can be quite powerful, it on its own is insufficient to provide certainty if cells are indeed interacting. However, the major limitation of culture-dependent approaches such as synthetic ecology is the scalability. Many synthetic ecology studies center around the technical ability to manipulate different isolate cultures to grow different sets of combinations and analyze the resulting co-cultures to analyze metabolites and community composition (86). However, as the number of members increases, the total number of combinations increases dramatically more. For a community of ten members, the total number of combinations is 2^{10} = 1,024. For 20 members, the total number if 2^{20} =1,048,576. Even for automation by a liquid handling robot, this can quickly become a large task. One of the main advantages of microdroplets is its ultra-high-throughput. The production of droplet is on the range of 1000 per second and can more easily generate the different combinations of cell types due to the stochastic encapsulation statistics, given that the cells of interest are of reasonably relative abundance in the community. Microfluidic encapsulation also allows for the manipulation of these cell types without the need to isolate individual members, which can very quickly become a labor-intensive effort on its own. With this implementation, synthetic ecology could be performed on natural samples of varying diversity and of varying previous characterization efforts.

As demonstrated in Chapter 3 of this dissertation, while co-cultivation of sub-communities has much potential, there is still limited information that can be gathered from one single sub-community. In contrast, this chapter is focused on the development of a platform to study the membership and co-cultivation of thousands of droplets, regardless of the degree of co-growth exhibited. For around two decades, 16S sequencing has allowed microbial ecologists to study membership of different environments (67), and advances in spike-in standards has allowed for quantitative amplicon sequencing (181). However, how can this be applied when the different samples are distinct droplets and how does one add a spike-in standard? This is a difficult technical microfluidics and molecular workflow problem. The most similar work to date is Sheth et al. which uses droplet microfluidics and barcoding to perform high-throughput biogeography mapping of the gut microbiome (182). This work was able to demonstrate the use of droplet microfluidics in combination with 16S sequencing to answer the basic question of "who is next to whom?". This study was highly creative and novel, but it did not incorporate co-cultivation or a quantitative signal. Incorporating a quantitative signal has been demonstrated in numerous transcriptomic droplet studies (183). However, these do not incorporate cultivation and the molecular workflow for 16S sequencing is very different from eukaryotic transcriptomics. Hsu et al. (94) demonstrated microbial interaction network inference in microdroplets (MINI-Drop) which used statistical analysis across a large number of droplets to quantify the magnitude of interactions between model co-cultures with fluorescence microscopy and computer vision to determine the absolute composition of cells in each droplet after co-cultivation and co-cultivation. However, while the proof of concept of interaction

inference using droplets was demonstrated well, the generalization of this method to natural communities is limited due to its reliance on fluorescence markers in each of the community members.

This chapter details the methodological workflow we developed to extend the conceptual motivation of MINI-Drop with sequencing (MINI-Drop-seq) for the study of interaction networks in natural communities. We benchmark the technical workflow of MINI-Drop-seq, specifically its ability to determine the absolute abundance of thousands of individual droplet communities. To do so, we applied MINI-Drop-seq on droplet communities derived from mock communities of defined compositions to benchmark its efficiency and accuracy. Finally, we suggest improvements to the current workflow for better sequencing efficiency and generalizability for different microbial systems. We also propose further benchmarking to better characterize the MINI-Drop-seq workflow and future microbial systems to study.

4.3 Materials and Methods

4.3.1 Microbial cultures

Benchmarking was performed with five microbial cultures: *E. coli* K12 BW25113, *B. subtilis* 168, *P. putida* KT2440, *B. thetaiotaomicron* VPI-5482, and *L. crispatus* ATCC 33820 to test the workflow on cells with different cell physiologies. *E. coli* K12 BW25113, *B. subtilis* 168, and *P. putida* KT2440 were grown overnight in LB miller media; *E. coli* and *B. subtilis* were grown at 37°C, and *P. putida* at 30°C. *B. thetaiotaomicron* VPI-5482 was grown overnight in RUM (modified YCFA) media (184) supplemented with 4 g/L of fructose at 37°C anaerobically. *L. crispatus* was grown overnight in MRS media at 37°C anaerobically. Anaerobic growth was performed in an

aerobic chamber (Coy) with a 5% carbon dioxide, 2-4% hydrogen, and balance nitrogen atmosphere. For each cell culture, 1 mL of overnight culture was centrifuged in a 1.5 mL microcentrifuge tube at 4000xg for 5 minutes, the supernatant was decanted, and the cell pellet was resuspended twice in PBS. Each washed culture was appropriately diluted and quantified with a C-Chip disposable haemocytomer (Fisher Scientific, #22-600-100) under a Nikon Ti-S microscope on the “Ph2” filter with a 40X objective lens. Mock communities were composed of four members - *E. coli*, *B. subtilis*, *P. putida*, and *B. thetaiotaomicron* - and were mixed together according to twice the desired community compositions for the two mock communities: (1) the “even” community with a λ of 50 cells of each species per droplet after 2 times dilution and (2) the “log” community with a λ of 180, 18, 1.8, and 0.2 cells per droplets after 2 times dilution, respectively. Monoculture suspensions were prepared for all cultures to a λ of 100 cells per droplet after 2 times dilution.

4.3.2 Droplet encapsulation

Microdroplet cell encapsulation was performed in a modified large oven incubator (VWR Scientific 1535) to maintain a temperature between 37-40ºC to keep low melting agarose suspensions liquid. To limit heat loss, the incubator had the outer door removed and the inner glass door replaced with an equivalent-sized, clear acrylic sheet (Optix, 30"x36"x.22") with holes cut out for the entry and exit of hands for the exchange of suspensions or adjustment of syringe pumps (Figure C.1). A 3% agarose solution with 30% Optiprep in PBS was prepared by melting low melting-point temperature (SeaPlaque GTG, Lonza) in 30% Optiprep in PBS at 70ºC in 50 mL Falcon tube in a water bath, filtering the solution through a sterile 0.45 µm PDVF filter into a new Falcon

tube, and keeping the tube at 70ºC in the water bath until use. Right before encapsulation, the 3% agarose solution was mixed with the cell suspensions in PBS in a 1:1 volume ratio to obtain the appropriate cell suspensions in 15% Optiprep and 1.5% agarose in PBS for droplet generation. Around 600 µL of cell suspension was quickly withdrawn into a sterile 1 mL syringe with a 24 G x 1" syringe needle (Terumo), any air bubbles were removed, and the syringe was placed into the heated incubator oven to equilibrate with temperature. Two syringe pumps (Kent Scientific Genie Touch) in the incubator were used to flow the cell suspension and the oil phase comprising of HFE7500 Novec engineering oil with 2% fluorosurfactant (008-fluorosurfactant, RAN Biotechnologies, Inc.) into the microfluidic device. The microfluidic device is a flow-focusing device with channel heights of 50 µm and channel widths of 25 µm (Figure A.1). The flow rate of the oil phase and agarose suspension into the microfluidic device is 10 µL/min and 5 µL/min, respectively. PFTE tubing (23g) is used to flow each phase into the inlets of the device and generated droplets and oil at the outlet to a microcentrifuge tube for collection. Approximately 200 µL of droplets were generated for each mock community and monoculture suspension and were set as microgels in droplets by placing the collection tube after generating ice for 10-20 minutes.

4.3.3 Agarose microdroplet processing

Oil phase below droplets in the collection tube was removed with a micropipette and 200 µL gel loading tip. We added 500 µL of PBS-wash buffer (1X PBS, 0.1% Triton X-100) and 500 µL of 20% perfluorooctanol in HFE7500 oil to the remaining droplet emulsion, vortexed well, centrifuged for 1 min at 300xg, and removed as much oil phase as possible without removing aqueous phase holding the collected microgels. We then

added 500 µL of 1% Span-80 in hexane, vortexed well, and centrifuged for 1 min at 300xg. The clear hexane layer on top was removed and residual hexane-water cloudy emulsion below was left. The residual hexane was removed by adding 500 µL of PBS-wash, vortexing, centrifuging, and skimming off the frothy/milky layer on top with a pipette tip. This was repeated 3 times until the supernatant is clear and the microgels formed a distinct pellet at the bottom. The suspension was mixed by pipetting and transferred to a clean 1.5 mL microcentrifuge tube. The tube was centrifuged, and supernatant was removed by pipetting without disturbing the gel pellet. Gels were washed with 1 mL 10 mM Tris-HCl three times; the resulting gel suspension was mixed with a 2X lysis buffer in a 1:1 ratio for a final concentration of 1 mM DTT, 1 mM Tris-HCl pH 8.0, 2.5 mM EDTA, 100 mM NaCl, 0.8% ready-lyse lysozyme (Lucigen, R1804M); and incubated in a 37°C incubator shaker overnight. Lysis buffer was removed from gels after centrifugation, and gels were washed three times in 10 mM Tris-HCl pH 8.0. The resulting suspension was mixed with a 2X digestion buffer in a 1:1 ratio for a final concentration of 30 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.8% Triton X-100 (v/v), 0.5% SDS, 1 ug/µL proteinase K (Lucigen, MPRK092), and incubated in a heat block for 30 minutes at 50°C. To deactivate proteinase, gels were centrifuged; digestion buffer was removed; the remaining gels were washed in 10 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.1% Tween-20 (v/v), 5 mM phenylmethylsulfonyl fluoride (PMSF) (Sigma-Aldrich 93482) and washed in 10 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.1% Tween-20 (v/v) (TET buffer) five times; and stored at 4°C until use. Occasionally, a fine, white precipitate would form when the digestion buffer was added. Since the precipitate was much smaller than the gels, it would settle at the bottom of the tube during

centrifugation and the gels would sit on top. The gels would be carefully removed by pipetting the microgel layer without removing precipitate into a separate microcentrifuge tube.

To check that genomic DNA was immobilized in the microgels after cell lysis, 10 µL of packed microgels were incubated for 30 minutes in 1 mL of 10 mM Tris HCl with 1X SYBR stain in the dark. Afterwards, gels were pelleted by centrifugation and washed in TET buffer three times, and visualized in a disposable haemocytomer under fluorescence microscopy with a FITC filter.

4.3.4 Barcode bead synthesis

Barcode hydrogel beads were purchased as a unit of 1 million suspended in TET buffer (RAN Biotechnology, CustomSeqReady-1M). The functional oligonucleotides on the beads do not include the 16S V4 forward primer, so it needs to be extended onto the oligonucleotides after purchase. 16S 515f primer extension is based off the protocol from Zilonis et al. (183) for barcode extension on hydrogel beads. The purchased 1 million beads were centrifuged at 1000xg; washed three times in hydrogel bead wash buffer (5 mM Tris-HCl pH 8.0, 5 mM EDTA, 0.05% Tween 20); and incubated in 1X isothermal amplification buffer (NEB B0537S), Bst 2.0 DNA polymerase (350 U/mL) (NEB M0537S), dNTP mix (650 µM each), and 10 µM of the extension oligo comprising of the annealing region to the existing oligonucleotide and 16S 515f primer overhang (5'-TTACCGCGGCKGCTGRCACCTCCTGTCATCTCACTCCTG) for 1 hour at 60ºC protected from light. Beads were centrifuged, supernatant was removed, incubated in STOP-25 buffer (10 mM Tris-HCl pH 8.0, 25 mM EDTA, 0.1% Tween-20, 0.1 M KCl) at room temperature for 30 minutes, and washed and incubated in STOP-10 buffer (10

mM Tris-HCl pH 8.0, 12.5 mM EDTA, 0.1% Tween-20, 0.1 M KCl) three times at room temperature. To remove any annealed oligo on the beads, beads were washed and incubated in denaturation solution (0.15 M NaOH, 0.5% Brij-35) three times and then washed with neutralization buffer (100 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.1% Tween-20, 0.1 M NaCl) twice. Beads were washed three times with hydrogel bead wash buffer and washed three times with hybridization buffer (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 0.1% Tween-20, 0.33 M KCl). Any unelongated oligonucleotides on the bead are removed by ExoI clean-up. To do so, beads are incubated for 30 minutes with 20 μM of protective oligos complementary to the 16S 515f reverse region, and then incubated in 1X ExoI buffer and ExoI (0.27 U/μL) (Thermo Fisher Scientific, EN0581) for two hours at room temperature for digestion of any single stranded oligos without hybridization of the protective oligo. To remove the protective oligos, beads were washed in STOP-25 buffer, incubated, washed in STOP-10 buffer three times, incubated and washed in denaturation buffer three times, washed in neutralization buffer twice, and washed in TET buffer three times for storage at 4ºC.

After the extension protocol, barcode beads were validated to ensure proper 16S forward primer addition on oligonucleotides. Two aliquots of 5 μL of packed, extended barcode beads were taken from the finished tube after centrifugation and placed in separate 1.5 mL microcentrifuge tubes, one for checking the extended 515f forward region and the other for the preexisting PE1 region. Respective FAM oligo probes (purchased from Integrated DNA Technologies, 5-/56-FAM/TTACCGCGGCKGCTGRCAC for the 515f region and 5-/56-FAM/AGATCGGAAGAGCGTCGTGTAGGGAAAGAG for the PE1 region) were

annealed to the bead oligonucleotides at a concentration of 10 μM in 1 mL QC buffer (5 mM Tris-HCl pH 8.0, 5 mM EDTA, 0.05% Tween-20, 1 M KCl) in each tube, incubated for 20 minutes at room temperature in the dark, and washed with QC buffer three times to wash off free FAM probes. Beads were concentrated with centrifugation and visualized on C-Chip cell count haemocytomers (SKC, Inc., DHCN015) for fluorescence with FITC filters to confirm the presence of both regions on the beads.

4.3.5 16S standard design and quantification

The 16S standard is an oligonucleotide composed of a modified 16S V4 sequence from *Thermus thermophilus* ATCC 33923. The full sequence is as follows: GTGCCAGCAGCCGCGGTAANNNNNNNGGCGCGAGCGTTACCCGGATTCACTGGG CGTAAAGGGCGTGTAGGCGGCCTGGGGCGTCCCATGTGAAAGACCACGGCTCAA CCGTGGGGGAGCGTGGGATACGCTCAGGCTAGACGGTGGGAGAGGGTGGTGGA ATTCCCGGAGTAGCGGTGAAATGCGCAGATACCGGGAGGAACGCCGATGGCGAA GGCAGCCACCTGGTCCACCCGTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAAC CGG*ATTAGATACCCGGGTAGTCC*. The first 19 and last 20 nucleotides of the sequence are conserved for annealing of the standard 16S 515f and *16S 816r* primer, respectively, and the seven nucleotides directly after the 515f region are replaced with seven degenerate oligonucleotides for use as unique molecular identifiers. The 16S standard was purchased from GenScript as a single strand DNA (ssDNA) lyophilized stock with PAGE purification and verification for quality control. The ssDNA stock was resuspended in molecular-grade water and quantified with an in-house digital droplet polymerase chain reaction (ddPCR) to determine the concentration of functional 16S standard molecules. To do so, we diluted the stock 10^4 to 10^9 fold and made PCR

reactions for each. The PCR reaction was composed of 0.6 µL of 816r 16S V4 reverse primer (10 µM), 0.6 µL of 515f 16S forward primer (10 µM), 5 µL of 5X SuperFi II Buffer, 0.6 µL of 10 mM dNTP (each) mixture, 0.6 µL of Platinum SuperFi II DNA polymerase, 0.375 µL of 20 mg/mL bovine serum albumin (BSA), 3 µL of 10% Pluronic F-68, 3 µL of 16S standard dilution, and 15.225 µL of molecular-grade water. This reaction volume was run through the flow-focusing microfluidic device with QX200 Droplet Generation Oil for EvaGreen (Biorad, cat # 1864005) to produce droplets with a diameter of approximately 40 µm in diameter. The droplets were collected in a 0.2 mL PCR tube for each dilution, excess oil was removed, and covered with 50 µL of mineral oil. The droplets were thermocycled using the following program: 95ºC for 2 min, 98 for 30 sec, 30 cycles of (98 for 10 seconds, 60 for 10 seconds, and 72 for 30 sec) with a 2 ºC/s ramp rates between steps and no lid heating. Mineral oil was removed, and the oil phase was replaced with 10X SYBR in 2% surfactant in HFE7500 oil (0.2 µm filtered to remove excess SYBR in DMSO) and incubated for 20 minutes and was washed twice in fresh 2% surfactant in HFE7500 for 30 minutes. Droplets were viewed in a C-Chip under a FITC filter to identify the correct dilution that would provide a λ of 0.1 template molecules/droplet. This condition was imaged, and images were processed with a custom MATLAB script to determine the concentration of functional 16S standard in the standard stock. The stock 16S standard concentration was approximately 2.91×10^6 molecules/nL.

4.3.6 Droplet barcoding

515f-extended barcode beads were washed in bead buffer (10 mM Tris HCl pH 8.0, 0.1% Tween 20, 50 mM KCl) by washing in 0.5 mL PCR tubes by centrifugation at

1000 g for 1 minute three times, and a gel-loading pipette tip was used to remove as much liquid as possible after centrifugation at 3000 g for 2 minutes. Mock community agarose gels were washed in 10 mM Tris-HCl pH 8.0 by centrifugation at 1000 g for 1 minute three times and as much liquid was removed with a gel-loading pipette tip. Agarose gels were diluted 1:2 by volume with 1.5X agarose gel suspension buffer (15% OptiPrep, 1.5% (w/v) Pluronic F-68, 0.45 mg/mL BSA, 15 mM Tris-HCl pH 8.0) for a 33% agarose suspension by volume in a 0.5 mL PCR tube.

A 2x concentrated PCR mastermix solution was made by combining 6 µL of PE2-816r reverse primer (10 µM) (for a final concentration of 0.4 µM), 60 µL of 5X SuperFi II buffer, 6 µL of 10 mM dNTPs (for a final concentration of 400 µM each base), 6 µL of Platinum SuperFi II DNA polymerase, 18 µL of 10% Pluronic F-68 (v/v), 2.25 µL of 20 mg/mL bovine serum albumin (BSA), 29.1 µL of the 10^{-4} dilution of the 16S standard stock, and 22.65 µL of molecular-grade water. The desired λ for the 16S standard was 50 molecules/droplet.

The device used for droplet barcoding is modified device available from Droplet Genomics Inc (purchase upon request). The device was originally designed for dual encapsulation of 60-70 µm diameter polyacrylamide beads. However, because our workflow requires the encapsulation of both polyacrylamide beads and agarose microgels, modifications were made (Figure C.2). Specifically, because of the differences in viscoelastic flow properties between a polyacrylamide bead suspension and an agarose microgel suspension, the inlet for the agarose microgel suspension was punched in a different location than originally designed.

The microscope used to monitor droplet encapsulation of beads and microgels was a Nikon Ti-S inverted microscope. A red band-pass filter (Midopt, BP635) was equipped into the condenser filter to prevent exposure of trace amount of UV-light from the microscope onto the UV-light sensitive barcode beads. For syringe pumps (three GenieTouch syringe pumps from Kent Scientific and one from KD Scientific) were used for microfluidic operation.

The barcoding device has four inlets for the PCR reagent carrier phase, the packed barcode beads, the agarose microgel suspension, and the oil, as well as one outlet for the droplets. For each inlet phase, each is controlled by a syringe pump with an Air-Tite 1 mL plastic syringe filled with Biorad droplet generation oil with syringe needle. To draw up each phase into PFTE tubing, PFTE tubing is cut and a transfer syringe - an empty syringe with needle - is inserted into one end of the tubing and is used to manually draw an appropriate amount of each phase from a microcentrifuge tube into the PFTE tubing without drawing up air. Before withdrawing the barcode bead or agarose microgel suspension, the suspensions are pulse-vortexed 10 times to ensure homogenous distribution and break up clusters of gels or beads. The end of the tubing is removed from the transfer syringe and placed onto the end of the syringe and needle filled completely with oil, without any air between the liquid interface and the needle syringe with oil. The other end of the PFTE tubing is attached to the inlet of the device. The syringe pump is run to slowly prime the liquid suspension to the inlet of the device at a steady flow of 10 µL/min. Approximately 100 µL of packed barcode beads were loaded into the PFTE tubing and covered with a black tubing sheath (McMaster-

Carr) to protect the barcode beads from light. Around 30 µL of agarose microgel suspension was loaded, and approximately 100 µL of 2X PCR mix was loaded.

The flow rates for the syringe pumps were as follows: the oil flow was 5 µL/min, the barcode beads at 0.5 µL/min, the agarose suspension at 1.5 µL/min, and PCR reagent at 2 µL/min. Occasionally, due to particle arching or larger microgels getting stuck, the flow for the agarose suspension needed pulsing manually to keep the suspension flowing.

For each condition, we collected 6 min of barcoded droplets from the outlet in 0.5 mL PCR tubes. For droplet thermocycling in 0.2 mL PCR tubes, each condition had 20 µL of Biorad oil and 30 µL of droplets from barcoding. Droplets were exposed to UV-light on ice under a 365-nm UV-light (Ted Pella Blak-Ray) for 10 min (with the lamp pre-heated for 10 minutes beforehand). 50 µL of mineral oil was placed on top of the droplets and the reaction tubes were briefly centrifuged on the mini-centrifuge. Droplets were thermocycled using the following program: 95 ºC for 2 min; 98 ºC for 30 sec; and 30 cycles of 98 ºC for 10 seconds, 60 ºC for 10 seconds, and 72 ºC for 30 sec), with a 2 ºC/s ramp rates between all steps with no lid heating. Mineral oil was removed immediately after thermocycling finished and as much excess oil was removed with a gel-loading pipette tip.

For PCR clean-up, 40 µL 1X ExoI buffer with 1 U/µL ExoI and 30 µL of perfluorooctanol was added for each condition and briefly centrifuged to merge droplets and incubated at 37 ºC for 30 minutes. Approximately 50 µL of aqueous phase was transferred without any oil to new 0.2 PCR tubes for clean-up with AMPure XP beads.

Following the AMPure XP protocol (Table C.1), we then eluted in 20 µL of 10 mM Tris HCl 8.0.

4.3.7 Library preparation and sequencing

Library preparation was performed with PCR in a 50 µL reaction comprising of 2.5 µL of each library primer with appropriate indices (0.5 µM stock) (Table C.2), 25 µL of 2X NEBNext Q5 HotStart HiFi PCR MasterMix, 0.125 µL of BSA (20 mg/mL), 9.875 µL of molecular grade water, and 10 µL of DNA clean-up elution. The thermocycler program was as follows: 98 ºC for 30 seconds; 10 cycles of 98 ºC for 10 seconds, 68 ºC for 20 seconds, and 65 ºC for 30 seconds; 65 ºC for 2 minutes; and a 12 ºC hold. PCR product was verified on a 1.5% agarose gel after running gel electrophoresis at 100 mV for 50 minutes, ethidium bromide staining, and 5 minute wash in TAE buffer. For PCR clean-up and removal of non-specific amplification, we used the QIAEXII Gel Extraction kit to selectively purify the expected ~490 bp DNA band. To ensure that the product was correct, the primer-library mix was sent to Eurofins Sequencing for Sanger sequencing with the P5 and P7 Illumina adaptors as primers. Verified libraries were sent to the University of Michigan Advanced Genomics Core for sequencing with the Illumina NextSeq 2000 with P1 300 cycle chemistry. Paired end sequencing was done with asymmetric read lengths, the forward read with 260 base pairs and the reverse with 40 base pairs.

4.3.8 Bioinformatics

A custom python script from Zilionis et al. (183) originally used for single-cell transcriptomics was modified and used to demultiplex droplet libraries and trim

barcodes, adaptors, and the 16S 515f forward primer from reads. A 16S V4 reference database for the cultured representatives and *T. thermophilus* for the 16S standard was generated with the pcr.seqs command in mothur with the reference genomes in NCBI. The operational taxonomic unit (OTU) table was produced by running the query reads from demultiplexing and trimming against the 16S V4 database with the usearch_global command in vsearch (v.2.13) (185). Samples with less than 100 reads were removed. Any samples with more than 90% and 75% of the reads for the “even” and “log” mock community libraries, respectively, identifying as 16S standard were considered droplets without genomic DNA and removed. In addition, any samples with less than 10 total 16S standard reads were also removed. Distances between individual mock community samples were calculated and a principle component analysis was performed in mothur (v.1.44.1) (144). Relative abundances calculations accounted for 16S copy numbers retrieved from rrnDB (186).

4.4 Results

4.4.1 Workflow overview

To provide insight into the community interactions responsible for co-growth, the developed workflow needs to answer two questions: "who is in each droplet?" and "to what extent did they grow?". To answer who is in each droplet, we can apply recent and established advances in 16S sequencing (67,143). However, having droplet-resolution, particularly of who is present is much more technically challenging and requires further methodological development. To do so, we used hydrogel beads with bead-unique oligonucleotides barcodes utilized in similar studies (182,183). These beads are encapsulated approximately one-to-one in microdroplets with the template DNA of

interest to allow for droplet-specific barcoding after tagmentation or extension of the barcodes onto sequences. To answer to what extent did they grow for each droplet necessitates a quantitative 16S sequencing platform. Typically 16S Illumina sequencing provides relative abundances of the communities in question, but studies using spike-in internal standards allow for the interpretation of absolute abundance (181). Typically for these internal standards, a known amount of a sequence standard is added to the sample, and this allows for reconstruction of the absolute abundances from relative abundance data based on the spike-in ratio. A similar strategy could be applied to droplets for individual 16S droplet quantitative sequencing, but one complication is the stochasticity of the number of exact standard molecules introduced in the droplets due to Poisson statistics. To be an effective standard, the precise number of standard molecules needs to be known for each droplet. One approach relevant from recent sequencing workflows is the usage of unique molecular identifiers (UMIs) (187). Unique molecular identifiers are degenerate oligonucleotide regions introduced in sequences that are randomly generated in the oligonucleotide synthesis process to quantify reads on a molecular level even after sequence amplification. Due to the sheer number of unique UMIs – determined by 4^n where n is the number of degenerate base pairs – the probability of the same UMI reoccurring twice in a small pool of UMIs is very low. Taking inspiration from this, we designed a 16S standard by introducing a UMI region into a known 16S sequence derived from the organism *Thermus thermophilus*, ordered as a ssDNA oligo, quantified the stock concentration of this oligo precisely with digital droplet polymerase chain reaction (ddPCR) for encapsulation into droplets.

The workflow incorporating both methodological strategies is shown in Figure 4.1. Sub-communities are generated by microdroplet generation, co-cultivated at appropriate conditions, and cells from each droplet are immobilized in an agarose microgel, microgels are collected, and cells are lysed in the microgels. The agarose matrix for each microgel prevents the diffusion of large molecular weight particles, including cells and genomic DNA. This allows everything to be suspended in one convenient aqueous phase yet still discretized in biologically-meaningful units that can be physically handled downstream for individual analysis. Droplet barcoding takes the lysed genomes immobilized in agarose microgels and co-encapsulates them in droplets with the barcode beads in a liquid phase containing the required PCR reagents and the 16S standard molecules with UMIs for quantification downstream. Droplet-specific barcodes are attached to the 16S V4 amplicons during droplet PCR. The PCR buffer is formulated to keep droplets stable during thermocycling to prevent barcode cross-talk in merged droplets. The droplets are then merged, barcoded 16S V4 amplicons are washed with a PCR clean-up protocol, and then sequencing libraries are generated with another PCR. After sequencing, droplet libraries are demultiplexed bioinformatically and quantified on an absolute basis with the spike-in 16S standard. Using presence and absence and the large number of samples, statistical inference is used to determine which co-culture combinations have interactions. Monoculture libraries can be generated from the microbial sample with a low-λ droplet encapsulation (e.g. 0.1 cells/droplet) and cultivation to compare with co-cultivation libraries to determine which co-cultures result in better growth compared with monoculture.

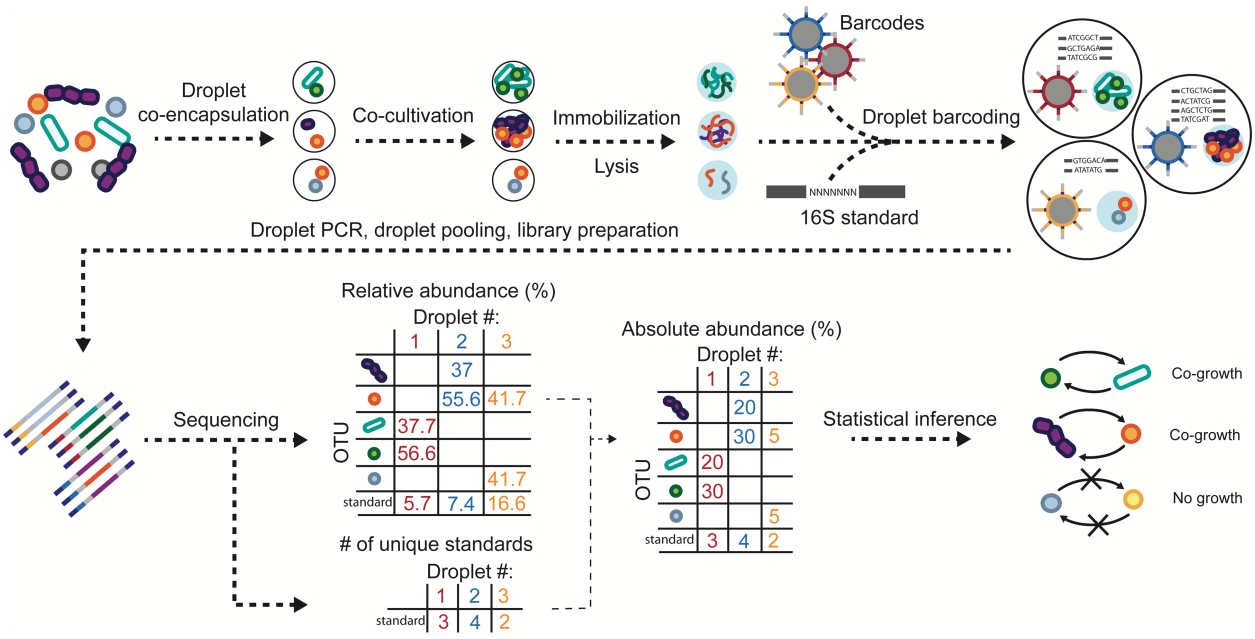


Figure 4.1: Workflow overview. Droplets are used to decompose a natural community into subset communities and are co-cultivated and immobilized in an agarose microgel. The agarose microgel allows for fixation of the cells and their genomic material during lysis and washing in the microgel. During droplet barcoding, the microgels are paired with barcode beads to introduce droplet-specific signals and a 16S standard with a unique molecular identifier (UMI) for absolute quantification after sequencing. To attach and amplify 16S regions, droplet PCR is performed, and droplets are pooled and subject to library preparation for Illumina paired-end sequencing. Bioinformatic analysis is used to determine the relative abundance of each droplet community and the number of unique standards based on 16S standard UMIs. Together this information is used to determine the absolute abundance of members in each droplet community for interaction inference.

4.4.2 Design of barcode beads

The detailed bead barcode design is presented in Figure 4.2. When purchased, the oligonucleotides compose of a photocleavable spacer, a T7 promoter, a PE1 region for downstream library preparation, two barcode regions connected by an adaptor sequence, a 6 base pair UMI sequence, and another adaptor. Two regions of the oligonucleotide on the beads are unutilized in our workflow: the T7 promoter sequence and the UMI. The original molecular design for these beads is for a single-cell

transcriptomics workflow which utilizes linear amplification with the T7 promoter region and absolute quantification of transcripts with the 6 bp UMI region (183). The original design uses a dual index "split and pool" design to generate 147,456 unique bead barcodes from two sets of barcode libraries comprising of 384 unique barcodes each (183). The last "custom sequence adaptor" is meant as a universal adaptor for users to extend their own region of interest onto the oligo. To do so, we hybridized an oligo with a region complementary to the custom sequence adaptor with a 16S V4 515f forward primer overhang and extended the oligo on beads to incorporate the 16S forward primer. The extension oligo is removed and to verify the addition of the 16S forward primer, we hybridized a fluorescent probe complementary to the 16S forward primer onto oligos on the barcode beads to ensure correct hybridization (Figure 4.2b). After droplet encapsulation of the beads with microgels with genomic material, the oligonucleotides are cleaved off the bead with UV-light for droplet PCR to amplify the 16S V4 region with the released oligo and 16S 816r reverse primer, and then library preparation is used to amplify the entire construct.

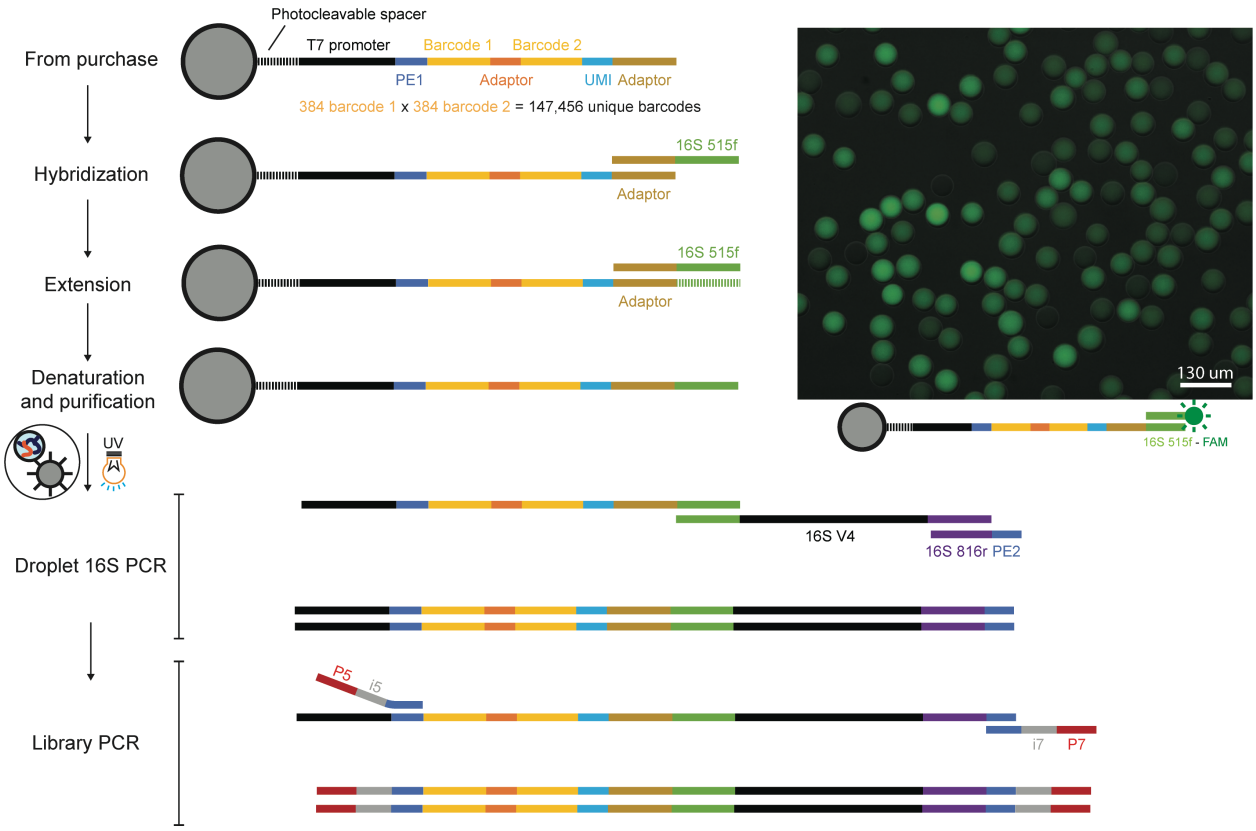


Figure 4.2: Barcode bead design. Barcodes are dual-index, composed of pairwise combinations between two sets of 384 barcodes for a total of 147,456 possible unique barcodes. The 16S primer is attached by hybridization and extension by polymerase. Bead oligonucleotides include a T7 promoter and a unique molecular identifier (UMI) which are not utilized in our workflow. During droplet barcoding, oligonucleotides are released from the bead by UV-light treatment. During library PCR, library adaptors (P5 and P7) are annealed onto the PE1 and PE2 regions. Microscopy image on the top right shows barcode beads after 16S extension with 16S primer FAM probes.

4.4.3 Microfluidic and molecular biology workflow

Since agarose is being using as the gelation agent in this workflow, all steps prior to gelation must be done at elevated temperatures. In particular, low-melting point temperature agarose remains liquid at 37 ºC. To do so, all relevant steps are performed in a modified oven incubator, described in the “Materials and Methods” section. After

gelation, microgels are distinct units that can be recovered from droplets and suspended in aqueous media (Figure 4.3a).

Cell lysis of the immobilized cells within the microgels is limited to enzymatic and chemical treatment since bead beating would break the microgels as well. To test the capability of cell lysis in microgels, we tested a broad cell lysis protocol on different cell cultures, both gram-positive and gram-negative and from different environments, encapsulated in microgels at a lambda of 100. For most strains – *E .coli*, *P. putida*, *B. subtilis*, and *B. thetaiotaomicron* - lysis was very efficient, with very little remaining intact cells left (Figure 4.3b). Staining these gels with a double-stranded DNA stain demonstrated that high-molecular weight genomes from lysed cells were retained in the microgel (Figure 4.3b). One strain of the five tested, *L. crispatus*, a gram-positive isolate, remained intact after the cell lysis protocol (Figure 4.2b). This demonstrates the need to understand the efficiency of lysis on the biological system of interest and optimize lysis protocols. Additional enzymes and chemical reagents could be used to target the specific physiology of *L. crispatus* (188), but this demonstration validates the proof-of-concept.

Droplet barcoding hinges on the pairing of the barcode beads with agarose microgels. The efficiency was determined and is presented in Figure 4.3c. There were four types of resulting droplets: empty droplets, droplets with one barcode bead, droplets with one agarose microgel, and droplets with both. The distribution of empty droplets and droplets with only agarose microgels do not result in any amplified sequences due to the lack of a forward primer from the barcode bead. However, sequencing inefficiency is introduced by droplets with just barcode beads; due to the

presence of the 16S standard, even without genomic template from the agarose microgels, there will be amplification of the 16S standard alone. Based on these encapsulation statistics, around 33-47% of droplets sequenced will be from informative droplet libraries, while the rest is from redundant 16S standard amplification.

The PCR amplification of the 16S V4 sequences in droplets is illustrated in Figure 4.3d. After exposure to UV light and high temperature to cleave oligos off the barcode bead and release microbial genomes from the agarose microgel, respectively, 16S amplification is performed in droplet, attaching barcodes to 16S amplicons. Droplets are merged and then subject to library preparation for Illumina sequencing.

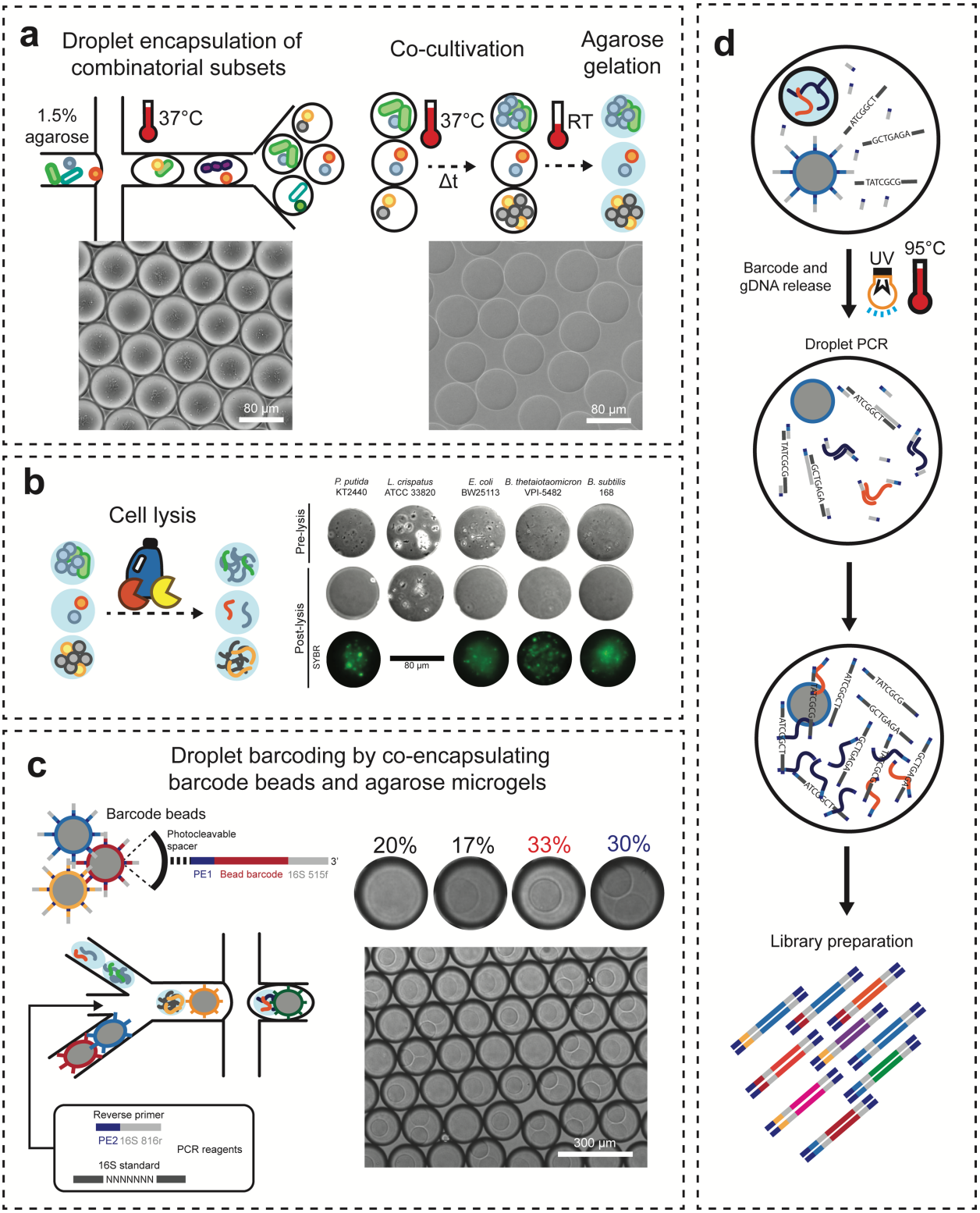

a
Droplet encapsulation of combinatorial subsets
1.5% agarose
37°C
80 µm
Co-cultivation
37°C
Δt
RT
Agarose gelation
80 µm
b
Cell lysis
P. putida KT2440
L. crispatus ATCC 33820
E. coli BW25113
B. thetaiotaomicron VPI-5482
B. subtilis 168
Pre-lysis
Post-lysis
SYBR
80 µm
c
Droplet barcoding by co-encapsulating barcode beads and agarose microgels
Barcode beads
Photocleavable spacer
PE1 Bead barcode 16S 515f 3'
20%
17%
33%
30%
Reverse primer
PE2 16S 816r
PCR reagents
16S standard
NNNNNNN
300 µm
d
ATCGGCT
GCTGAGA
TATCGCG
Barcode and gDNA release
UV
95°C
Droplet PCR
Library preparation

Figure 4.3: Detailed microfluidic and molecular workflow. (a) Droplets are generated with cells in a water and oil flow-focusing microfluidic device. The aqueous suspension is cells in a 1.5% low melting-point agarose suspension in the media or buffer of choice. To maintain the agarose as a liquid, the generation and co-cultivation environment is maintained at 37 C. After co-cultivation, droplets are incubated at room temperature or on ice for gelation. Photos show cells in droplets before gelation (left) and microgels after gelation suspended in an aqueous phase (right). (b) Cells are lysed in the microgels and nucleic acid is immobilized inside. To demonstrate, five different cell types were encapsulated in different gels, exposed to lysis and imaged before and after. Gels with lysed cells were stained with SYBR to stain double stranded genomic DNA. (c) Individual microgels are then barcoded by co-encapsulation with barcode beads. The carrier phase is a PCR reagent with the reverse primer and 16S standard for quantification. Microscopy images show examples of droplets from barcoding. Above the image are encapsulation statistics for empty droplets, droplets with just agarose microgels, droplets with just barcode beads, and both, respectively from left to right. (d) Molecular workflow for droplet PCR and attachment of barcodes to 16S amplicons and library preparation.

4.4.4 Benchmarking with mock communities

To assess the developed workflow, we applied it to defined mock communities. These mock communities were comprised of four cultured representatives comprising of *E. coli*, *B. subtilis*, *P. putida*, and *B. thetaiotaomicron*, which were verified to be effectively lysed by our workflow. There were two mock communities: (1) the “even” community with a λ of 50 cells of each species per droplet and (2) the “log” community with a λ of 180, 18, 1.8, and 0.2 cells per droplets, respectively (Figure 4.4). The “even” community is comprised of a somewhat equal portion of these different species in each droplet at a lambda of 50 cells of each type, although Poisson encapsulation statistics introduces quite a large amount of variability. The “log” community is much more uneven, being primarily dominated by *E. coli* with fewer members of *B. subtilis*, with the occasional *P. putida*, and even more rare *B. thetaiotamicron*. Due to random Poisson statistics, the exact composition of individual droplets is indeterminable, but the overall expected distribution can be used to compare with the distribution generated from sequencing the final libraries generated from the workflow.

Species	λ	
	Even	Log
E. coli	50	180
B. subtilis	50	18
P. putida	50	1.8
B. thetaiotaomicron	50	0.2

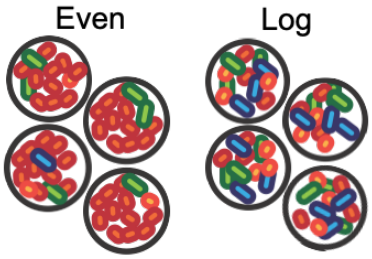


Figure 4.4: Mock communities utilized in benchmarking. Both are composed of the same four members, but in different community compositions. The "even" community has an approximately 1:1:1:1 ratio between the four members and the "log" community is expected to have E. coli *and* B. subtilis *in each,* P. putida *in a fraction, and* B. thetaiotaomicron *occasionally.*

Considerations for benchmarking included verifying the distribution of sequencing across different droplets after 16S amplification and sequencing and the accuracy of the community composition data. Based on the number of droplets generated and the proportion of droplets with the correct encapsulation of both the agarose microgel and barcode bead, we expected around 7000 droplets to be present in the sequencing dataset. We wanted to verify if approximately that number of droplets was represented in the sequencing or not. Additionally, the accuracy of the data needs to be validated as well. This includes checking if the distribution of community composition resembles the expected distribution from the mock communities. From our quantification of the 16S standard, we expected a λ of 50 unique 16S standards per droplet and would be compared with actual distributions after bioinformatic processing.

The sequencing depth across droplets from the "even" mock community is presented in Figure 4.5. After removing droplet libraries with less than 100 reads, there were 1325 droplets barcoded, less than expected by an approximate scale of five. There is a high level of variation in sequencing depth across droplets, with most

droplets being sequenced at depths of 100-300, with a long tail-end of certain samples that amplified to depths of tens of thousands. Although not shown here, a large proportion of droplets were removed when samples with less than 100 reads were removed. While 7000 droplets may have been processed, some droplets were amplified much more than others, resulting in a reduction of total droplets effectively sequenced. Although the barcoding is done in monodisperse droplets which are largely very uniform and identical in composition, the degree of variability of the number of functional oligos per barcode bead observed in Figure 4.2 may explain some of the variation between droplets.

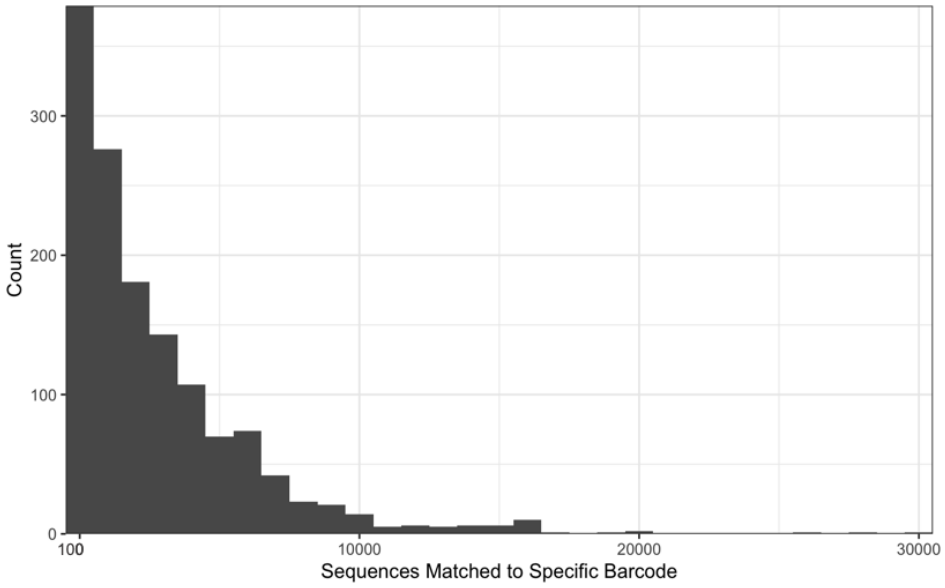


Figure 4.5: Histogram for the number of reads associated with each barcode in the "even" mock community after demultiplexing. Sequence from droplets with less than 100 sequences are not included.

Another aspect of the sequencing dataset that is nontrivial is the determination of sequences resulting from real barcoded droplets with the mock community versus droplets with just the 16S standard. A large portion of droplet libraries have reads from

the 16S standard contributing higher than 90% of the total reads. Droplets that only amplified the 16S standard and not genomic DNA comprise a significant part of this. However, many droplets have a substantial number of reads from the mock community and are still heavily dominated by the 16S standard. Some of this may be due to the uneven sequencing depth across droplets, where droplets with only the 16S standard are unevenly and heavily sequenced so that a significant portion of free mock community DNA is also sequenced. To proceed with analysis, I set a threshold for the determination of these standard-only droplets, as explained in the "Material and Methods" section of this chapter. In the future, more work should be done to characterize this distribution and what can be done to improve it for more rational screening. Using the number of unique 16S standards for each may help resolve the uneven sequencing depth to some extent but will likely not resolve it entirely.

To verify the general accuracy of the method, I analyzed the mock community portion of the sequenced droplet libraries. To represent the variation of community structures across and between the droplet mock communities, a principle component analysis with the Yue and Clayton (189) measure of community dissimilarity is presented in Figure 4.6. The expected distributions were generated *in silico* from the Poisson distribution and represented as well. A very stark difference is seen between "log" droplets and "even" droplets, with "log" droplets having a relatively low variation and "even" droplets having much higher variation. Variation within the "log" communities is expected to be much lower because of the smaller degrees of variation contributed by less abundant members, while the much higher relative variation contributed from each species in the "even" community results in a higher overall community variation. The

comparison of the expected distribution with the actual dataset shows relatively good fit. The “log” droplets match expectations almost precisely while the “even” droplets approximately match their expectations with a much wider distribution. This may be explained by deviations from the Poisson distribution during droplet encapsulation due to cell adhesion, which may be particularly pronounced in the “even” droplets whereas the “log” droplets are much more constrained due to the inherent tighter variation imposed by the highly uneven community profile, even with clustering of cells.

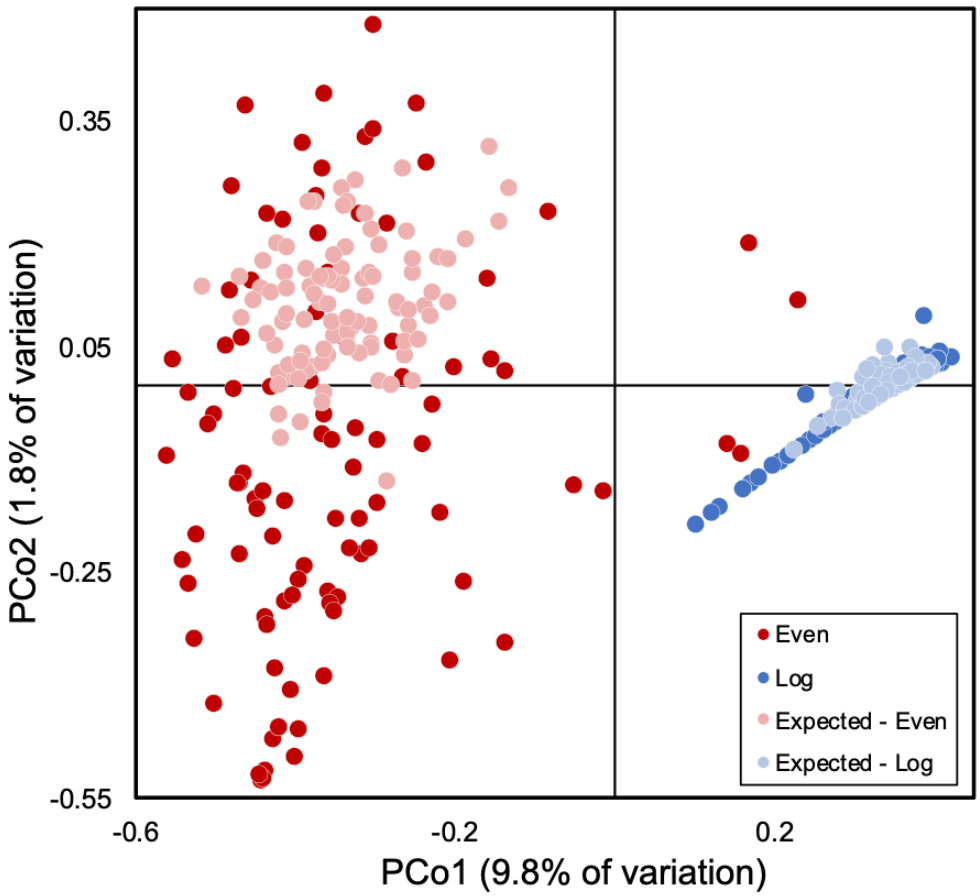


Figure 4.6: Comparing community compositions between the “even” and “log” mock droplet communities in a principal component analysis of Yue and Clayton distances. Expectations from the Poisson distribution are provided in lighter corresponding colors.

4.5 Discussion

This method demonstrates the first high-throughput co-cultivation platform for the study of microbial interactions for natural microbiomes. We have designed a workflow for the combinatorial generation, co-cultivation, and analysis of microbial sub-communities and benchmarked it on simple mock communities. The quantitative aspect of the method is still in the process of benchmarking, particularly the determination of unique 16S standards from the sequencing library, but the workflow is expected to require some more validation and development. Overall, despite the areas for improvement, the methodology demonstrated a significant capability for the high-throughput, compositional analysis of 4,071 droplets, from both the “even” and “log” mock communities. This is a significant technical advancement towards high-throughput system biology methods for the study of natural microbial communities.

4.5.1 Further benchmarking analysis

Due to time, the distribution of unique 16S standard molecules in each droplet was not determined, although the 16S standard is indeed present in each droplet. With the datasets already gathered, we will perform this analysis. Because some droplets were sequenced more deeply than others, we can perform rarefaction curves to analysis how much sequencing depth for each droplet is necessary to recover all unique 16S standards and the full diversity of the community. This may allow us to set thresholds for the minimum number of reads per each droplet community more systematically. Mock communities are certainty not just for methodology benchmarking, but also for determination of these thresholds and parameters for quality control, just as mock communities are routinely sequenced with traditional 16S sequencing.

One issue with using droplet mock communities is the high variation from Poisson statistics. A precise assessment of the accuracy and precision of the method is difficult to assess with the “even” mock community but is more apparent with the “log” community where variability is lower. The “log” community libraries seem to suggest that the method is quite accurate and precise, even for detecting rare members, but more resolved benchmarking is needed. The most ideal benchmark standard is with the actual measurement of droplets, instead of being based on expectations from distributions. The next benchmarking experiments are to grow a model bi-culture with a known interaction in the droplets, measure the growth of each species per each droplet with a reliable method, perform our developed workflow in parallel, and then compare the sequencing signal with the expectations from the reliable measurement. We plan to do this with a *B. subtilis* and *E. coli* auxotrophic, syntrophic co-culture from Hsu et al. (94). Because each strain is fluorescently labelled, fluorescence imaging will provide more accurate expectations from measurement.

4.5.2 Further technical improvements

This demonstration is a technical proof-of-concept demonstration. The mock community studied has four members, which pale in comparison to most natural communities. Additionally, to conserve sequencing effort, Illumina paired end sequencing only partially sequenced the V4 region, specifically the first 160 base pairs. it was not a very deep sequencing effort, utilizing the NextSeq P1 chemistry that provides around 100 million reads. For benchmarking with our simple four-member mock community, this resolution was adequate. However, for the phylogenetic resolution of a natural community, the entire V4 region is recommended, and much

deeper Illumina paired end sequencing efforts are available, such as the NovaSeq 6000 SP at 800 million reads.

The variability in droplet sequencing depth has been measured and some improvements to the workflow can be made to ameliorate it. This bias is most likely from the variability in functionalized barcode beads. One potential strategy to reduce the variability is to hybridize fluorescent probes to the functionalized oligonucleotides, sort barcode beads based on the degree of functionalization with flow cytometry, and only use beads within the top 10% of functionalization for barcoding. Alternatively, if this cannot be amended, the variation in sequencing depth can be accounted for in future efforts through appropriately deeper sequencing efforts.

Another consideration is the disparity in amplification between the relatively low molecular weight 16S standard and 16S sequence within larger genomic DNA fragments. Thermodynamically, it may be more difficult for the annealing of primers onto the 16S sequence within larger genomic DNA than on shorter 16S standards. The PCR efficiency should be compared between the 16S standard and in genomic DNA, as the absolute quantification only works if there is a similar efficiency between the two. If not, this efficiency can be measured and accounted for.

As explained before, the device used to pair barcode beads to the agarose microgels was not designed with our workflow in mind, leading to inefficient pairing and wasted sequencing. Other designs that better accommodate the flow of an agarose gel suspension can inform our design of new devices (190). Improving this aspect will limit the number of droplets with excessive signal from the 16S standard and will improve the interpretability of data.

4.5.3 Generalizability

The current scheme works given the assumption that the community is agarose-agnostic, meaning the community does not consume agarose or that agarose does not affect the metabolic state of the community. This is certainly not the case in aquatic or marine communities in which agarose is a naturally occurring carbon source. In addition, to maintain planktonic growth, the community must grow in temperatures above 37 ºC. Otherwise, the agarose will set, immobilize cells prematurely, and constrain cell growth. The nature of interactions between cells may change if they are not allowed to physically-interact. Ultra-low melting point agarose is an alternative hydrogel since it can remain liquid at room temperature or 30 ºC. If agarose affects cell physiology, an alternative is high-throughput picoinjection of agarose or acrylamide into individual aqueous droplets after co-cultivation (191). Lastly, as demonstrated in Aim 1, droplet size must be considered for effective growth of certain biological systems. Due to this, the droplet sizes utilized in this proof-of-concept method may not be generalizable to all biological systems. Droplet dimensions may need to be adjusted depending on the system of study and which questions are being investigated.

4.5.4 Future applications

One biological system that we are excited to apply this method to with the fewest alterations to the technical workflow is the vaginal microbiome. The system is expected to be agarose-agnostic and can be co-cultivated at 37 ºC, which is compatible with our current workflow. A previous member of our lab has also applied microdroplets of similar sizes to the co-cultivation of isolates from the vaginal microbiome (102). The vaginal microbiome is a relatively well-characterized system with many interesting

biological questions (192,193). For example, we know about the vaginal microbiome's high prevalence of negative interaction. *Lactobacillus* species are known to inhibit the growth of other anaerobic bacteria by keeping the vaginal environment acidic through the excretion of lactic acid (12). However, *Lactobacillus* species are also known to produce other bactericidal compounds (194), and the specificities of those agents is unknown. With our high-throughput interaction inference method, we can co-cultivate various *Lactobacillus* cells from vaginal samples with isolates of known problematic vaginal isolates such as *Gardnerella*, *Prevotella*, and *Mobiluncus* to characterize negative pairwise interactions in a much more resolved and high-throughput manner.

The study of more complex communities such as the human gut microbiome with this method is possible if the throughput is scalable to the diversity of the system. Additionally, with larger, more comprehensive datasets comes the potential to apply machine learning, as some research groups have (88), which may help us in understanding how these cell-cell interactions contribute to higher-order emergent properties.

Chapter 5 : Concluding Remarks and Perspectives

This chapter will summarize the work in this dissertation, perspectives on future work, and reflections regarding parallels between phenomena observed in microbial communities and the scientific community studying them.

5.1 Summaries of Completed Work

5.1.1 Characterizing the Effect of Droplet Size on Syntrophic Dynamics

Using a two-member, syntrophic *E. coli* bi-culture and different droplet sizes under different environmental contexts provided much insight regarding the effect of droplet size on syntrophic co-growth. The specificity of the response to the change in droplet size between the different environmental contexts was particularly stunning. In the lowest interaction scenario tested, increasing the droplet size resulted in a prolonged lag time with less significant effects on the apparent maximum specific growth rate. In the intermediate interaction scenario, increasing the droplet size resulted in a more dramatic effect on the apparent maximum specific growth rate and an insignificant rate on the lag time. In the most strenuous interaction condition, a very drastic binary "establishment or not" effect was seen. These observations inspired investigations into the mechanism for these phenomena. However, even with this simplified bi-culture system, the mechanisms are not intuitive. These observations can be replicated to a limited extent in an ordinary differential equation model consisting of

two-members growing in a population-dependent manner according to Monod kinetics on limited substrates produced by the other member. Because this model assumes immediate mass transfer, these observations are not a result of diffusion but more likely due to nonlinear responses to changes in starting parameters which are a function of droplet size. The behavior of the system is quite complex, and a more systematic numerical analysis to characterize the system response would be helpful. For example, in inspiration from fluid mechanics and chemical engineering, dimensionless numbers provide a metric to characterize to what extent opposing factors are contributing the system behavior. Studying how the availability of specific substrates as a composite of secretion and uptake compares with the biosynthetic requirement for growth is one potential. Additionally, to verify findings from our numerical analyses, it is important to also verify their relevance experimentally, which would require measurement of these parameters.

Additionally, the staggering amount of droplet-to-droplet variation confirms the degree of cell stochasticity introduced from low initial cell number populations. The high variation observed in the bi-cultures compared to in monoculture conditions suggests the sensitivity of the bi-culture to initial parameters. This sensitivity to initial conditions is a non-linear behavior known as chaos. This chaotic behavior in co-culture conditions has been utilized to select for high secretion strains in co-culture assays, manifesting as a signal of much higher total co-culture growth (100). This phenomenon has been studied extensively in physics and is characteristic of ecological phenomenon such as founder effects. Utilizing droplets is certainly an exciting way to quantify these characteristics in microbial systems.

Despite their growing popularity, the droplet as a cultivation environment has still not been fully investigated. Some anecdotal evidence claims that some strains behave differently in microdroplets than in bulk conditions. There probably remains other factors that have not been elucidated yet. For example, one major difference is the lack of convection on fluid flow. While mass transport may be similar due to the length scales, cells may respond differently to different flow regimes. Bulk cultures are subject to convective flow for mixing, and cells have been observed to react differently to different kinds of flow (195,196).

5.1.2 High-resolution Metagenomic Dissection on Selected Microdroplets to Study Microbial "Dark Matter"

We successfully applied singlet droplet manipulation, collection, and individual droplet shot-gun metagenomics to recover genomes *de novo.* While other studies have applied this to single cell cultivation, to our knowledge, this was the first application to the recovery of genomes from a community grown in a single droplet. The experimental design and limited number of droplets studied do not allow for interaction inference, but it is likely that the organisms that were grown may have benefited from certain undefined interactions. The genomes represented uncharacterized diversity and demonstrated that microdroplets are indeed a method for the cultivation of rare strains that may be outcompeted by cultivation in bulk cultures.

Assuming that sequencing effort is limited, it is difficult to recover all genomes from the sub-community. While the recovery rate of genomes demonstrated here is still quite impressive compared to single-cell genomics, multiple displacement amplification (MDA) introduces extensive bias in bulk reactions that make the reconstruction of entire

communities extremely difficult. Advances such as digital droplet MDA (ddMDA) may allow for the unbiased amplification of genomic material from single droplets and recovery of more genomes from a limited sequencing effort. Still, metagenomic efforts of communities with this method will require significant sequencing effort and will be limited to a smaller number (10-100s) of sub-communities. For natural communities of exemplary high diversity where recovery of genomes through conventional means may not be feasible, reducing the complexity through this methodology to identify interactions may be more informative. This could be performed with more targeted approaches, such as the selection of droplets with species of interest by flow cytometry or droplet sorting (104). For example, there is interest in identifying functional consortia for plastic-degradation or carbohydrate degradation and their interactions. This could be done by the selection of droplets with a positive signal (e.g. detection of certain fluorescent metabolites or the growth of fluorescent reporter strains) and processing of individual droplets with this methodology.

We are currently applying the metagenomic approach with ddMDA on single aggregates from the Lake Erie cyanobacterial harmful algal blooms. Single aggregates provide an ecologically meaningful unit to study due to their self-aggregation in nature. We are using this approach to study interactions between *Microcystis aeruginosa* hosts and their microbiomes through genomic inferences and metabolic complementarity.

5.1.3 Droplet-based, Quantitative 16S Profiling for High-throughput Combinatorial Decomposition

Chapter 5 demonstrated the design of a quantitative droplet-based 16S profiling workflow. The lack of this methodology has limited the application of microfluidic

droplets for the high-throughput interaction inferencing of microbiomes. Instead of using metagenomic shot-gun sequencing on a smaller pool of droplets, this approach is high-throughput, targeted sequencing. Utilizing polyacrylamide barcode beads to retain barcode signal in sequencing and a 16S standard to quantify community composition were the main technical capabilities incorporated.

The methodology's ability to study the relative abundance of different droplet communities has been demonstrated well with mock communities. We plan to benchmarking the accuracy and precision of the workflow by incorporating co-growth and more precise validation with fluorescence signals. Analysis of the mock communities by absolute quantification has yet to be done and will determine how much more prototyping is necessary for accurate quantification. Further work to generalize the method for other biological systems will require accounting for different droplet sizes and microfluidic picoinjection.

Additionally, we have realized the utility of mock community droplet standards for identifying thresholds for eliminating sequencing noise and sequencing coverage required to capture the entire community profile. The addition of a mock community droplet library standards should be standard protocol for determining sequencing efficiency, distribution across droplet libraries, and error.

5.2 Future Work

5.2.1 Study of single-cell level variation on interactions and community outcomes

Chapter 2 demonstrated how microdroplet cultivation can elucidate high-levels of phenotypic resolution with simple mono- or co-cultures. Our study of a model *E. coli* autoxtroph system demonstrated the spectrum of community states that result from

supposed single-cell variability and random Poisson encapsulation statistics but is unable to decouple the two. One potential research opportunity is how variability at the single-cell level contributes to different community structure outcomes. The inoculation of an open-niche, localized environment by a small number of establishing organisms may have chaotic implications for the succession principles for a community outcomes (197). To do so requires the ability to study the initial and final compositions of the same microdroplets and visual reporters for strain phenotype. Currently, with our technical capability it is difficult to store or track droplets over time and fluorescence reporters are constitutively expressed. Adapting microfluidic devices for the storage of individual microdroplets (198) and designing strains with genetic circuitry to link fluorescent reporters with the biosynthesis of cross-fed metabolites may allow for the higher resolution characterization of how strains and their phenotypic properties contribute to their interactions and community outcomes.

5.2.2 Beyond prokaryotic cells

While this dissertation discussed utilizing microdroplets for the study of interactions within microbiomes, all work was exclusively demonstrated or applied to prokaryotic systems. Microbiomes include fungi, viruses, and protists as well. However, the application of microdroplet technologies to these kingdoms of life will require technical advancements and the design of unique microfluidic and molecular biology workflows. Specifically, if the high-throughput quantitative droplet profiling approach in Chapter 4 is to be applied in this context, it will require the inclusion of other marker genes for multiplexed amplification and sequencing. More specifically, barcode beads would incorporate not just primers for the 16S gene but other marker genes including

the internal transcribed spacer (ITS) region or 18S for fungi and protists and other phylogeny-specific genes. This kind of approach is necessary for the study of microbiomes associated with different eukaryotic phytoplankton (199) as well as with the study of fungal-bacterial interactions for degradation of lignocellulose (200).

5.2.3 Beyond genomics

The work presented in this thesis is primarily genomics-driven. While metagenomic-approaches have greatly advanced our knowledge of natural systems, 'omics approaches also include transcriptomics, metabolomics, and proteomics. There are certainly new advances in single-cell bacterial transcriptomics (201). The application of single-cell eukaryotic transcriptomic workflows to prokaryotic systems was made possible by utilizing polyadenylation enzymes for the addition of polyA tails onto bacterial mRNA transcripts (202). From this, single cell bacterial transcriptomics have been developed but applying them in a low-biomass, community, or high-throughput droplet context will be difficult. Regarding metabolomics, high-throughput droplet workflows have been developed (203), and one study has applied Raman spectroscopy for the identification of pure cultures grown in droplets that produce certain natural products (204). A critical challenge is combining metagenomic, transcriptomic, and metabolomic workflows. This requires methods to divide individual droplets for analyses with different high-throughput droplet platforms and the ability to link metagenomic and transcriptomic signals to corresponding spectroscopy and chromatography analyses. Even less work has been done to incorporate proteomics for single-cell or water-in-oil microdroplet systems due to the amount of sample needed for detection and the precise and intensive sample processing required. Sequencing-based workflows, such as

metagenomics and transcriptomics, are certainty the easiest to adapt into droplet-based methodologies. However, metabolomic approaches are needed for more resolved phenotypic information.

5.3 The Scientific Community and Microbial Communities

Throughout my PhD, I honed my skills as a methodologist. Many of the questions that I asked were along the lines of "what can I do with this technology that I am developing?" rather than those of a traditional scientist who desire to answer questions to deeply understand a system of interest. There are benefits in being a methodologist. For example, scientists are limited to pursuing the questions that can be answered rather than which questions are important to answer because of the limitations of available methods. Because I am in disciplines that support methodology development, I can spend the time to do so. However, the amount of time to answer questions is very limited. Scientific disciplines would rather train their students to be competent scientists rather than methodologists. As such, doing both is extremely difficult. In this regard, I have learned that successful science in microbial ecology is performed by a highly mutualistic community. Just as microbial communities form very intimate mutualistic interactions, we should as well! The degree of the analogous nature of how interactions between cells in microbial communities apply to the scientific community is quite extensive.

5.3.1 Division of labor with frequent cross-feeding and communication

Divisions will naturally emerge in science because people are limited in their capacity, and this is not a bad thing. Differences in scientific approach have generated

the spectrum of top-down and bottom-up investigations explained in Chapter 1. Another division has emerged between individuals who are adept at studying the detailed mechanisms (e.g. a molecular biologist for a specific disease) and others who are able to make generalizations from a broader study of many systems (e.g. ecosystem ecologists and their comparison of different biomes). The previously mentioned disparity between methodologically-orientated scientists and inquiry-centered scientists is the same. One of the strongest disparities is the one between experimental scientists and computational/modeling scientists. Like how microbial communities have a strong division of labor, our scientific community does as well! However, at least from my empirical observations, the lack of progress to answer critical questions most likely results from the lack of communication and collaboration between these groups. I have been quite fortunate to be part of institutions that realize this and attempt to promote collaboration by establishing interdisciplinary centers. Examples for the microbiome sciences include the University of Michigan's Integrated Training for Microbial Systems and the Ohio State University's Center for Microbiome Sciences. However, I would argue that the change must be in our individual mentalities, rather than structural changes. Being able to communicate interdisciplinary requires a willingness to "translate" when certain semantics do not work across disciplinary lines, and even inform ourselves about the fundamental principles of something entirely new to us so we can discuss.

The biggest barrier is the acknowledgement that interdisciplinary work may require extensive commitment. Developing sustainable collaborations and partnerships will not always result in a temporary win on the individual level. Usually, it requires the

intensive investment of resources (time, money, personnel) to explore what the collaboration might even provide. Starting to see the payoff from the collaboration takes at least a couple of years in my experience. Sadly, I believe the structuring of principal investigators in academia and funding timelines do not encourage this long-term endeavor. Despite this, the community must grow to have a sense of commitment to collaborative endeavors and especially to one another. When principal investigators are committed beyond convenience, there is more extensive communication and commitment between research groups of the principal investigators as well. As someone doing methodological work, other research groups and principal investigators tend to turn aside when the tool does not address the questions that they are interested in. Alternatively, they assume that methodologies developed are "one fit all" solutions to all systems without realizing that much development work needs to be done to tailor a methodology to the needs of a specific system or question. Reciprocally, I have noticed my own dismissal of other's work when my methodology is not applicable either. Strong collaborations need a sense of flexibility as well, knowing that methods need time for development to tailor workflows to specific questions. High-reward, risky research endeavors like this can be difficult to fund, but seed funding grants that foster collaboration and innovation are necessary as well.

5.3.2 Horizontal gene transfer for greater expertise training and reproducibility

One of the major issues in the adoption of microfluidics is the expertise needed to do it. Numerous times when presenting my work, principal investigators and researchers want to implement this technology in their own laboratories. However, if a laboratory doesn't have some embodying the expertise, they simply cannot implement

it. The advances in microfluidic analyses for microbial ecology are impressive but despite the impact factor of the journal these advances are published in, they are only impactful when reproducible by other research groups. Research equipment companies are realizing this deficiency and seek to provide services and goods, but the cost can be exorbitantly high (as discovered by myself numerous times when performing this dissertation work). Thankfully, most major research institutions have at least one research group specializing in micro- or nanofluidics. Additionally, trainees such as myself hold the responsibility to train others in other research laboratories. In this way, our expertise is exchanged and passed along the entire community, similarly to how a plasmid encoding an antibiotic-resistant gene is passed by horizontal gene transfer throughout the microbiome. While bacteria rely on phage for transduction or pili for conjugation, the scientific community relies on the exchange of trained researchers who can provide training to others who will also then provide training to others.

5.3.3 Functional diversity and niche differentiation

One tension within the scientific community results from individuals trying to answer the same, critical questions. As a result, we are going to overlap in terms of our expertise and approach. Competitive exclusive suggests to us that we will inevitably outcompete one another if we occupy the same niche. I have experienced this personally when talking with another expert in the field doing similar work to myself. I found myself getting fearful upon hearing that a similar methodology in Chapter 3 was trying to be developed by their laboratory. The first tendency is keep details hidden. And while that may need to be the case to a certain extent, even in these relationships, we need to be transparent with one another to find the opportunities for niche

differentiation. Even when we study the same questions, we approach them from different perspectives with different motivations. Instead of getting defensive, we should acknowledge and embrace our diversity in scientific approaches and look for ways for collaboration rather than competition. Certainly, one of the greatest surprises in microbial community research is the degree of coexistence amongst functionally similar bacteria and how they differentiate along numerous niche dimensions (205). Functional redundancy can certainly also exist and even bolster and stabilize our research community.

5.3.4 Conclusion

Personally, I believe the future for microbial ecology and the community that studies it is quite exciting. While the approaches to address them may not be intuitive, the challenges moving forward are clear. The ability for us to understand microbiomes is not limited only by funding and technological advances, but how well we interact within our own research community. May the communities we study challenge ourselves to be in intimate, diverse, nuanced, and communicative communities as well.

Appendix A. Supporting Information for Chapter 2

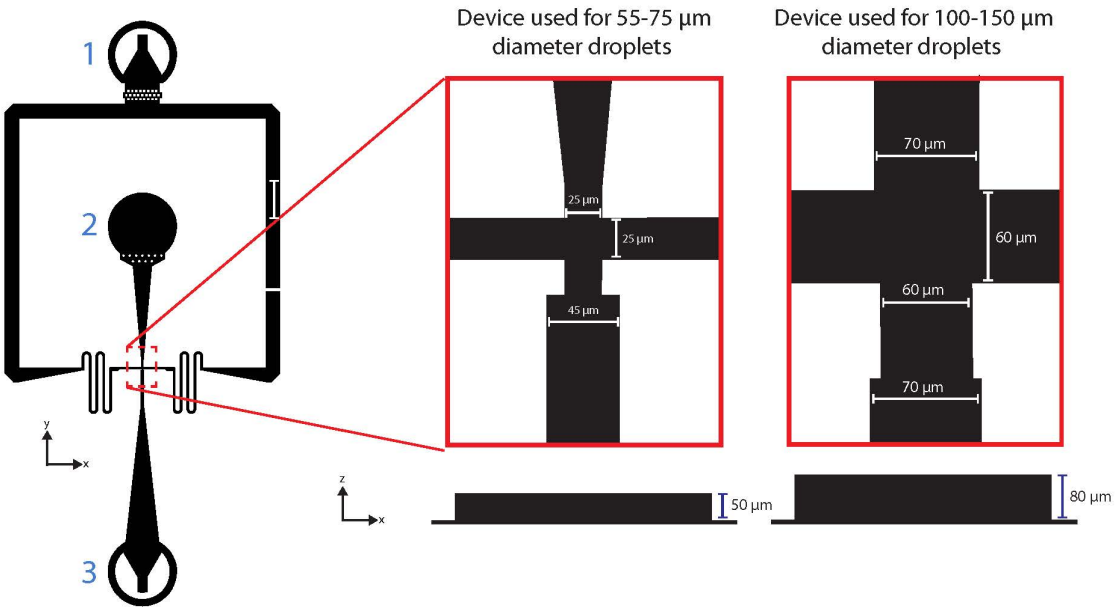


Figure A.1: Schematic layout of the SU-8 molds used to make the flow-focusing microfluidic PDMS devices used for droplet generation. There were two devices utilized in the study which had similar design features but had different channel dimensions. All devices had an oil inlet (1), cell suspension inlet (2), flow-focusing channel intersection (boxed in red), and a droplet outlet (3). The first device mold (left) had channel widths of 25 µm and a channel height (blue scale bar) of 50 µm, while the second device mold (right) had channel widths of 60 or 70 µm and a channel height (blue scale bar) of 80 µm. The first device generated 55 µm and 75 µm diameter droplets, while the second device was used to generate 100, 125, and 150 µm diameter droplets.

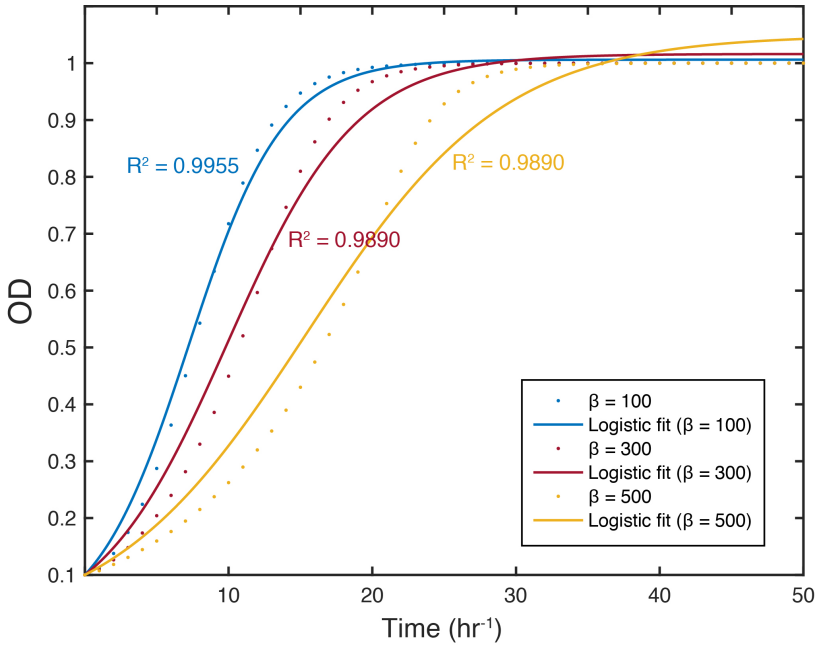


Figure A.2: Comparison between simulated growth dynamics of an auxotroph participating in mutual cross-feeding with another auxotroph and the profiles fitted using the logistic equation. The cellular requirement for the cross-fed metabolite is represented by β and was adjusted from 100 to 300 and lastly to 500 µg/L-OD while all other parameters of the model were kept constant. The simulated growth curves of the auxotroph are plotted in dotted lines, while the respective logistic fits are provided in solid ones. The R^2 value indicative of the quality of the fit is also provided for each fit.

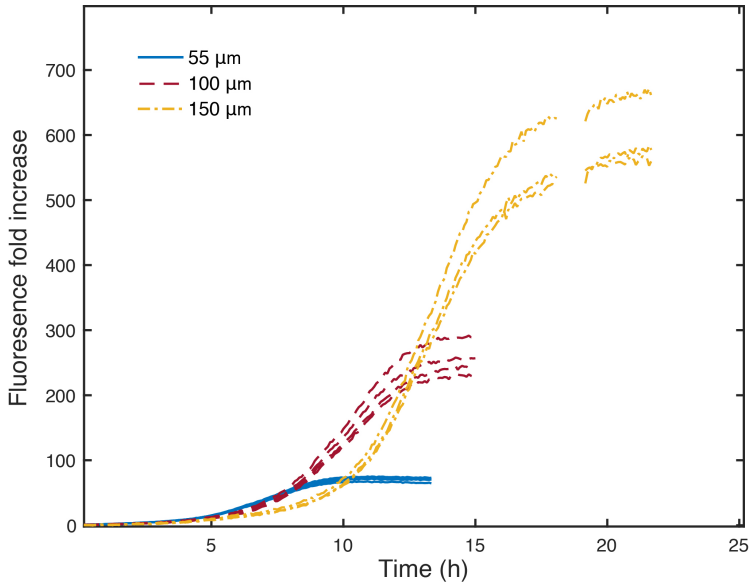


Figure A.3: The full set of growth curves for the monoculture microdroplet cultivation of S1 ΔilvD. The initial λ was 5 cells/droplet. Each curve is the aggregate growth of droplets within a single well in a 96-well plate, representing a large population of droplets. Fluorescence fold increase is the fluorescence at a time

point normalized by the initial fluorescence of the sample well. Each condition had 4 replicates, with 3 for 150 μm diameter droplets due to one replicate having inaccurate initial measurements.

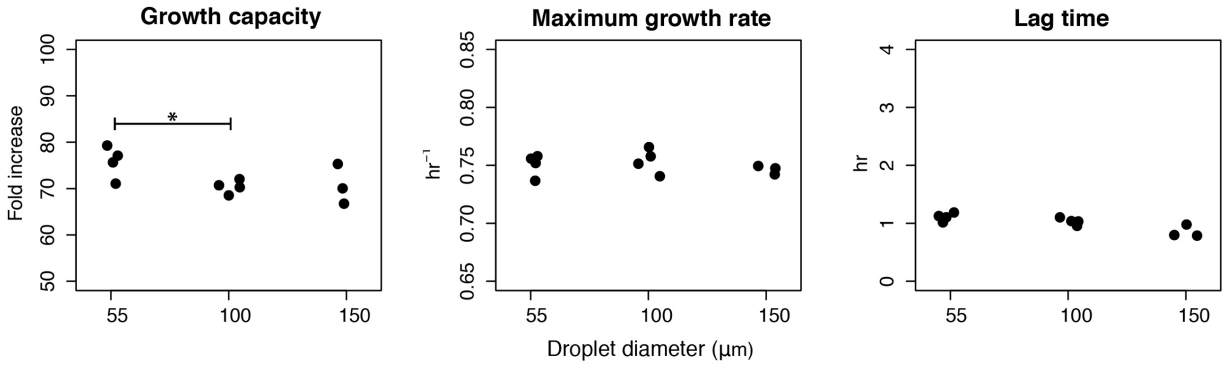


*Figure A.4: Fig S4. Growth model parameters estimated from fluorescence data of S1 ΔilvD monoculture grown with the same initial cell density in droplets of different sizes. The average initial cell number (i.e. λ parameter of the Poisson distribution) was 5, 30, and 100 cells/droplet in 55, 100, and 150 μm diameter droplets, respectively. * indicates p-value < 0.05 for statistical significance.*

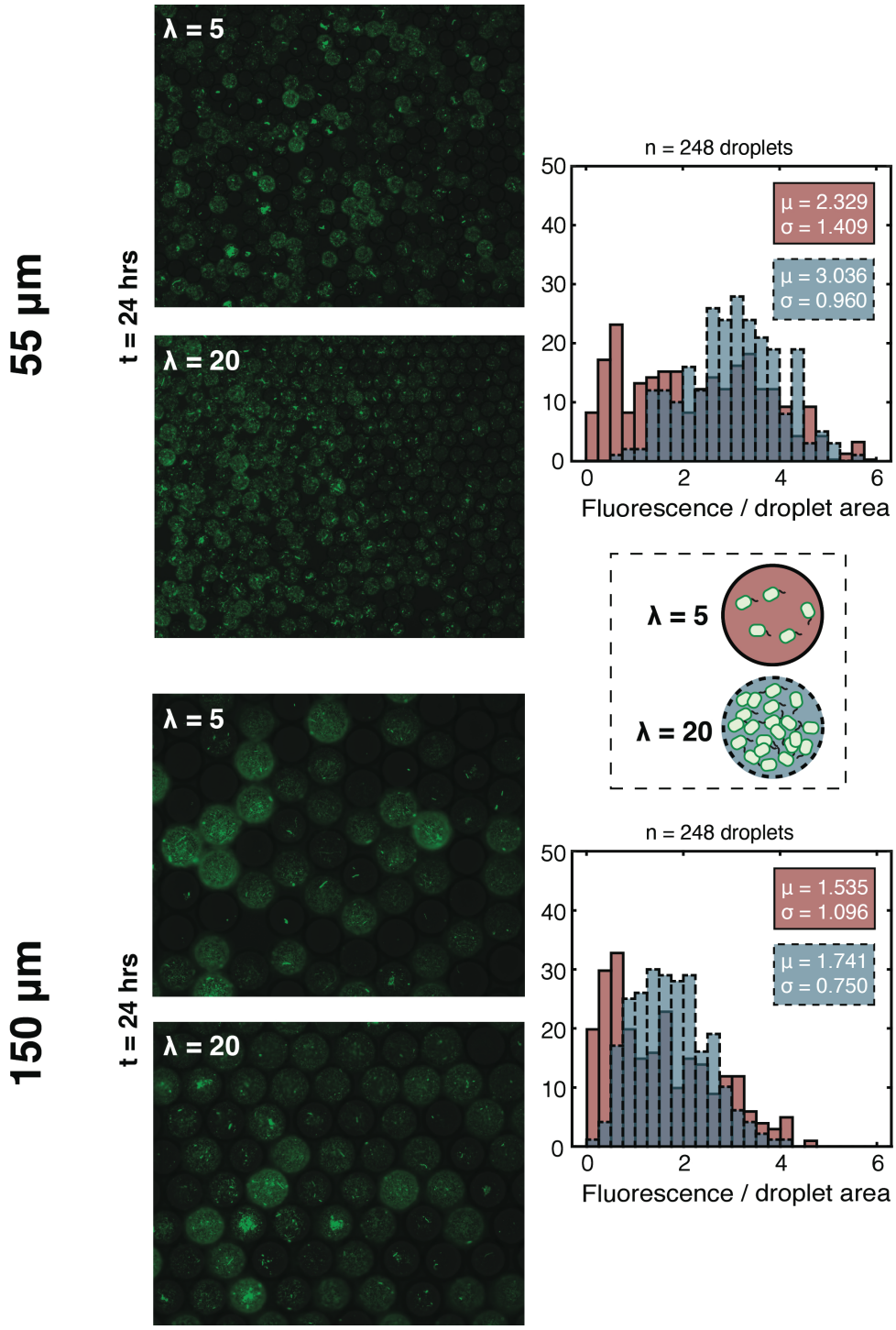


Figure A.5: Droplet-to-droplet variance of fluorescence for the cultivation of S1 ΔilvD with initial λ=5 and λ=20 cells/droplet in droplets with diameters of 55 and 150 µm. Droplet-to-droplet variation is observed in representative images of populations of droplets after cultivation for 24 hours under both conditions. Image analysis of a large population of droplets (248 droplets for each) was performed to quantify the degree of droplet-to-droplet fluorescence variation through histograms with associated statistics (mean and standard deviation). The distribution and statistics of the λ=5 cells/droplet condition is magenta with a solid boundary. The distribution and statistics of the λ=20 cells/droplet condition is blue with a dashed boundary.

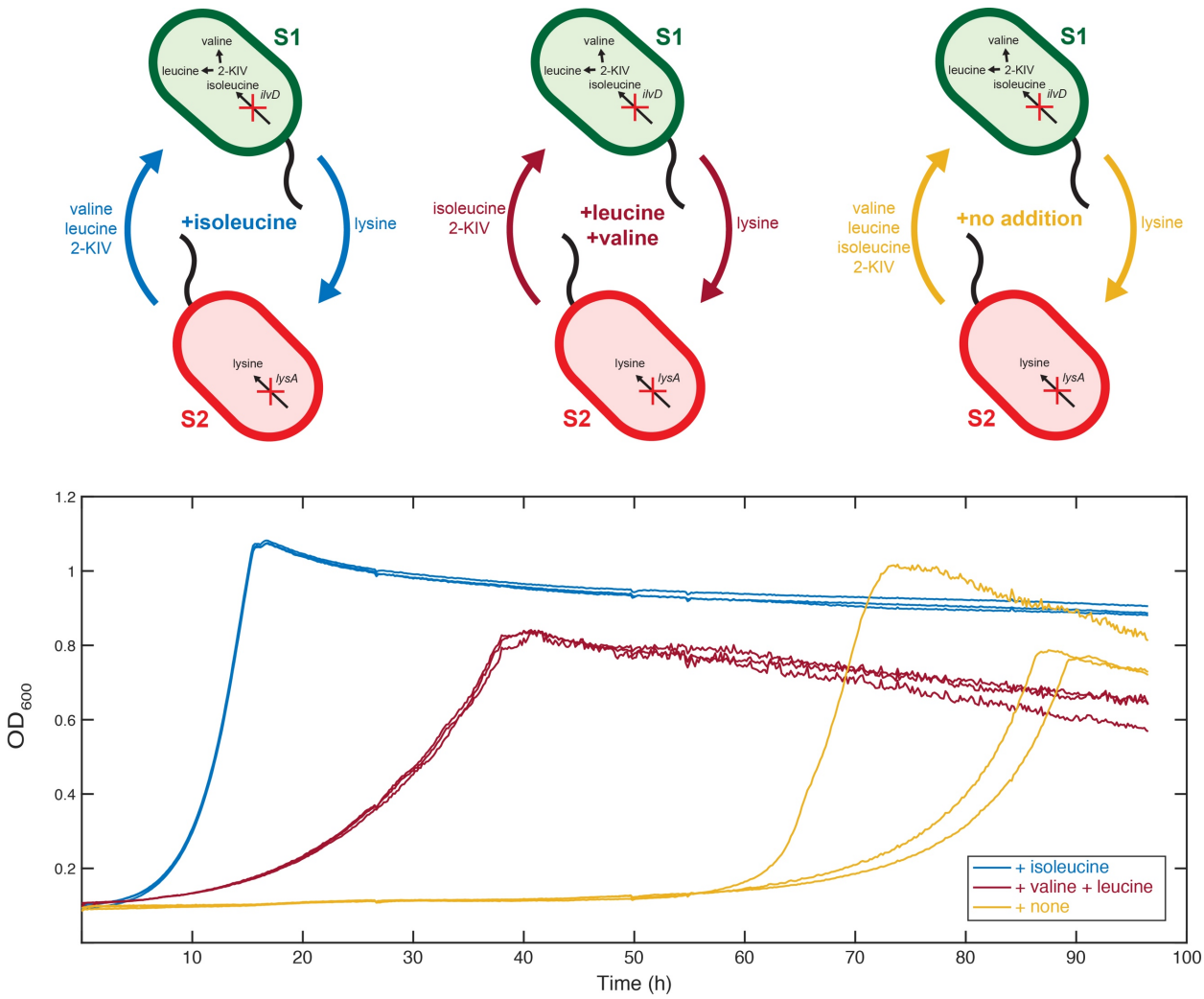


Figure A.6: Growth of co-cultures of S1 ΔilvD and S2 ΔlysA in bulk. Cultivation was done in microwell plates under three different amino acid supplementation conditions to modulate the degree of interaction between the two auxotrophic partners: (1) with 3 mM isoleucine, (2) with 3 mM leucine and 3 mM valine, and (3) with no addition of amino acids. Each condition had three replicates.

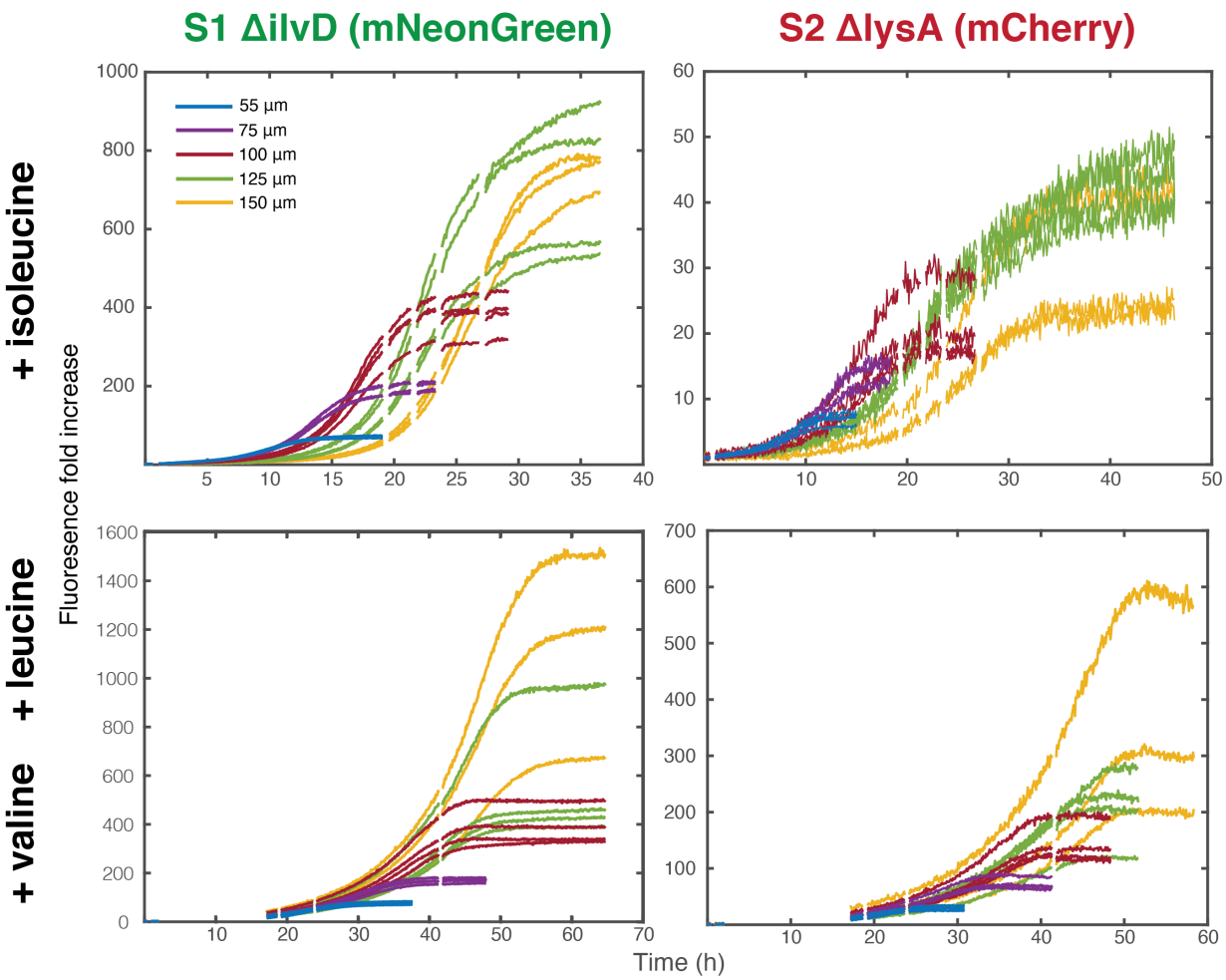


Figure A.7: The full set of growth curves from co-cultivation of S1 ΔilvD and S2 ΔlysA in droplets. The initial λ value was 5 cells/droplet of each strain under the two amino acid supplementation conditions. Each curve is the aggregate growth of droplets within a single well in a 96-well plate, representing a large population of droplets. Fluorescence fold increase is the fluorescence at a time point normalized by the initial fluorescence in the same well. Each condition had 4 replicates, with 3 for 150 µm diameter droplets due to one replicate having inaccurate initial measurements.

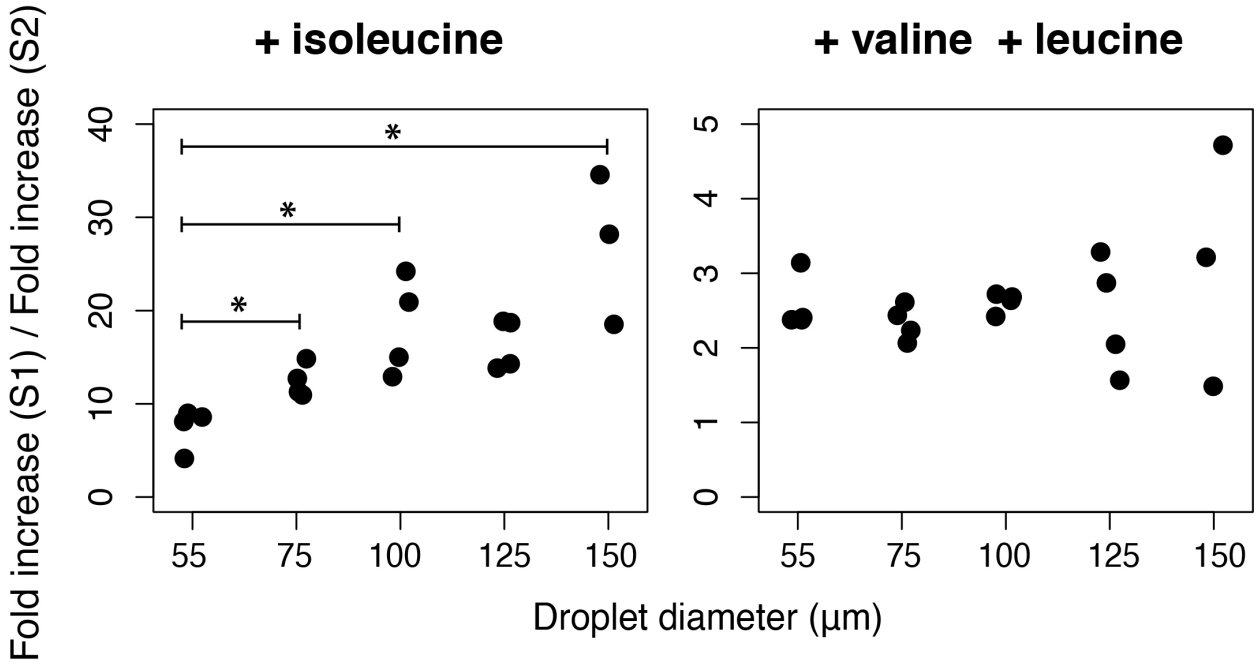


Figure A.8: The effect of droplet size on community composition between S1 ΔilvD and S2 ΔlysA. Community composition is provided as the ratio between fold increases of S1 ΔilvD and S2 ΔlysA in droplets of different sizes when 3 mM isoleucine or 3 mM valine and 3 mM leucine is supplemented. For each well in the co-cultivation experiments, a ratio between the fold increase of S1 and that of S2 was calculated. Statistical significance is defined by a p-value of less than 0.05 (*).

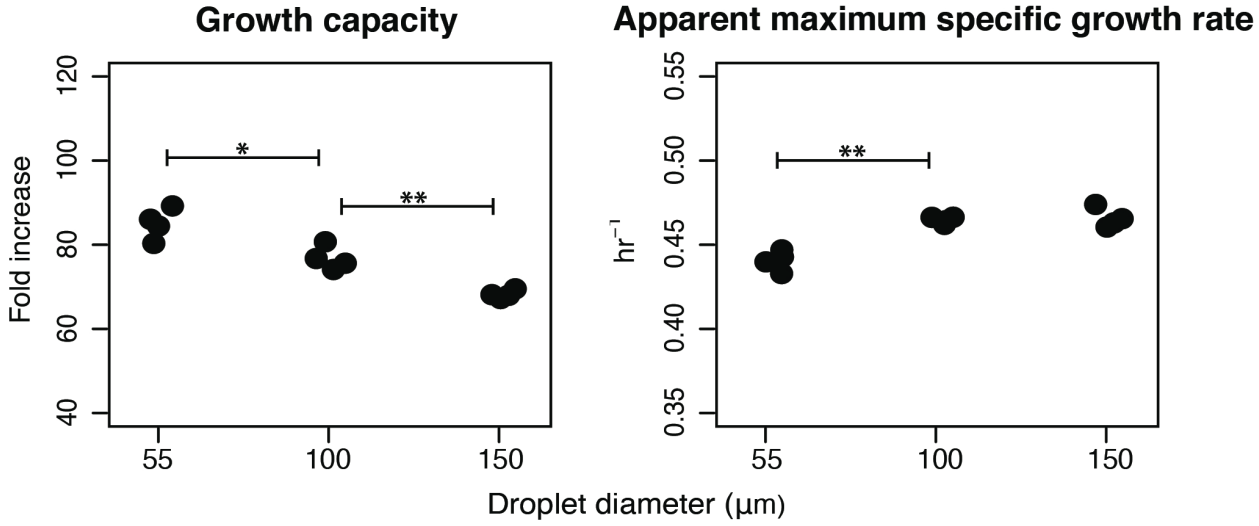


Figure A.9: Growth model parameters estimated from fluorescence data of S1 ΔilvD grown in co-culture with S2 ΔlysA with the same initial cell density in droplets of different sizes. The average initial cell number (i.e. λ parameter of the Poisson distribution) was 5, 30, and 100 cells/droplet for each strain in 55,

100, and 150 μm diameter droplets, respectively. Lag time was not evaluated due to there being an insignificant lag time observed. Statistical significance is defined by a p-value of less than 0.05 (*) and a p-value of less than 0.01 (**).

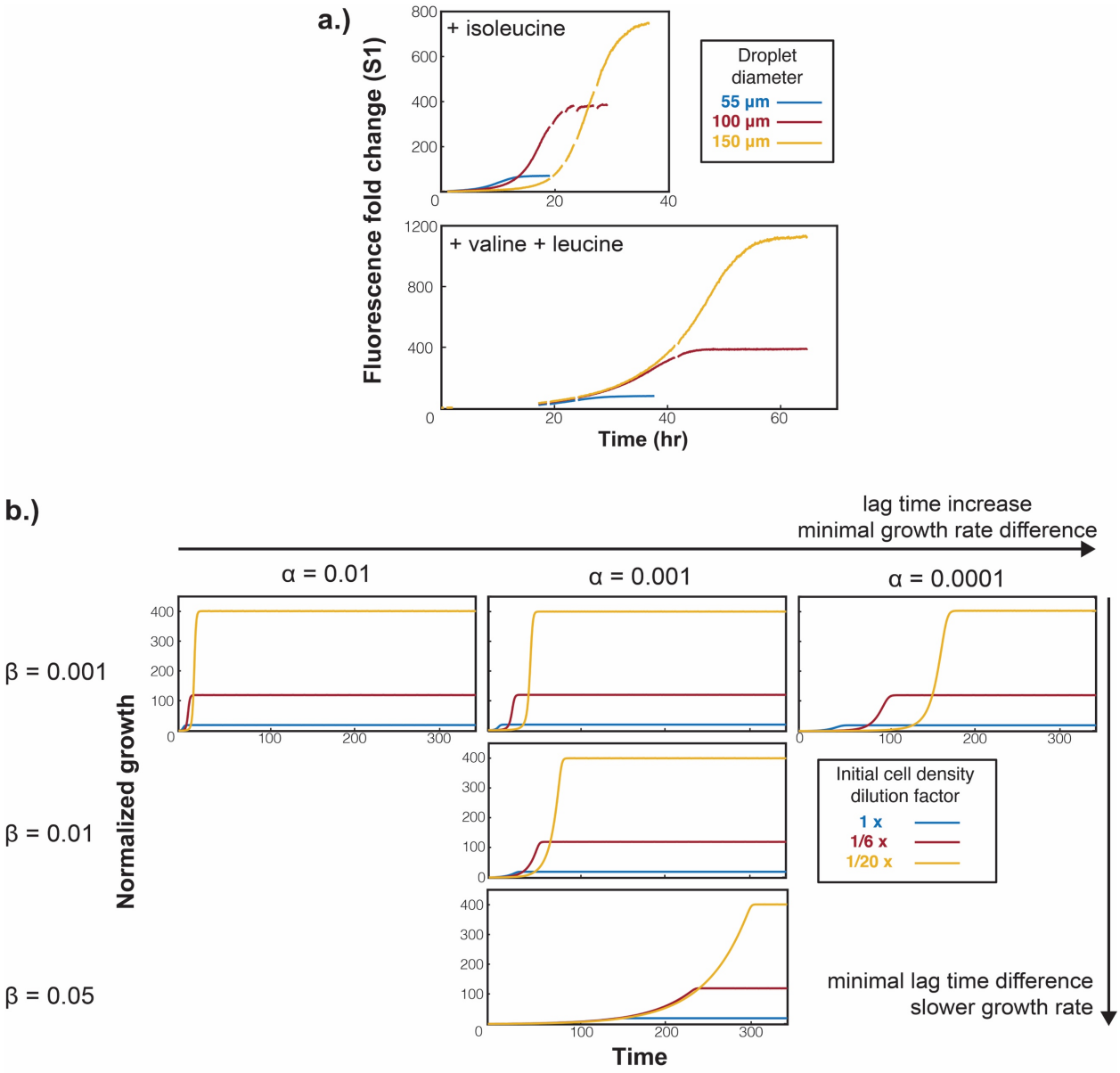


Figure A.10: Different parameter values of a mathematical model lead to different patterns of growth dynamics, recapitulating in part experimental observations. (a) Experimental data for S1 ΔilvD in co-culture with S2 ΔlysA with supplementation of isoleucine and valine/leucine (from Fig 3c and Fig 4c, respectively). Under the isoleucine supplemented condition, lag time is extended as droplet size increases. Under the valine and leucine supplemented condition, lag time remains largely the same and the growth curves overlap during the earliest phase. (b) Growth curves generated from the cross-feeding

autotroph ODE model, with different parameter values for amino acid secretion (α) and cellular requirement of the other amino acid for growth (β). For each scenario, three sets of initial conditions were specified corresponding to changes of the initial cell density when the droplet diameter was increased from 55 to 100, and then to 150 μm. It was noted that the growth dynamics in the scenario of α = 0.0001 and β = 0.001 (top right sub-plot) exhibited qualitatively similar pattens to those in experimental profiles under the isoleucine supplemented condition (top sub-plot in a.), whereas the simulated growth profiles in the scenario of α = 0.001 and β = 0.05 (bottom center sub-plot) greatly resemble those in experimental profiles under the valine and leucine supplemented condition (bottom sub-plot in a.). Other parameters of the model were set as: μ_{max} = 1, K_s = 0.1, and K = 100.

Note A: Dynamic model of cross-feeding auxotrophs and parameter fitting

We observed that the growth profile of each auxotroph in the cross-feeding bi-culture exhibited the "S"-shape, characteristic of logistic growth commonly assumed for monocultures. To test, from a theoretical point of view, the validity of employing the logistic equation to approximate the growth of cross-feeding amino acid auxotrophs, we modified the mechanistic dynamic model by Kerner et al. (108) *and carried out simulations to investigate growth dynamics resulting from interactions between two cross-feeding auxotrophs.*

We assumed logistic growth for each auxotroph and the Monod equation relating the specific growth rate to the concentration of a limiting substrate (in this case the cross-fed molecule). The governing equations are as follows:

$$\frac{dn_1}{dt} = \mu_1 n_1 \left(1 - \frac{n_1}{K_1}\right)$$

$$\frac{dn_2}{dt} = \mu_2 n_2 \left(1 - \frac{n_2}{K_2}\right)$$

$$\frac{dc_1}{dt} = \alpha_2 n_2 - \beta_1 \frac{dn_1}{dt}$$

$$\frac{dc_2}{dt} = \alpha_1 n_1 - \beta_2 \frac{dn_2}{dt}$$

$$\mu_1 = \frac{\mu_1^{max} c_1}{K_{s1} + c_1}$$

$$\mu_2 = \frac{\mu_2^{max} c_2}{K_{s2} + c_2}$$

where dynamic variables n_1 and n_2 are the cell densities of the two auxotrophs, c_1 and c_2 are the concentrations of the cross-fed amino acids, and μ_1 and μ_2 are the (instantaneous) specific growth rates of the two auxotrophs. Model parameters K_1 and K_2 are the carrying capacities of the two strains following the logistic growth assumption, α_1 and α_2 represent the auxotrophs' export rate of the amino acids, and β_1 and β_2 are their cellular requirement for the corresponding amino acids, respectively. μ_1^{max} and μ_2^{max} are the maximum specific growth rates of the two strains (i.e. at an infinitely high concentration of the limiting substrate); K_{s1} and K_{s2} are the "half-rate" constants in the Monod equation (i.e. concentration of the limiting substrate at which the growth rate is half of the maximum) for the two strains, respectively.

Using the above ordinary differential and algebraic equations with an initial condition of $c_1 = c_2 = 0$ (i.e. no supplementation of the cross-fed molecules at the start), we carried out dynamic simulations to explore various scenarios representing different degrees of interactions between the two cross-feeding auxotrophs. Specifically, one of the β parameters was changed to represent different levels of demand for the cross-fed metabolite. Our simulation results, as illustrated in Fig S2, demonstrate that the growth profile of each auxotroph in the cross-feeding bi-culture follows a general S-shape and can be approximated reasonably well with a logistic fit (R^2 value higher than 0.98). The fit is particularly strong at the lowest degree of cross-feeding (i.e. when β is set at the smallest value).

Appendix B. Supporting Information for Chapter 3

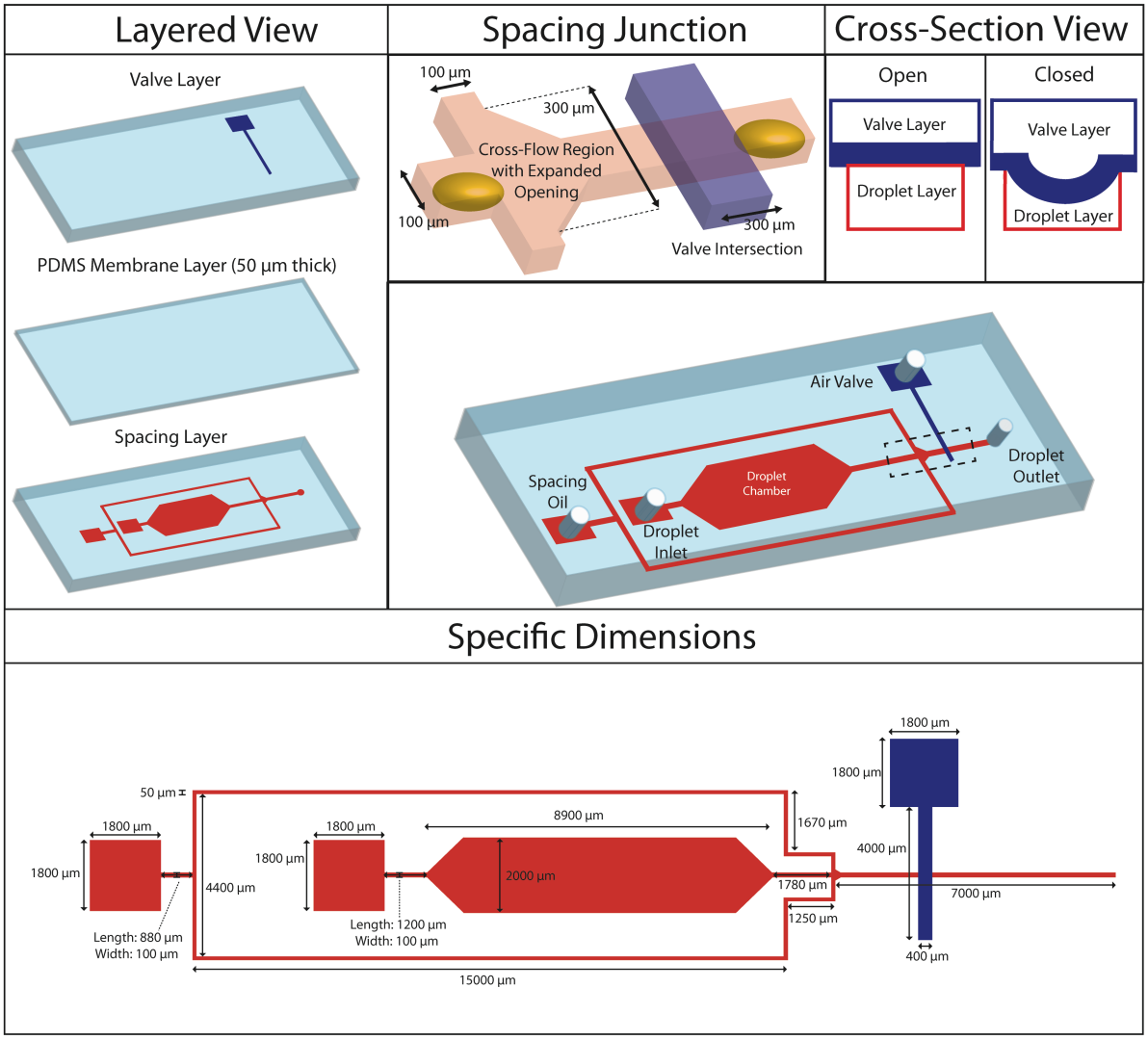


Figure B.1: Schematic of the droplet spacing device and the dimensions of the channels. The device is composed of three layers from bottom to top: the spacing layer, the thin PDMS membrane, and the valve layer. When assembled, the layers provide a membrane valve on top of the flow channels. When pressure from an external pump is applied, the membrane expands, preventing the flow of larger droplets while allowing oil flow to continue. When pressure is closed, the membrane contracts back to its normal configuration and allows droplets to pass. Droplets are flowed from the inlet into a larger droplet chamber to alleviate droplet merging that occurs with pressure build-up. Spacing oil with surfactant flows around to the spacing junction. The spacing junction, specified in the dashed box region of the device where spacing oil and droplet flow intersect, has an expanded opening to prevent droplets from shearing when passaged.

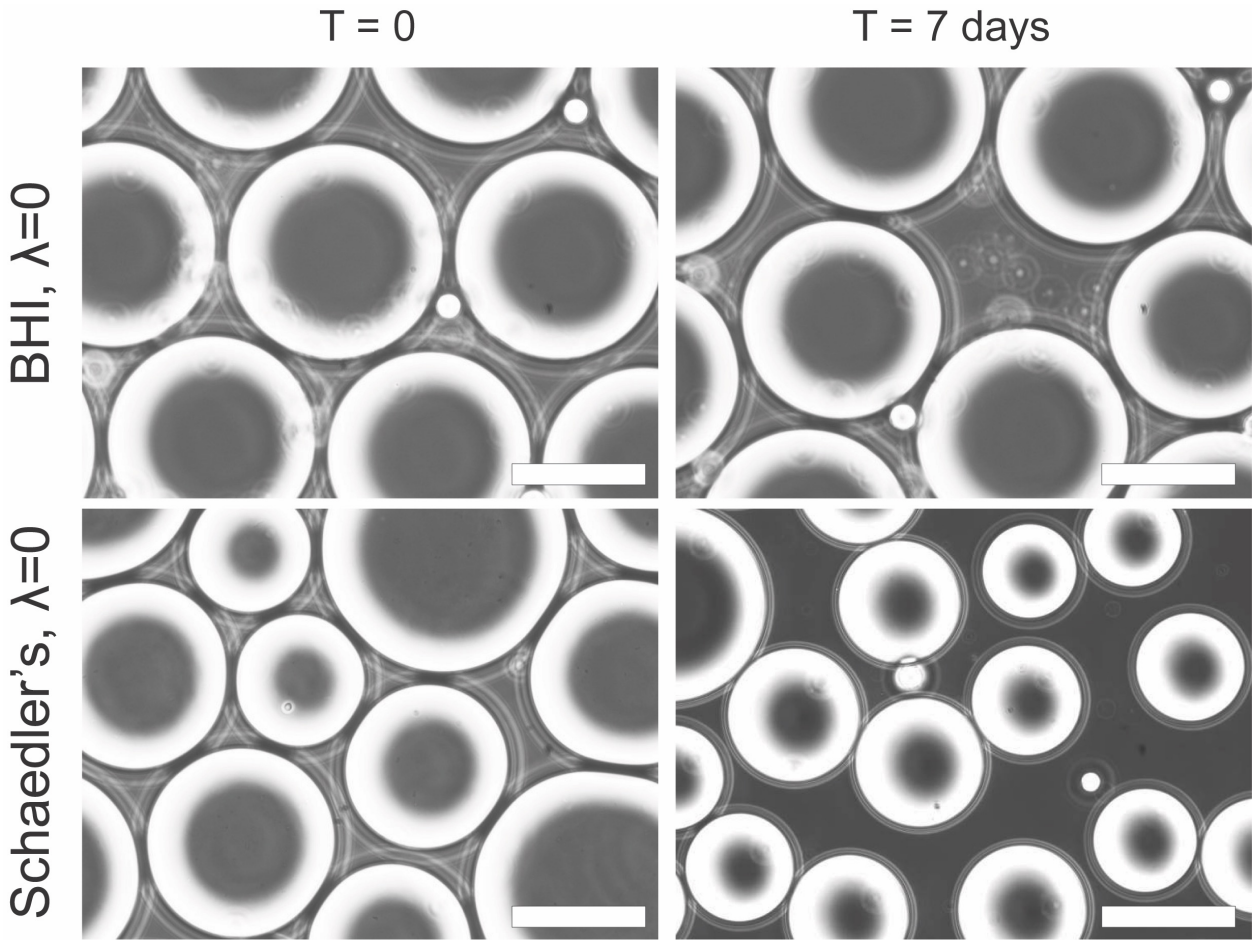


Figure B.2: Control droplets without encapsulated cells to demonstrate absence of cultivation of contaminant bacteria from reagents, tubing, and the device. Droplets were generated in the same fashion as specified with droplets with λ of 2 or 10, but no cells from the fecal sample were suspended in the BHI or Schaedler's media. Droplets were incubated anaerobically for a week, similarly to the droplets with λ of 2 or 10. Scale bar is 100 μm.

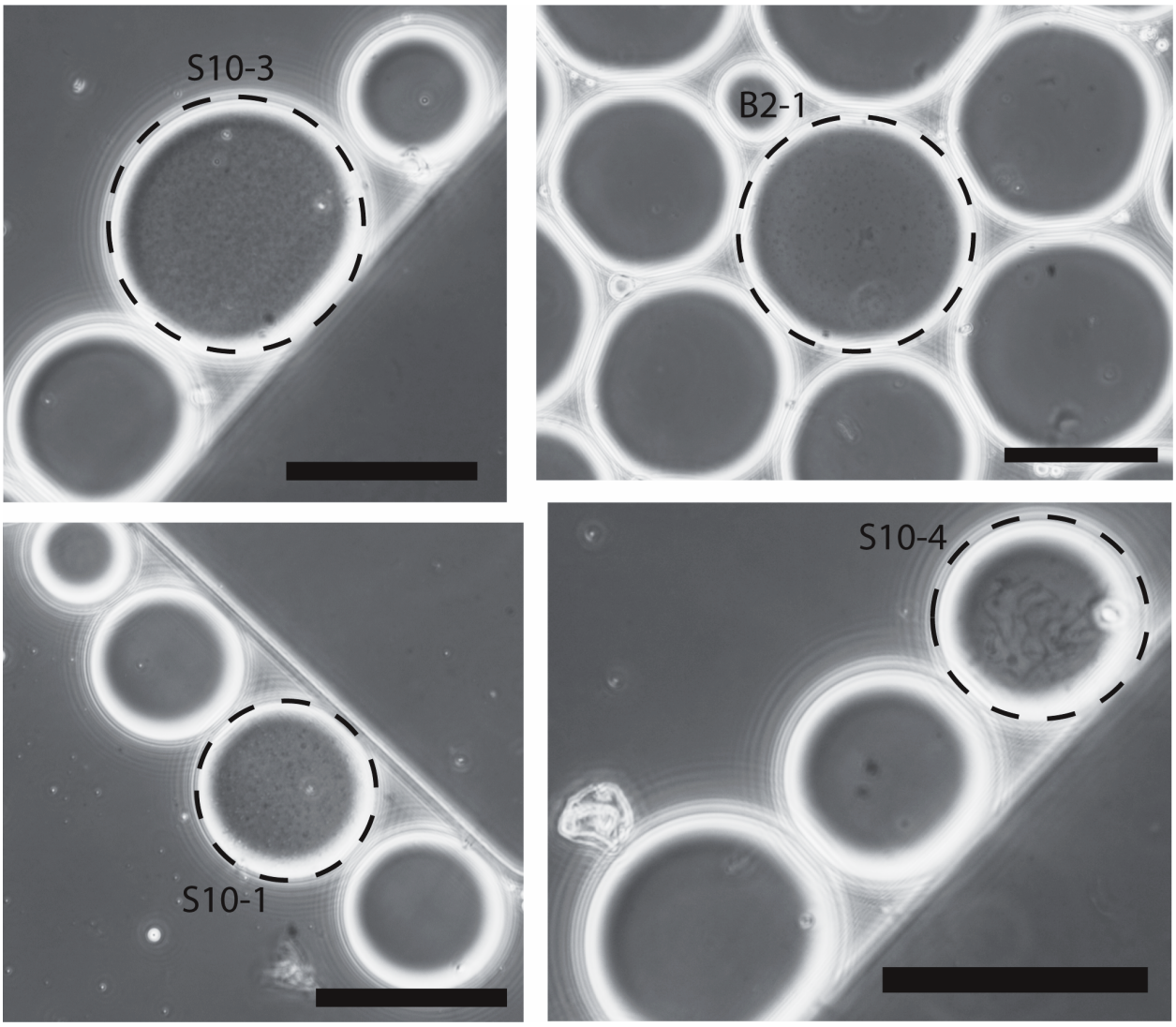


Figure B.3: Images of four selected droplets in the droplet spacing device's chamber after co-cultivation, but before individual droplet spacing. The droplets that were selected for downstream sequencing are specified with dashed borders and labels among the other droplets without co-growth. The selected droplets have clear degrees of co-growth that other surrounding droplets do not have, allowing for relatively easy visual determination of droplet growth and manual selection. Scale bar is 100 μm.

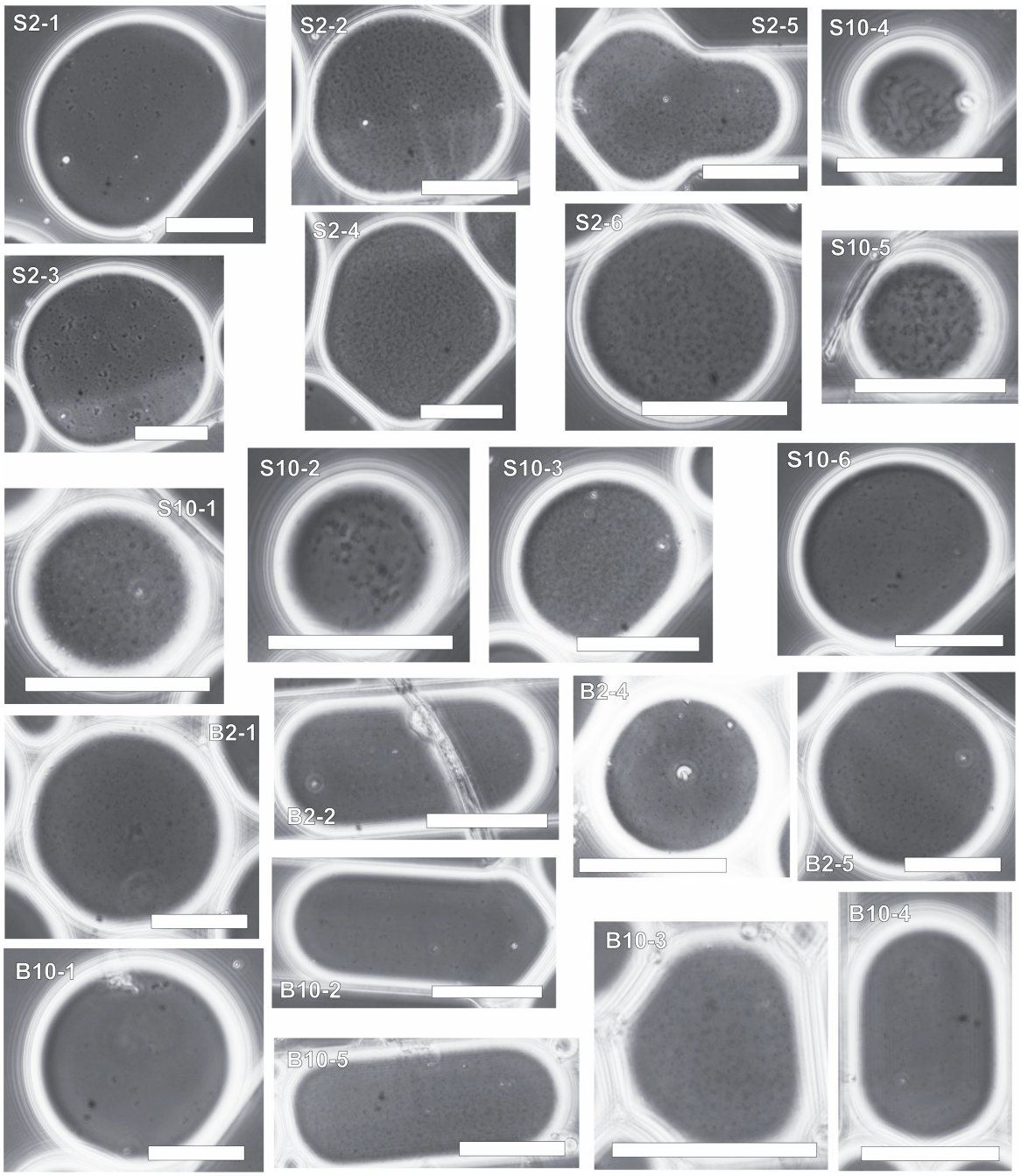


Figure B.4: Manually selected droplets with high degree of co-growth from droplet spacing. Individual droplets were passaged into separate microcentrifuge tubes, and the communities underwent whole-genome amplification with MDA. Droplet naming follows the convention of media used ("B" for BHI or "S" for Schaedler's), the initial λ (2 or 10), and a numerical value. All selected droplets are present, except for B2-3, whose image was lost due to an error in the imaging software. Scale bar is 100 μm.

OTU	Comparing between groups:	Indicator for group:	Indicator value	p value
Otu005	**S vs B**	**S**	**90.909088**	**0.021**
Otu003	**S vs B**	**S**	**75.248756**	**0.046**
Otu002	S vs B	S	61.274506	0.121
Otu027	S vs B	B	57.85714	0.191
Otu013	S vs B	S	53.571426	0.187
Otu004	S vs B	B	51.702126	0.369
Otu006	S vs B	S	51.282055	0.182
Otu015	B2 vs B10	B2	62.5	0.326
Otu001	B2 vs B10	B10	55.555557	0.581
Otu004	B2 vs B10	B2	55.555557	0.587
Otu027	B2 vs B10	B2	55.555557	0.587
Otu012	B2 vs B10	B10	53.333328	0.403
Otu016	S2 vs S10	S10	66.666672	0.17
Otu039	S2 vs S10	S2	66.666672	0.182
Otu010	S2 vs S10	S10	59.523808	0.272
Otu001	S2 vs S10	S10	54.545456	0.571
Otu003	S2 vs S10	S10	54.545456	0.57
Otu026	S2 vs S10	S10	53.333336	0.259
Otu041	S2 vs S10	S10	53.333336	0.264
Otu004	S2 vs S10	S2	52.083332	0.403
Otu006	S2 vs S10	S10	52.083332	0.41
Otu005	S2 vs S10	S10	50	1
Otu011	S2 vs S10	S2	50	0.246
Otu036	S2 vs S10	S10	50	0.259

Table B.1: Indicator analysis indicating the strength of the associations of specific OTUs to specific groups of droplets. Analysis was done on a binary presence/absence of OTUs. There were three comparisons conducted: (1) between the co-cultivated droplets with Schaedler's Media (S) and droplets with Brain-Heart Infusion Media (B), (2) within the Schaedler's Media droplets, droplets with the initial λ of 2 (S2) and droplets with the initial λ of 10 (S10), (3) and within the Brain-Heart Infusion Media droplets, droplets with the initial λ of 2 (B2) and droplets with the initial λ of 10 (B10). Higher indicator values (up to 100) signify stronger exclusive association of that OTU to that specific group.

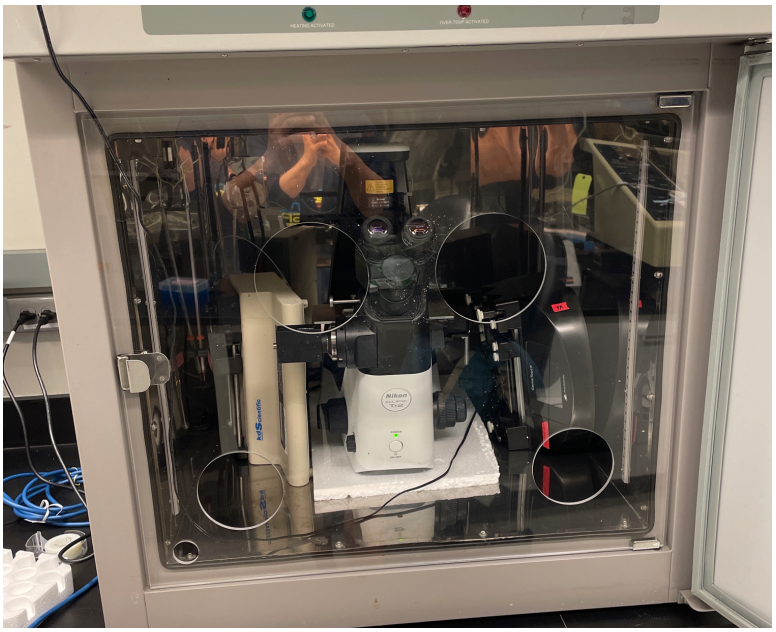
OTU Number	OTU Taxonomy	NCBI Top hit	Identity	Notes
Otu001	Staphylococcus(99)	Staphylococcus aureus strain S33 R 16S ribosomal RNA, complete sequence	1	
Otu002	Bacteroides(100)	Bacteroides timonensis strain AP1 16S ribosomal RNA, partial sequence	1	
Otu003	Butyricimonas(100)	Butyricimonas faecihominis strain 180-3 16S ribosomal RNA gene, partial sequence	0.99	
Otu004	Corynebacterium(100)	Corynebacterium tuberculostearicum strain Medalle X 16S ribosomal RNA, partial sequence	1	
Otu005	Parabacteroides(100)	Parabacteroides merdae strain JCM 9497 16S ribosomal RNA gene, partial sequence	1	
Otu006	Alistipes(100)	Alistipes shahii strain JCM 16773 16S ribosomal RNA gene, partial sequence	1	
Otu007	Bacteria_unclassified(97)	Traorella massiliensis strain Marseille-P3110 16S ribosomal RNA, partial sequence	0.99	
Otu008	Bacillales_unclassified(96)	Bacillus tropicus strain MCCC 1A01406 16S ribosomal RNA, partial sequence	0.99	
Otu009	Streptococcus(100)	Streptococcus pneumoniae strain NBRC 102642 16S ribosomal RNA gene, partial sequence	0.99	
Otu010	Alistipes(100)	Alistipes ihumii strain AP11 16S ribosomal RNA, partial sequence	0.97	
Otu011	Chryseobacterium(95)	Chryseobacterium endophyticum strain CC-YTH209 16S ribosomal RNA, partial sequence	1	
Otu012	Enterobacteriaceae_unclassified(80)	Escherichia marmotae strain HT073016 16S ribosomal RNA, partial sequence	1	
Otu013	Ruminococcaceae_unclassified(82)	Ruthenibacterium lactatiformans strain 585-1 16S ribosomal RNA, partial sequence	1	
Otu014	Streptococcus(100)	Streptococcus dentisani strain 7747 16S ribosomal RNA gene, partial sequence	1	
Otu015	Actinomycetales_unclassified(100)	Lawsonella clevelandensis strain X1036 16S ribosomal RNA, partial sequence	1	
Otu016	Bacteroides(100)	Bacteroides thetaiotaomicron strain VPI-5482 16S ribosomal RNA, partial sequence	1	
Otu017	Phascolarctobacterium(100)	Phascolarctobacterium faecium strain ACM 3679 16S ribosomal RNA gene, partial sequence	1	
Otu018	Neisseriaceae_unclassified(67)	Neisseria mucosa strain DSM 17611 16S ribosomal RNA gene, partial sequence	1	
Otu019	Comamonadaceae_unclassified(90)	Comamonas terrigena strain NBRC 12685 16S ribosomal RNA gene, partial sequence	1	
Otu020	Lactococcus(100)	Lactococcus taiwanensis strain 0905C15 16S ribosomal RNA gene, partial sequence	1	7 *Lactococcus lactis* hits at 100% identity
Otu021	Betaproteobacteria_unclassified(61)	Snodgrassella alvi strain wkB2 16S ribosomal RNA, complete sequence	0.93	
Otu022	Rikenellaceae_unclassified(77)	Alistipes ihumii strain AP11 16S ribosomal RNA, partial sequence	1	
Otu023	Anaerococcus(100)	Anaerococcus nagyae strain ENR0686 16S ribosomal RNA, partial sequence	0.99	
Otu024	Acinetobacter(100)	Acinetobacter johnsonii strain ATCC 17909 16S ribosomal RNA, partial sequence	1	
Otu025	Akkermansia(100)	Akkermansia muciniphila strain ATCC BAA-835 16S ribosomal RNA, partial sequence	1	
Otu026	Parabacteroides(100)	Parabacteroides distasonis strain ATCC 8503 16S ribosomal RNA, partial sequence	0.99	
Otu027	Propionibacterium(100)	Propionibacterium acnes subsp. elongatum strain K124 16S ribosomal RNA, partial sequence	1	
Otu028	Pseudomonas(59)	Pseudomonas japonica strain NBRC 103040 16S ribosomal RNA gene, partial sequence	0.99	
Otu029	Prevotella(100)	Prevotella buccalis strain JCM 12246 16S ribosomal RNA gene, partial sequence	1	
Otu030	Enterobacteriaceae_unclassified(100)	Klebsiella grimontii strain SB73 16S ribosomal RNA, partial sequence	1	
Otu031	Chryseobacterium(100)	Chryseobacterium cucumeris strain GSE06 16S ribosomal RNA, partial sequence	1	
Otu032	Sphingobacterium(100)	Sphingobacterium detergens strain 6.2S 16S ribosomal RNA gene, partial sequence	1	
Otu033	Clostridiales_unclassified(80)	[Clostridium] cellobioparum strain DSM 1351 16S ribosomal RNA gene, partial sequence	0.9	
Otu034	Truepera(100)	Truepera radiovictrix strain RQ-24 16S ribosomal RNA, partial sequence	0.95	
Otu035	Flavobacterium(100)	Flavobacterium johnsoniae strain UW101 16S ribosomal RNA, partial sequence	1	
Otu036	Firmicutes_unclassified(100)	Papillibacter cinnamivorans strain CIN1 16S ribosomal RNA gene, partial sequence	0.9	
Otu037	Pedobacter(89)	Pedobacter ureilyticus strain THG-T11 16S ribosomal RNA, partial sequence	1	
Otu038	Ezakiella(100)	[Bacteroides] coagulans strain EUH 581-73 16S ribosomal RNA gene, partial sequence	0.99	2 hits to *Ezakiella* at 95%
Otu039	Bacteria_unclassified(100)	Defluviimonas aestuarii strain BS14 16S ribosomal RNA gene, partial sequence	0.76	
Otu040	Varibaculum(100)	Varibaculum anthropi strain CCUG 31793 16S ribosomal RNA, partial sequence	0.99	
Otu041	Bacteria_unclassified(67)	Clostridium tepidum strain IEH 97212 16S ribosomal RNA, complete sequence	0.92	
Otu042	Comamonadaceae_unclassified(77)	Delftia tsuruhatensis strain NBRC 16741 16S ribosomal RNA gene, partial sequence	1	
Otu043	Actinomycetales_unclassified(54)	Propionibacterium acnes subsp. elongatum strain K124 16S ribosomal RNA, partial sequence	0.96	
Otu044	Alistipes(100)	Alistipes putredinis strain JCM 16772 16S ribosomal RNA gene, partial sequence	1	

Table B.2: Nucleotide BLAST identification of the 16S V4 region of the representative OTUs from the droplet against the NCBI 16S database.

Microfluidic OTUs Taxonomy	Bulk Pyroseuqencing OTU Taxonomy	Distance between	Relative abundance in patient sample (%)
Akkermansia(100)	Akkermansia(100)	0.0000	0.01
Parabacteroides(100)	Bacteria_unclassified(100)	0.0000	0.01
Comamonadaceae_unclassified(77)	Comamonadaceae_unclassified(100)	0.0000	0.01
Bacteroides(100)	Bacteroides(100)	0.0074	0.01
Enterobacteriaceae_unclassified(100)	Enterobacteriaceae_unclassified(93)	0.0074	11.18
Bacteroides(100)	Bacteroides(100)	0.0148	0.16
Bacteria_unclassified(67)	Christensenellaceae_unclassified(80)	0.0148	0.85
Rikenellaceae_unclassified(77)	Alistipes(95)	0.0148	0.09
Parabacteroides(100)	Parabacteroides(100)	0.0148	0.01
Enterobacteriaceae_unclassified(80)	Enterobacteriaceae_unclassified(93)	0.0222	11.18
Alistipes(100)	Alistipes(97)	0.0222	0.01
Firmicutes_unclassified(100)	Clostridiales_vadinBB60_group_ge(96)	0.0222	0.42
Alistipes(100)	Alistipes(97)	0.0222	0.01
Clostridiales_unclassified(80)	Ruminococcaceae_unclassified(100)	0.0296	0.01
Ruminococcaceae_unclassified(82)	Ruminococcaceae_unclassified(100)	0.0370	0.29

Table B.3: OTUs from the droplet B2-2 that appear in the bulk pyrosequencing library and the relative abundances they occur in the bulk. Overlap between representative V4 amplicons of the droplet OTUs and the partial V45 bulk pyrosequencing OTUs dataset were examined for distances less than 5% to determine whether an OTU was present in the bulk dataset.

Appendix C. Supporting Information for Chapter 4



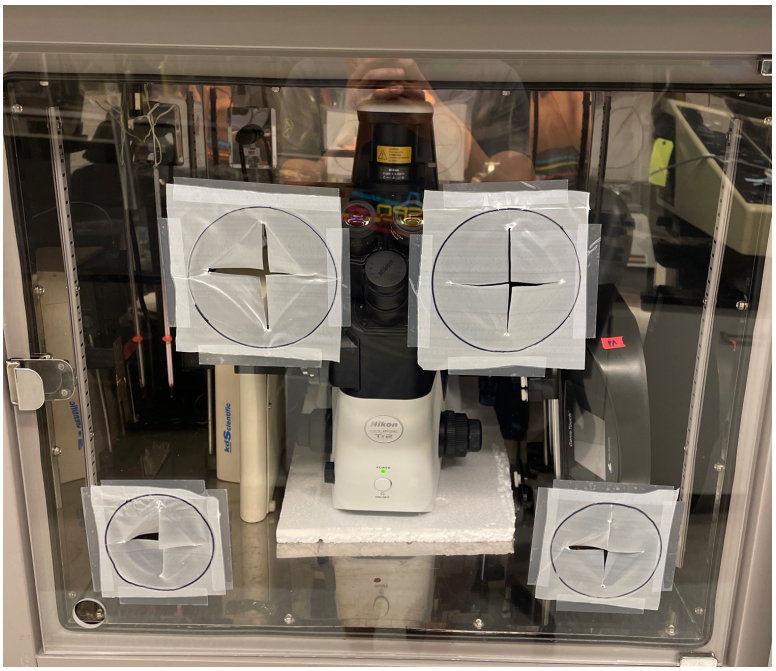


Figure C.1: Modified oven incubator for processing of agarose microdroplets above 37 ºC (Top) Glass door of the oven incubator was replaced with a clear acrylic sheet with holes cut out for entry of hands for microfluidic device operation. (Bottom) To prevent excess heat loss through the holes, thin plastic sheets with slits were taped onto the holes.

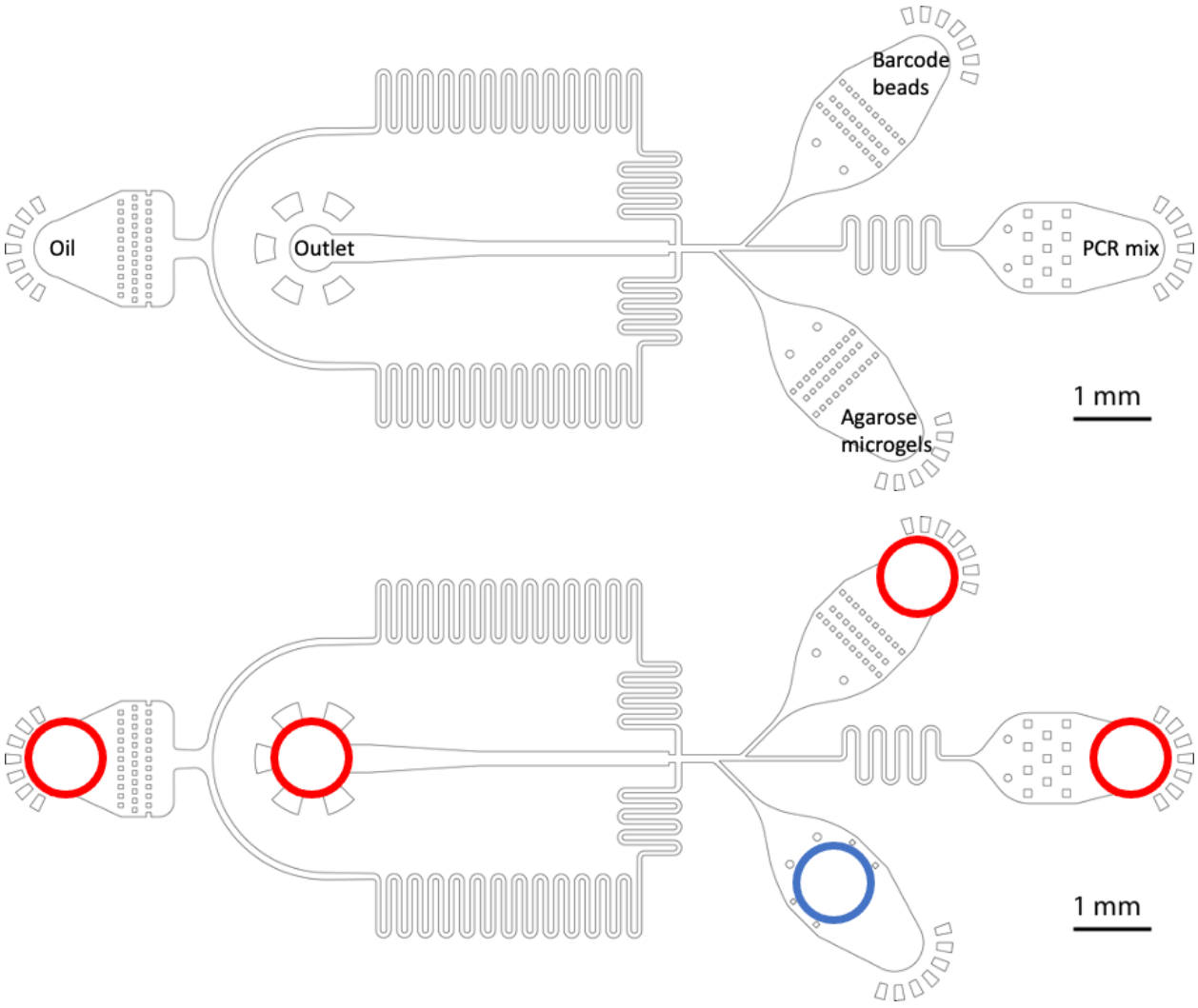


Figure C.2: Schematic of microfluidic device used for droplet barcoding, specifically the pairing of the barcode bead with agarose microgels with immobilized genomic DNA. (Top) Specifications for what each inlet and outlet are for. (Bottom) Holes are punched for the inserted of tubing for inlets and outlets to the PDMS device. Locations labelled in red are the holes are typically punched according to design specifications. For our protocol, an alternative location in blue was punched. Schematic provided by Dr. Emilis Gegevicius from Droplet Genomics Inc.

Detailed protocol for dsDNA with AMPure XP beads
Vortex the bottle of AMPure XPs vigorously to ensure all beads have been suspended
Add 75 uL AMPure XP per 50 uL of the sample you want to clean in individual 0.25 mL PCR tubes
Mix by pipetting and incubate the mixture at room temp for 20 minutes
Spin down the mixture briefly and place tubes on magnet rack for at least 10 minutes
Discard unbound fraction by pipetting
Wash beads with 200 uL of 80% vol/vol **FRESH** EtOH **twice**
Remove EtOH after beads are pelleted by the magnet rack, remove as much residual EtOH as possible with 20 uL pipette tip
Dry tubes with caps open for 30 minutes (cracking is okay)
Add and mix each sample's beads in 20 uL of 10 mM Tris HCl 8.0 by pipette-mixing, vortex, spin down briefly
Incubate at room temperature for 5 min for elution
Place tubes on magnet to concentrate after 2 minutes, transfer the eluate into a new PCR tube for processing downstream, store in fridge overnight

Table C.1: Protocol for clean-up of double-stranded DNA using AMPure XP beads.

primer_name	primer_seq	bc_seq
p7_1	CAAGCAGAAGACGGCATACGAGATTCGATGAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	CTCATCGA
p7_2	CAAGCAGAAGACGGCATACGAGATAACGATCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	GGATCGTT
p7_3	CAAGCAGAAGACGGCATACGAGATTAACGTGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	CCACGTTA
p5_1	AATGATACGGCGACCACCGAGATCTACACTAGATCGCACACTCTTTCCCTACACGACGCTCTTCCGATCT	TAGATCGC
p5_2	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTCTTCCGATCT	CTCTCTAT
p5_3	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	TATCCTCT

Table C.2: Primers used for library preparation of droplet barcode 16S amplicon libraries. "bc-seq" refers to the indices used for demultiplexing after Illumina sequencing. Taken from (182)

Bibliography

1. Dixon B. Power Unseen: How microbes rule the world. Oxford University Press; 1998.

2. Madsen EL. Identifying microorganisms responsible for ecologically significant biogeochemical processes. Nat Rev Microbiol. 2005;3(5):439–46.

3. Madsen EL. Microorganisms and their roles in fundamental biogeochemical cycles. Vol. 22, Current Opinion in Biotechnology. 2011. p. 456–64.

4. Strom SL. Microbial ecology of ocean biogeochemistry: A community perspective. Vol. 320, Science. American Association for the Advancement of Science; 2008. p. 1043–5.

5. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA. 1990 Jun 1;87(12):4576–9.

6. Castelle CJ, Banfield JF. Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. Vol. 172, Cell. 2018. p. 1181–97.

7. Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y, et al. Isolation of an archaeon at the prokaryote–eukaryote interface. Nature. 2020;577(7791):519–25.

8. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol. 2016;14(8).

9. Kaoutari A El, Armougom F, Gordon JI, Raoult D, Henrissat B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat Rev Microbiol. 2013;11(7):497–504.

10. Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. Nat Rev Microbiol. 2012;10.

11. Chung H, Pamp SJ, Hill JA, Surana NK, Edelman SM, Troy EB, et al. Gut immune maturation depends on colonization with a host-specific microbiota. Cell. 2012;149(7):1578–93.

12. Aldunate M, Srbinovski D, Hearps AC, Latham CF, Ramsland PA, Gugasyan R, et al. Antimicrobial and immune modulatory effects of lactic acid and short chain fatty acids produced by vaginal microbiota associated with eubiosis and bacterial vaginosis. Front Physiol. 2015;6:1–23.

13. Ma B, Forney LJ, Ravel J. Vaginal Microbiome: Rethinking Health and Disease. Annu Rev Microbiol. 2012;66(1):371–89.

14. Moeller AH, Caro-Quintero A, Mjungu D, Georgiev A V., Lonsdorf E V., Muller MN, et al. Cospeciation of gut microbiota with hominids. Science. 2016;353(6297):380–2.

15. Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, et al. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature. 2007;450:560–5.

16. Dunlap P V., Davis KM, Tomiyama S, Fujino M, Fukui A. Developmental and microbiological analysis of the inception of bioluminescent symbiosis in the marine fish Nuchequula nuchalis (Perciformes: Leiognathidae). Appl Environ Microbiol. 2008;74(24):7471–81.

17. Nyholm S V., McFall-Ngai MJ. The winnowing: Establishing the squid - Vibrios symbiosis. Vol. 2, Nature Reviews Microbiology. Nature Publishing Group; 2004. p. 632–42.

18. Hentschel U, Piel J, Degnan SM, Taylor MW. Genomic insights into the marine sponge microbiome. Nat Rev Microbiol. 2012;10(9):641–54.

19. Oldroyd GED. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. Nat Rev Microbiol. 2013;11(4):252–63.

20. Kwak MJ, Kong HG, Choi K, Kwon SK, Song JY, Lee J, et al. Rhizosphere microbiome structure alters to enable wilt resistance in tomato. Nat Biotechnol. 2018;36(11):1100–16.

21. Mylona P, Pawlowski K, Bisseling T. Symbiotic Nitrogen Fixation. Plant Cell. 1995;10:869–85.

22. Mueller UG, Juenger TE, Kardish MR, Carlson AL, Burns KM, Edwards JA, et al. Artificial Selection on Microbiomes To Breed Microbiomes That Confer Salt Tolerance to Plants. mSystems. 2021;6(6).

23. Dick G. Genomic Approaches in Earth and Environmental Sciences. 1st ed. John Wiley & Sons, Inc.; 2019.

24. Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive earth’s

biogeochemical cycles. Science. 2008:320;1034–9.

25. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: Integrating terrestrial and oceanic components. Science. 1998;281(5374):237–40.

26. Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, et al. Scientists' warning to humanity: microorganisms and climate change. Nat Rev Microbiol. 2019;17(9):569–86.

27. Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, et al. Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. Proc Natl Acad Sci USA. 2013;110(24):9824–9.

28. Buesseler KO, Lamborg CH, Boyd PW, Lam PJ, Trull TW, Bidigare RR, et al. Revisiting carbon flux through the ocean's twilight zone. Science. 2007;316(5824):567–70.

29. Boyd PW, Claustre H, Levy M, Siegel DA, Weber T. Multi-faceted particle pumps drive carbon sequestration in the ocean. Nature. 2019:568;327–35.

30. Suttle CA. Marine viruses - Major players in the global ecosystem. Nat Rev Microbiol. 2007;5(10):801–12.

31. Zhang M, Kong F, Tan X, Yang Z, Cao H, Xing P. Biochemical, morphological, and genetic variations in Microcystis aeruginosa due to colony disaggregation. World J Microbiol Biotechnol. 2007;23(5):663–70.

32. Singh BK, Bardgett RD, Smith P, Reay DS. Microorganisms and climate change: Terrestrial feedbacks and mitigation options. Nat Rev Microbiol. 2010;8(11):779–90.

33. Zimov SA, Schuur EAG, Stuart Chapin F. Permafrost and the global carbon budget. Vol. 312, Science. American Association for the Advancement of Science; 2006. p. 1612–3.

34. Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. Nature. 2018;558(7710):440–4.

35. Hover BM, Kim SH, Katz M, Charlop-Powers Z, Owen JG, Ternei MA, et al. Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. Nat Microbiol. 2018;3(4):415–22.

36. Baker BJ, Banfield JF. Microbial communities in acid mine drainage. FEMS Microbiology Ecology. 2003:44;139–52.

37. Webster TM, Reddy RR, Tan JY, Van Nostrand JD, Zhou J, Hayes KF, et al. Anaerobic Disposal of Arsenic-Bearing Wastes Results in Low Microbially Mediated Arsenic Volatilization. Environ Sci Technol. 2016;50(20):10951–9.

38. Qin J, Rosen BP, Zhang Y, Wang G, Franke S, Rensing C. Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. Proc Natl Acad Sci USA. 2006;103(7):2075–80.

39. Chiao TH, Clancy TM, Pinto A, Xi C, Raskin L. Differential resistance of drinking water bacterial populations to monochloramine disinfection. Environ Sci Technol. 2014;48(7):4038–47.

40. Little BJ, Blackwood DJ, Hinks J, Lauro FM, Marsili E, Okamoto A, et al. Microbially influenced corrosion—Any progress? Corrosion Science. 2020:170.

41. Keller J, Yuan Z, Blackall LL. Integrating process engineering and microbiology tools to advance activated sludge wastewater treatment research and development. Rev Environ Sci Biotechnol. 2002;1(1):83–97.

42. Albertsen M, Hansen LBS, Saunders AM, Nielsen PH, Nielsen KL. A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. ISME J. 2012;6(6):1094–106.

43. Kraft B, Tegetmeyer HE, Sharma R, Klotz MG, Ferdelman TG, Hettich RL, et al. The environmental controls that govern the end product of bacterial nitrate respiration. Science. 2014;345(6197):676–9.

44. Xu J, Hao J, Guzman JJL, Spirito CM, Harroff LA, Angenent LT. Temperature-Phased Conversion of Acid Whey Waste Into Medium-Chain Carboxylic Acids via Lactic Acid: No External e-Donor. Joule. 2018;2(2):280–95.

45. Herren CM, McMahon KD. Cohesion: a method for quantifying the connectivity of microbial communities. ISME J. 2017;1191:2426–38.

46. Minty JJ, Singer ME, Scholz S a, Bae C-H, Ahn J-H, Foster CE, et al. Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. Proc Natl Acad Sci USA. 2013;110(36):14592–7.

47. Shahab RL, Brethauer S, Davey MP, Smith AG, Vignolini S, Luterbacher JS, et al. A heterogeneous microbial consortium producing short-chain fatty acids from lignocellulose. Science. 2020;369(6507).

48. Zengler K, Zaramela LS. The social network of microorganisms — how auxotrophies shape complex communities. Nat Rev Microbiol. 2018 Jun 29;16(6):383–90.

49. McGlynn SE, Chadwick GL, Kempes CP, Orphan VJ. Single cell activity reveals direct electron transfer in methanotrophic consortia. Nature. 2015;526(7574):531–5.

50. Ruaud A, Esquivel-Elizondo S, de la Cuesta-Zuluaga J, Waters JL, Angenent LT, Youngblut ND, et al. Syntrophy via interspecies H2 transfer between christensenella and methanobrevibacter underlies their global cooccurrence in the human gut. MBio. 2020;11(1).

51. Embree M, Liu JK, Al-Bassam MM, Zengler K. Networks of energetic and metabolic interactions define dynamics in microbial communities. Proc Natl Acad Sci USA. 2015;112(50):15450–5.

52. Zelezniak A, Andrejev S, Ponomarova O, Mende DR, Bork P, Patil KR. Metabolic dependencies drive species co-occurrence in diverse microbial communities. Proc Natl Acad Sci USA. 2015;112(20):6449–54.

53. Lamont RJ, Koo H, Hajishengallis G. The oral microbiota: dynamic communities and host interactions. Nat Rev Microbiol. 2018;1.

54. Men Y, Feil H, VerBerkmoes NC, Shah MB, Johnson DR, Lee PKH, et al. Sustainable syntrophic growth of Dehalococcoides ethenogenes strain 195 with Desulfovibrio vulgaris Hildenborough and Methanobacterium congolense: global transcriptomic and proteomic analyses. ISME J. 2012;6(2):410–21.

55. Mee MT, Collins JJ, Church GM, Wang HH. Syntrophic exchange in synthetic microbial communities. Proc Natl Acad Sci USA. 2014;111(20):E2149-56.

56. Madigan MT, Martinko JM, Bender KS, Buckley DH, Stahl DA. Brock Biology of Microorganisms. 14th ed. Pearson; 2014.

57. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, et al. A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. Cell. 2016;167(5):1339-1353.e21.

58. Wang J, Qin J, Li Y, Cai Z, Li S, Zhu J, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490(7418):55–60.

59. Harke MJ, Steffen MM, Gobler CJ, Otten TG, Wilhelm SW, Wood SA, et al. A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, Microcystis spp. Harmful Algae. 2016;54:4–20.

60. Huisman J, Codd GA, Paerl HW, Ibelings BW, Verspagen JMH, Visser PM. Cyanobacterial blooms. Nat Rev Microbiol. 2018;1.

61. Rossetti S, Tomei MC, Nielsen PH, Tandoi V. “Microthrix parvicella", a filamentous bacterium causing bulking and foaming in activated sludge systems: a review of current knowledge. FEMS Microbiol Rev. 2005 Jan 1;29(1):49–64.

62. Evans MR, Grimm V, Johst K, Knuuttila T, de Langhe R, Lessells CM, et al. Do simple models lead to generality in ecology? Trends Ecol Evol. 2013;28(10):578–83.

63. Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, et al. Challenges in microbial ecology: building predictive understanding of community function and dynamics. ISME J. 2016;10:2557-2568.

64. Staley J. Measurement of In Situ Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats. Annu Rev Microbiol. 1985;39(1):321–46.

65. Eilers H, Pernthaler J, Glöckner FO, Amann R. Culturability and in situ abundance of pelagic Bacteria from the North Sea. Appl Environ Microbiol. 2000;66(7):3044–51.

66. Hugenholtz P, Tyson GW. Metagenomics. Nature. 2008 Sep 24;455(7212):481–3.

67. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA. 2011;108:4516–22.

68. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11(11):1144–1146.

69. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature. 2004;428(6978):37–43.

70. Dick GJ, Banfield JF. Community-wide analysis of microbial genome sequence signatures. Genome Biol. 2009;10:R85

71. Bashiardes S, Zilberman-Schapira G, Elinav E. Use of metatranscriptomics in microbiome research. Bioinform Biol Insights. 2016;10:19–25.

72. Mashego MR, Rumbold K, De Mey M, Vandamme E, Soetaert W, Heijnen JJ. Microbial metabolomics: Past, present and future methodologies. Biotechnology Letters. 2007;29:1–16.

73. Mosier AC, Li Z, Thomas BC, Hettich RL, Pan C, Banfield JF. Elevated temperature alters proteomic responses of individual organisms within a biofilm community. ISME J. 2014;9(1):1–15.

74. Mark Welch JL, Hasegawa Y, McNulty NP, Gordon JI, Borisy GG. Spatial organization of a model 15-member human gut microbiota established in gnotobiotic mice. Proc Natl Acad Sci USA. 2017;201711596.

75. Mark Welch JL, Rossetti BJ, Rieken CW, Dewhirst FE, Borisy GG. Biogeography of a human oral microbiome at the micron scale. Proc Natl Acad Sci USA. 2016;113(6):E791–800.

76. Orphan VJ. Methods for unveiling cryptic microbial partnerships in nature. Vol. 12, Current Opinion in Microbiology. 2009:12;231–7.

77. Fike DA, Gammon CL, Ziebis W, Orphan VJ. Micron-scale mapping of sulfur cycling across the oxycline of a cyanobacterial mat: a paired nanoSIMS and CARD-FISH approach. ISME J. 2008;2(7):749–59.

78. Hatzenpichler R, Connon SA, Goudeau D, Malmstrom RR, Woyke T, Orphan VJ. Visualizing in situ translational activity for identifying and sorting slow-growing archaeal−bacterial consortia. Proc Natl Acad Sci USA. 2016;201603757.

79. Harder W, Dijkhuizen L. Strategies of mixed substrate utilization in microorganisms. Philos Trans R Soc Lond B Biol Sci. 1982 Jun 11;297(1088):459–80.

80. Morowitz MJ, Denef VJ, Costello EK, Thomas BC, Poroyko V, Relman DA, et al. Strain-resolved community genomic analysis of gut microbial colonization in a premature infant. Proc Natl Acad Sci USA. 2011;108(3):1128–33.

81. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, et al. Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. Science. 2014 Apr 25;344(6182):416–20.

82. Arevalo P, Vaninsberghe D, Elsherbini J, Gore J, Polz Correspondence MF. A Reverse Ecology Approach Based on a Biological Definition of Microbial Populations. Cell. 2019;178:820-834.e14.

83. Cordero OX, Datta MS. Microbial interactions and community assembly at microscales. Curr Opin Microbiol. 2016;31:227–34.

84. Leventhal GE, Boix C, Kuechler U, Enke TN, Sliwerska E, Holliger C, et al. Strain-level diversity drives alternative community types in millimetre-scale granular biofilms. Nat Microbiol. 2018;3(11):1295–303.

85. Smith DJ, Tan JY, Powers MA, Lin XN, Davis TW, Dick GJ. Individual Microcystis colonies harbour distinct bacterial communities that differ by Microcystis oligotype and with time. Environ Microbiol. 2021;23(6):3020–36.

86. Venturelli OS, Carr A, Fisher G, Hsu R, Lau R, Bowen BP, et al. Deciphering microbial interactions in synthetic human gut microbiome communities. Mol Syst Biol. 2018;14(e8157).

87. Clark RL, Connors BM, Stevenson DM, Hromada SE, Hamilton JJ, Amador-Noguez D, et al. Design of synthetic human gut microbiome assembly and butyrate production. Nat Commun. 2021 May 31;12(1):1–16.

88. Baranwal M, Clark RL, Thompson J, Sun Z, Hero AO, Venturelli OS. Recurrent neural networks enable design of multifunctional synthetic human gut microbiome dynamics. Elife. 2022;11:1–36.

89. Ponomarova O, Sevin D, Mülleder M, Zirngibl K, Bulya K. Yeast creates a stable niche for symbiotic lactic acid bacteria through nitrogen overflow. Cell Syst. 2017;1–13.

90. Lawson CE, Harcombe WR, Hatzenpichler R, Lindemann SR, Löffler FE, O'Malley MA, et al. Common principles and best practices for engineering microbiomes. Nat Rev Microbiol. 2019;17(12):725–41.

91. Mazutis L, Gilbert J, Ung WL, Weitz DA, Griffiths AD, Heyman JA. Single-cell analysis and sorting using droplet-based microfluidics. Nat Protoc. 2013;8(5):870–91.

92. Park J, Kerner A, Burns MA, Lin XN. Microdroplet-Enabled Highly Parallel Co-Cultivation of Microbial Communities. Herrera-Estrella A, editor. PLoS One. 2011 Feb 25;6(2):e17019.

93. Co AD, van Vliet S, Kiviet DJ, Schlegel S, Ackermann M. Short-range interactions govern the dynamics and functions of microbial communities. Nat Ecol Evol. 2020;4(3):366–75.

94. Hsu RH, Clark RL, Tan JW, Romero PA, Venturelli OS. Rapid microbial interaction network inference in microfluidic droplets. Cell Syst. 2019;9(3):229–42.

95. Joanicot M, Ajdari A. Droplet control for microfluidics. Science. 2005;309(5736):887–8.

96. Tan JY, Saleski TE, Lin XN. The effect of droplet size on syntrophic dynamics in droplet-enabled microbial co-cultivation. Papadimitriou K, editor. PLoS One. 2022;17(3):e0266282.

97. Raina J-B, Fernandez V, Lambert B, Stocker R, Seymour JR. The role of microbial motility and chemotaxis in symbiosis. Nat Rev Microbiol. 2019;17(5):284–94.

98. Longsworth LG. Diffusion Measurements, at 25°, of Aqueous Solutions of Amino Acids, Peptides and Sugars. J Am Chem Soc. 1953;75(22):5705–9.

99. Hsu RH, Clark RL, Tan JW, Ahn JC, Gupta S, Romero PA, et al. Microbial Interaction Network Inference in Microfluidic Droplets. Cell Syst. 2019;9(3):229-242.e4.

100. Saleski TE, Kerner AR, Chung MT, Jackman CM, Khasbaatar A, Kurabayashi K, et al. Syntrophic co-culture amplification of production phenotype for high-throughput screening of microbial strain libraries. Metab Eng. 2019;54:232–43.

101. Kehe J, Kulesa A, Ortiz A, Ackerman CM, Thakku SG, Sellers D, et al. Massively parallel screening of synthetic microbial communities. Proc Natl Acad Sci USA. 2019;116(26):12804–9.

102. Jackman CM, Deans KW, Forney LJ, Lin XN. Microdroplet co-cultivation and interaction characterization of human vaginal bacteria. Integr Biol. 2019;11(3):69–78.

103. Carruthers DN, Byun CK, Cardinale BJ, Lin XN. Demonstration of transgressive overyielding of algal mixed cultures in microdroplets. Integr Biol. 2017;9(8):687–94.

104. Baret JC, Miller OJ, Taly V, Ryckelynck M, El-Harrak A, Frenz L, et al. Fluorescence-activated droplet sorting (FADS): Efficient microfluidic cell sorting based on enzymatic activity. Lab Chip. 2009;9(13):1850–8.

105. Atsumi S, Wu TY, Eckl EM, Hawkins SD, Buelter T, Liao JC. Engineering the isobutanol biosynthetic pathway in Escherichia coli by comparison of three aldehyde reductase/alcohol dehydrogenase genes. Appl Microbiol Biotechnol. 2010;85(3):651–7.

106. Atsumi S, Hanai T, Liao JC. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature. 2008;451(7174):86–9.

107. Scholz SA, Diao R, Wolfe MB, Fivenson EM, Lin XN, Freddolino PL. High-Resolution Mapping of the Escherichia coli Chromosome Reveals Positions of High and Low Transcription. Cell Syst. 2019;8(3):212-225.e9.

108. Kerner A, Park J, Williams A, Lin XN. A Programmable Escherichia coli Consortium via Tunable Symbiosis. Sandler SJ, editor. PLoS One. 2012 Mar 30;7(3):e34032.

109. Swinnen IAM, Bernaerts K, Dens EJJ, Geeraerd AH, Van Impe JF. Predictive modelling of the microbial lag phase: A review. Int J Food Microbiol. 2004;94(2):137–59.

110. Kaprelyantis A, Kell D. Do bacteria need to communicate with each other for growth? Trends Microbiol. 1996 Jun;4(6):237–42.

111. Voloshin SA, Kaprelyants AS. Cell-cell interactions in bacterial populations. Biochem. 2004;69(11):1268–75.

112. Ackerley DF, Barak Y, Lynch S V, Curtin J, Matin A. Effect of chromate stress on Escherichia coli K-12. J Bacteriol. 2006;188(9):3371–81.

113. Pande S, Shitut S, Freund L, Westermann M, Bertels F, Colesie C, et al. Metabolic cross-feeding via intercellular nanotubes among bacteria. Nat Commun. 2015;6(1):6238.

114. Shitut S, Ahsendorf T, Pande S, Egbert M, Kost C. Nanotube-mediated cross-feeding couples the metabolism of interacting bacterial cells. Environ Microbiol. 2019;21(4):1306–20.

115. Jahn M, Vorpahl C, Hübschmann T, Harms H, Müller S. Copy number variability of expression plasmids determined by cell sorting and droplet digital PCR. Microb Cell Fact. 2016;15(1):211.

116. Münch KM, Müller J, Wienecke S, Bergmann S, Heyber S, Biedendieck R, et al. Polar fixation of plasmids during recombinant protein production in Bacillus megaterium results in population heterogeneity. Appl Environ Microbiol. 2015;81(17):5976–86.

117. Boedicker JQ, Vincent ME, Ismagilov RF. Microfluidic confinement of single cells of bacteria in small volumes initiates high-density behavior of quorum sensing and growth and reveals its variability. Angew Chemie - Int Ed. 2009;48(32):5908–11.

118. Hansen RH, Timm AC, Timm CM, Bible AN, Morrell-Falvey JL, Pelletier DA, et al. Stochastic Assembly of Bacteria in Microwell Arrays Reveals the Importance of Confinement in Community Development. Li X, editor. PLoS One. 2016 May 6;11(5):e0155080.

119. Pin C, Baranyi J. Kinetics of single cells: Observation and modeling of a stochastic process. Appl Environ Microbiol. 2006;72(3):2163–9.

120. Li J, Attila C, Wang L, Wood TK, Valdes JJ, Bentley WE. Quorum sensing in Escherichia coli is signaled by AI-2/LsrR: Effects on small RNA and biofilm architecture. J Bacteriol. 2007;189(16):6011–20.

121. González Barrios AF, Zuo R, Hashimoto Y, Yang L, Bentley WE, Wood TK. Autoinducer 2 controls biofilm formation in Escherichia coli through a novel motility quorum-sensing regulator (MqsR, B3022). J Bacteriol. 2006;188(1):305–16.

122. Mahler L, Niehs SP, Martin K, Weber T, Scherlach K, Hertweck C, et al. Highly parallelized droplet cultivation and prioritization of antibiotic producers from natural microbial communities. Elife. 2021;10.

123. Watterson WJ, Tanyeri M, Watson AR, Cham CM, Shan Y, Chang EB, et al. Droplet-based high-throughput cultivation for accurate screening of antibiotic resistant gut microbes. Elife. 2020 Jun 1;9:1–22.

124. Tan JY, Wang S, Dick GJ, Young VB, Sherman DH, Burns MA, et al. Co-cultivation of microbial sub-communities in microfluidic droplets facilitates high-resolution genomic dissection of microbial 'dark matter.' Integr Biol. 2020 Nov 18;12(11):263–74.

125. Lagier J-C, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, et al. Culture of previously uncultured members of the human gut microbiota by culturomics. Nat Microbiol. 2016;1:16203.

126. Cross KL, Campbell JH, Balachandran M, Campbell AG, Cooper SJ, Griffen A, et al. Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. Nat Biotechnol. 2019;37(11):1314–21.

127. Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, et al. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature. 2016;533(7604):543–6.

128. Stewart EJ. Growing unculturable bacteria. J Bacteriol. 2012;194(16):4151–60.

129. McInerney MJ, Struchtemeyer CG, Sieber J, Mouttaki H, Stams AJM, Schink B, et al. Physiology, ecology, phylogeny, and genomics of microorganisms capable of syntrophic metabolism. Ann N Y Acad Sci. 2008;1125:58–72.

130. Pulschen AA, Bendia AG, Fricker AD, Pellizari VH, Galante D, Rodrigues F. Isolation of uncultured bacteria from antarctica using long incubation periods and low nutritional media. Front Microbiol. 2017;8:1346.

131. Chen F, Zhan Y, Geng T, Lian H, Xu P, Lu C. Chemical transfection of cells in picoliter aqueous droplets in fluorocarbon oil. Anal Chem. 2011;83(22):8816–20.

132. Zhan Y, Wang J, Bao N, Lu C. Electroporation of cells in microfluidic droplets. Anal Chem. 2009;81(5):2027–31.

133. McLean JS, Lombardo M-J, Badger JH, Edlund A, Novotny M, Yee-Greenbaum J, et al. Candidate phylum TM6 genome recovered from a hospital sink biofilm provides genomic insights into this uncultivated phylum. Proc Natl Acad Sci USA. 2013;110(26):E2390–9.

134. Dodsworth JA, Blainey PC, Murugapiran SK, Swingley WD, Ross CA, Tringe SG, et al. Single-cell and metagenomic analyses indicate a fermentative and saccharolytic lifestyle for members of the OP9 lineage. Nat Commun. 2013;4:1854.

135. Ma L, Kim J, Hatzenpichler R, Karymov M a, Hubert N, Hanan IM, et al. Gene-targeted microfluidic cultivation validated by isolation of a gut bacterium listed in Human Microbiome Project's Most Wanted taxa. Proc Natl Acad Sci USA. 2014;111(27):9768–73.

136. Berghuis BA, Yu FB, Schulz F, Blainey PC, Woyke T, Quake SR. Hydrogenotrophic methanogenesis in archaeal phylum Verstraetearchaeota reveals the shared ancestry of all methanogens. Proc Natl Acad Sci USA. 2019;116(11):5037–44.

137. Yu FB, Blainey PC, Schulz F, Woyke T, Horowitz MA, Quake SR. Microfluidic-based mini-metagenomics enables discovery of novel microbial lineages from complex environmental samples. Elife. 2017;6(650):1–20.

138. Wang S, Rossion M, Park J, Burns MA, Lin XN. Multiple displacement amplification of single droplet-cultivated microbes.

139. Wang S. Characterizing microbial communities using droplet microfluidic technology. 2017.

140. Drasar BS. Cultivation of anaerobic intestinal bacteria. J Pathol Bacteriol. 1967;94(2):417–27.

141. Schaedler RW, Dubos R, Costello R. The development of the bacterial flora in the gastrointestinal tract of mice. J Exp Med. 1958;122(1):59–66.

142. Starr SE, Killgore GE, Dowell VR. Comparison of Schaedler agar and trypticase soy-yeast extract agar for the cultivation of anaerobic bacteria. Appl Microbiol. 1971;22(4):655–8.

143. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. Appl Environ Microbiol. 2013;79(17):5112–20.

144. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537–41.

145. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014;42(D1).

146. Via M, Gignoux C, Burchard EG. The 1000 Genomes Project: New opportunities for research and social challenges. Genome Medicine. 2010:2.

147. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: A new versatile metagenomic assembler. Genome Res. 2017;27(5):824–34.

148. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: Quality assessment tool for genome assemblies. Bioinformatics. 2013 Apr 15;29(8):1072–5.

149. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015;3:e1319.

150. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: Rapid and sensitive classification of metagenomic sequences. Genome Res. 2016;26(12):1721–9.

151. Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, et al. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. Proc Natl Acad Sci USA. 2013;110(14):5540–5.

152. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55.

153. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: Rapid annotations using subsystems technology. BMC Genomics. 2008;9.

154. Karp PD, Latendresse M, Caspi R. The pathway tools pathway prediction algorithm. Stand Genomic Sci. 2011;5(3):424–9.

155. Karp PD, Paley S, Romero P. The Pathway Tools software. Bioinformatics. 2002;18:S225–32.

156. Caspi R. MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res. 2006;34(90001):D511–6.

157. Segata N, Börnigen D, Morgan XC, Huttenhower C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun. 2013;4.

158. Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH, et al. Environmental whole-genome amplification to access microbial populations in contaminated sediments. Appl Environ Microbiol. 2006;72(5):3291–301.

159. Wexler AG, Goodman AL, Voigt CA, Lu TK, Hennet T. An insider's perspective: Bacteroides as a window into the microbiome. Nat Microbiol. 2017;2(5):17026.

160. Everard A, Belzer C, Geurts L, Ouwerkerk JP, Druart C, Bindels LB, et al. Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. Proc Natl Acad Sci USA. 2013;110(22):9066–71.

161. Belzer C, Chia LW, Aalvink S, Chamlagain B, Piironen V, Knol J, et al. Microbial metabolic networks at the mucus layer lead to diet-independent butyrate and vitamin B12production by intestinal symbionts. MBio. 2017;8(5).

162. Ventura M, O'Flaherty S, Claesson MJ, Turroni F, Klaenhammer TR, van Sinderen D, et al. Genome-scale analyses of health-promoting bacteria: probiogenomics. Nat Rev Microbiol. 2009;7(1):61–71.

163. Gilliland, Stanley, Morelli, Lorenzo and Reid G. Health and Nutritional Properties of Probiotics in Food including Powder Milk with Live Lactic Acid Bacteria. Joint FAO/WHO Expert Consultation on Evaluation of Health and Nutritional Properties of Probiotics in Food Including Powder Milk with Live Lactic Acid Bacteria. 2001.

164. Shreiner AB, Kao JY, Young VB. The gut microbiome in health and in disease. Curr Opin Gastroenterol. 2015;31(1):69–75.

165. Lopetuso LR, Scaldaferri F, Petito V, Gasbarrini A. Commensal Clostridia: leading players in the maintenance of gut homeostasis. Gut Pathog. 2013;5(1):23.

166. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods. 2012;9(8):811–4.

167. Iqbal J, Hussain MM. Intestinal lipid absorption. Am J Physiol Metab. 2009;296(6):E1183–94.

168. Daniel H, Gholami AM, Berry D, Desmarchelier C, Hahne H, Loh G, et al. High-fat diet alters gut microbiota physiology in mice. ISME J. 2014;8(2):295–308.

169. Hildebrandt MA, Hoffmann C, Sherrill–Mix SA, Keilbaugh SA, Hamady M, Chen Y, et al. High-Fat Diet Determines the Composition of the Murine Gut Microbiome

Independently of Obesity. Gastroenterology. 2009;137(5):1716-1724.e2.

170. Koeth RA, Levison BS, Culley MK, Buffa JA, Wang Z, Gregory JC, et al. γ-butyrobetaine is a proatherogenic intermediate in gut microbial metabolism of L-carnitine to TMAO. Cell Metab. 2014;20(5):799–812.

171. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, et al. Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat Med. 2013;19(5):576–85.

172. Romano KA, Vivas EI, Amador-noguez D, Rey FE. Intestinal Microbiota Composition Modulates Choline Bioavailability. MBio. 2015;6(2):1–8.

173. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature. 2011;472(7341):57–65.

174. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, et al. Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. Science. 2014;344(6182):416–20.

175. Mußmann M, Hu FZ, Richter M, De Beer D, Preisler A, Jørgensen BB, et al. Insights into the genome of large sulfur bacteria revealed by analysis of single filaments. PLoS Biol. 2007;5(9):1923–37.

176. Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, et al. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. MBio. 2015;6(2).

177. Byrd AL, Belkaid Y, Segre JA. The human skin microbiome. Nat Rev Microbiol. 2018;16(3):143–55.

178. Cao Z, Chen F, Bao N, He H, Xu P, Jana S, et al. Droplet sorting based on the number of encapsulated particles using a solenoid valve. Lab Chip. 2013;13(1):171–8.

179. Mittelbach G. Community Ecology. Sinauer Associates; 2012. 46–50 p.

180. Pande S, Merker H, Bohl K, Reichelt M, Schuster S, de Figueiredo LF, et al. Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. ISME J. 2014;8(5):953–62.

181. Stämmler F, Gläsner J, Hiergeist A, Holler E, Weber D, Oefner PJ, et al. Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. Microbiome. 2016 Jun 21;4(1):1–13.

182. Sheth RU, Li M, Jiang W, Sims PA, Leong KW, Wang HH. Spatial metagenomic characterization of microbial biogeography in the gut. Nat Biotechnol. 2019;37(8):877–83.

183. Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, Klein AM, et al. Single-cell barcoding and sequencing using droplet microfluidics. Nat Protoc. 2017 Dec 8;12(1):44–73.

184. Ze X, David Y Ben, Laverde-Gomez JA, Dassa B, Sheridan PO, Duncan SH, et al. Unique organization of extracellular amylases into amylosomes in the resistant starch-utilizing human colonic firmicutes bacterium ruminococcus bromii. MBio. 2015;6(5):e01058-15.

185. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: A versatile open source tool for metagenomics. PeerJ. 2016;2016(10):e2584.

186. Stoddard SF, Smith BJ, Hein R, Roller BRK, Schmidt TM. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Res. 2015;43:D593.

187. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods. 2013;11(2):163–6.

188. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. Evaluation of methods for the extraction and purification of DNA from the human microbiome. PLoS One. 2012;7(3).

189. Yue JC, Clayton MK. A similarity measure based on species proportions. Commun Stat - Theory Methods. 2005;34(11):2123–31.

190. Lan F, Demaree B, Ahmed N, Abate AR. Single-cell genome sequencing at ultra-high-throughput with microfluidic droplet barcoding. Nat Biotechnol. 2017;35(7):640–6.

191. Eastburn DJ, Sciambi A, Abate AR. Picoinjection Enables Digital Detection of RNA with Droplet RT-PCR. PLoS One. 2013 Apr 26;8(4):e62961.

192. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci USA. 2011;108:4680–7.

193. Gajer P, Gajer P, Brotman RM, Brotman RM, Bai G, Bai G, et al. Temporal Dynamics of the Human Vaginal Microbiota. Sci Transl Med. 2012;4(132):132ra52--132ra52.

194. Rampersaud R, Planet PJ, Randis TM, Kulkarni R, Aguilar JL, Lehrer RI, et al. Inerolysin, a Cholesterol-Dependent Cytolysin Produced by Lactobacillus iners. J Bacteriol. 2011;193(5):1034–41.

195. Wheeler JD, Secchi E, Rusconi R, Stocker R. Not just going with the flow: The effects of fluid flow on bacteria and plankton. Annual Review of Cell and Developmental Biology. 2019;35:213–37.

196. Kim MK, Ingremeau F, Zhao A, Bassler BL, Stone HA. Local and global consequences of flow on bacterial quorum sensing. Nat Microbiol. 2016;1(1):1–5.

197. Liu F, Mao J, Kong W, Hua Q, Feng Y, Bashir R, et al. Interaction variability shapes succession of synthetic microbial ecosystems. Nat Commun. 2020;11(1):1–13.

198. Toprakcioglu Z, Knowles TPJ. Sequential storage and release of microdroplets. Microsystems Nanoeng. 2021;7(1):1–10.

199. Seymour JR, Amin SA, Raina JB, Stocker R. Zooming in on the phycosphere: The ecological interface for phytoplankton-bacteria relationships. Nat Microbiol. 2017;2.

200. Peng X, Wilken SE, Lankiewicz TS, Gilmore SP, Brown JL, Henske JK, et al. Genomic and functional analyses of fungal and bacterial consortia that enable lignocellulose breakdown in goat gut microbiomes. Nat Microbiol. 2021;6(4):499–511.

201. Kuchina A, Brettner LM, Paleologu L, Roco CM, Rosenberg AB, Carignano A, et al. Microbial single-cell RNA sequencing by split-pool barcoding. Science (80-). 2021;371(6531).

202. Amara RR, Vijaya S. Specific polyadenylation and purification of total messenger RNA from Escherichia coli. Nucleic Acids Res. 1997;25(17):3465–70.

203. Holland-Moritz DA, Wismer MK, Mann BF, Farasat I, Devine P, Guetschow ED, et al. Mass Activated Droplet Sorting (MADS) Enables High-Throughput Screening of Enzymatic Reactions at Nanoliter Scale. Angew Chemie - Int Ed. 2020;59(11):4470–7.

204. Kogawa M, Miyaoka R, Hemmerling F, Ando M, Yura K, Ide K, et al. Single-cell metabolite detection and genomics reveals uncultivated talented producer. PNAS Nexus. 2022;1(1):1–13.

205. Louca S, Polz MF, Mazel F, Albright MBN, Huber JA, O'Connor MI, et al. Function and functional redundancy in microbial systems. Nat Ecol Evol. 2018;2:936-943.